Psychological Methods
2011, Vol. 16, No. 2, 127–148

1082-989X/11/$12.00 DOI: 10.1037/a0021764

Accuracy in Parameter Estimation for Targeted Effects in Structural Equation Modeling: Sample Size Planning for Narrow Confidence Intervals

Keke Lai and Ken Kelley
University of Notre Dame

In addition to evaluating a structural equation model (SEM) as a whole, often the model parameters are of interest and confidence intervals for those parameters are formed. Given a model with a good overall fit, it is entirely possible for the targeted effects of interest to have very wide confidence intervals, thus giving little information about the magnitude of the population targeted effects. With the goal of obtaining sufficiently narrow confidence intervals for the model parameters of interest, sample size planning methods for SEM are developed from the accuracy in parameter estimation approach. One method plans for the sample size so that the expected confidence interval width is sufficiently narrow. An extended procedure ensures that the obtained confidence interval will be no wider than desired, with some specified degree of assurance. A Monte Carlo simulation study was conducted that verified the effectiveness of the procedures in realistic situations. The methods developed have been implemented in the MBESS package in R so that they can be easily applied by researchers.

Keywords: structural equation modeling, sample size planning, confidence interval, accuracy in parameter estimation, power analysis

Structural equation modeling (SEM) is a widely used method in the behavioral, educational, social, and managerial sciences, where variables are measured with error and/or latent constructs are theorized to exist (for reviews of SEM, see, e.g., Bentler & Dudgeon, 1996; Bollen, 2002; MacCallum & Austin, 2000). Studies that use SEM generally evaluate the performance of the hypothesized model with chi-square likelihood ratio tests and/or fit indices. For example, if the overall model fit is of interest, a chi-square test can be performed to examine whether the null hypothesis H_0: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ can be rejected at the designated significance level, where $\boldsymbol{\Sigma}$ is the population covariance matrix of the manifest variables and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the population model-implied covariance matrix. If a specific part of the model (e.g., a direct path from one latent variable to another, the covariance between two latent variables, a factor loading) is of interest, one can conduct a Wald test or a likelihood ratio test to probabilistically infer if the null hypothesis that those specific paths are zero can be rejected (for statistical theories of Wald test and likelihood ratio test applied to SEM, see, e.g., Mulaik, 2009, pp. 373–381).

However, as has been echoed many times in the literature, there are limitations to null hypothesis significance tests (NHSTs; see, e.g., Nickerson, 2000, which provides a comprehensive historical review; see also Cohen, 1994; Meehl, 1997; Schmidt, 1996). Many times, even before conducting NHST, it is known from substantive theories and experience that the null hypothesis is almost certainly false. For example, the correlation between two exogenous variables is almost certainly not zero for many psychological constructs, but a null hypothesis that the correlation is zero is usually tested anyway. Admittedly, NHST informs researchers of the directionality of the population effect and thus helps to understand the phenomenon under study (e.g., if the null value of zero, without loss of generality, is rejected, the population effect is inferred to be either positive or negative). However, the results of a dichotomous (reject or fail-to-reject) significance test should not be the endpoint of scientific inquiry, because knowing whether the population effect is larger than zero does not answer the question about the magnitude of the population effect, which is often of ultimate interest. It would better facilitate the cumulation of knowledge of a discipline if studies also report plausible values for the population effect size of interest, which can be fulfilled via confidence intervals (CIs). A CI indicates probabilistically not only what the population parameter is not (i.e., those values excluded by the CI) but a range of plausible values (i.e., those values contained within the CI), and thus generally provides more information than does NHST (for more detailed discussions on the differences between CI and NHST, see, e.g., Harlow, Mulaik, & Steiger, 1997; Kelley & Maxwell, 2003; Kelley, Maxwell, & Rausch, 2003; Maxwell, Kelley, & Rausch, 2008).

Recent authoritative sources have urged that applied research report effect sizes and their corresponding CIs. For example, the *Publication Manual of the American Psychological Association* (American Psychological Association [APA], 2001) states that the "reporting of confidence intervals . . . for effect sizes . . . can be . . . extremely effective. . . . The use of confidence intervals is therefore strongly recommended" (p. 22). Moreover, in the latest *Publication Manual* (APA, 2010), the APA continues to increase its stress on using CIs, by stating that NHST is "but a starting point" and additional reporting elements such as effect sizes and CIs are necessary to "convey the most complete meaning of the results" and are "minimum expectations for all APA journals" (p. 33). In

This article was published Online First March 21, 2011.

Keke Lai, Department of Psychology, University of Notre Dame; Ken Kelley, Department of Management, University of Notre Dame.

Correspondence concerning this article should be addressed to Ken Kelley, Department of Management, University of Notre Dame, Notre Dame, IN 46556. E-mail: kkelley@nd.edu

the same vein, the American Educational Research Association (2006) states in its research guidelines that, for each of the major statistical results, there should be (a) an effect size of some kind and (b) an indication of the uncertainty of that effect size estimation, such as a CI (p. 37). Similar arguments are also available in the medical sciences (e.g., International Committee of Medical Journal Editors, 2004) and physiology (e.g., Curran-Everett & Benos, 2004), among others.

Effect sizes in the SEM context can be categorized into two types: (a) omnibus effects, which refer to the fit of the model as a whole, such as the nonnormed fit index (Bentler & Bonett, 1980) and the root mean square error of approximation (Browne & Cudeck, 1992; Steiger & Lind, 1980); and (b) targeted effects, which refer to the effect of a specific part of the model, such as the coefficient of a structural path, the covariance between two latent variables, or a factor loading. In addition to evaluating the fit of the model as a whole, many times a specific part of the model is also of interest. Moreover, a finding of a good overall fit does not imply that the targeted effects of interest are strong and/or practically important. It is possible for such effects to be weak or even trivial in a model with a good overall fit, if the dynamics in the model are in fact instead explained by large residual variances for endogenous variables (MacCallum & Austin, 2000). Therefore, it is critically important to study the targeted effects in a model so that the researcher can better understand the strength of the specific relationships of interest and evaluate substantive theories. To achieve this goal, instead of performing the traditional NHST to infer if some path coefficients and/or covariances or correlations equal zero, CIs can be formed, because CIs can convey not only the direction but the magnitude of the population parameters of interest.

Whenever an interval estimate is of interest, all other things being equal, it is more desirable to observe a narrow CI, as compared with a wider one, because a narrow CI includes a narrower range of plausible parameter values and thus is more informative. Congruent with the increasing emphasis on reporting CIs for effect sizes, a study can be designed with the goal of obtaining a narrow CI. Planning the sample size for a study with the goal of obtaining a narrow CI dates back to at least Guenther (1965) and Mace (1964), and is becoming popular recently as an alternative to, or supplement for, power analysis, as a result of the increasing emphasis on effect size estimation and CI formation. This approach to sample size planning has been set in a framework termed *accuracy in parameter estimation* (AIPE; e.g., Kelley, 2007c, 2008; Kelley & Maxwell, 2003; Kelley & Rausch, 2006), where the goal is to achieve a sufficiently narrow CI so that the parameter estimate will have a high degree of expected accuracy (for a review of applications of AIPE, see Maxwell et al., 2008). The desired CI width depends on the goals of the study and is determined by the researcher on a case-by-case basis, much like setting the desired level of statistical power if one was planning sample size from the power approach.

Throughout this article, we emphasize the input–output relationship of the sample size planning procedure. The input in the present context is usually a set of presumed population values. The focus of the present article is to show, given a certain set of input specifications, how to calculate the necessary sample size (i.e., output). To what extent the sample size calculated is approximate (i.e., the CI width obtained in a study compared with the desired width) depends on the quality of the input. To help estimate the necessary input values, we then propose several practical methods as a supplement to the literature review, which is generally the primary source for estimating the input parameters.

The present article develops methods to plan sample sizes for targeted effects in SEM so that the CIs for the parameters of interest will be sufficiently narrow. The standard method plans for the sample size so that the expected value of a CI width will be no larger than desired. Because the CI width is a random variable, setting the expected width to be sufficiently narrow does not guarantee that a CI observed in a particular study will be sufficiently narrow. This standard method is extended so that the CI obtained in a particular study will be no wider than desired, with some specified, usually high (e.g., 80%, 90%, 99%), degree of assurance. Because the sample size planning methods are based on standard CI formation methods in the existing SEM literature, which are technically approximate (i.e., based on asymptotic distributions but used at finite sample sizes), a Monte Carlo simulation study is conducted to evaluate the effectiveness of the proposed sample size planning methods. These sample size planning methods have been implemented in specialized R functions (R Development Core Team, 2010) in the MBESS package (Kelley, 2007a, 2007b; Kelley & Lai, 2010), so that they can be easily applied by researchers.[1]

CI Formation for SEM Model Parameters

There are several methods to construct CIs for SEM parameters; two of the most widely used are (a) maximum likelihood estimation (MLE; for a thorough discussion on likelihood methods, see, e.g., Pawitan, 2001; for applications in the SEM context, see, e.g., Bollen, 1989) and (b) the bootstrap approach (Efron & Tibshirani, 1993; for applications of the bootstrap to SEM, see Beran & Srivastava, 1985; Bollen & Stine, 1993; Yuan & Hayashi, 2006). Moreover, maximum likelihood standard errors can be based on the expected information matrix, the observed information matrix, or a sandwich-type covariance matrix (e.g., Arminger & Schoenberg, 1989; Bollen, 1989; Browne & Arminger, 1995; Shapiro, 1983; White, 1982; Yuan & Hayashi, 2006). The present article bases CI formation on the expected information matrix to develop sample size planning methods, for the following reasons: (a) Other methods require raw data and thus are not suitable for sample size planning, which is necessarily performed before data collection; (b) almost all SEM programs (e.g., AMOS, EQS, LISREL, Mplus, SAS PROC CALIS, the sem package in R) use (at least by default) the expected information matrix to construct CIs; and (c) different methods tend to give similar CIs if the assumptions (e.g., multivariate normality) are satisfied and the model is correctly specified.

We briefly review the CI formation in this section, because it is important when we formally present the sample size planning methods. Detailed analytical derivation of the standard errors of MLE for SEM model parameters is available in Yuan and Hayashi (2006). We developed the sample size planning methods based on the inferential methods most commonly used in the SEM literature

[1] MBESS was originally an acronym for Methods for Behavioral, Educational, and Social Sciences, but is used as a stand-alone package title now.

and by SEM software (i.e., maximum likelihood with the expected information matrix). The performance of the sample size planning procedures depends on the performance of the MLE, and in situations where MLE is less effective, the sample size planned may be different from the idealized sample size. Because maximum likelihood techniques are based on the asymptotic properties of the estimation, at any finite sample size inferences given by MLE are technically approximate. Therefore, we empirically evaluate the performance of our sample size planning procedures and MLE with a Monte Carlo simulation study in a later section. It should be made clear that the errors arising from using maximum likelihood asymptotic properties to approximate the behavior of MLE at finite sample sizes exist in any situation where maximum likelihood is applied. Such errors are not caused by sample size planning methods and are not unique to this article.

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ denote a random sample from a p-variate multivariate normal distribution with population covariance matrix $\boldsymbol{\Sigma}$. The present interest is to model the covariance matrix with a proposed structure, and the covariance matrix implied by the model is denoted as $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector containing the model parameters. If the model is correct, that is, $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}$, and $\mathbf{x}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ $(i = 1, \ldots, n)$, the log-likelihood function used to fit the model is (e.g., Bollen, 1989, pp. 132–134)

$$l(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} l_i(\hat{\boldsymbol{\theta}})$$

$$= -\frac{n}{2}\left\{\ln|\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})| + tr[\mathbf{S}\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})^{-1}]\right\} + c, \tag{1}$$

where

$$l_i(\hat{\boldsymbol{\theta}}) = -\frac{1}{2}\ln|\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})| - \frac{1}{2}(\mathbf{x}_i - \bar{\mathbf{x}})^T\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}) + c', \tag{2}$$

$\mathbf{S}$ is the observed $p \times p$ covariance matrix based on a sample of size n, $\bar{\mathbf{x}}$ is the sample mean vector, and c and c' are constants that have no influence on the choice of $\hat{\boldsymbol{\theta}}$.

Let $\hat{\boldsymbol{\theta}}_n$ denote the maximum likelihood estimator of $\boldsymbol{\theta}$ based on a sample of size n. Standard asymptotic theories state that $\hat{\boldsymbol{\theta}}_n$, the MLE for $\boldsymbol{\theta}$, converges in distribution to a multivariate normal distribution as $n \to \infty$ (e.g., Lehmann, 1998; Pawitan, 2001):

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{L} N(\mathbf{0}, \mathbf{I}(\boldsymbol{\theta})^{-1}), \tag{3}$$

where $\mathbf{I}(\boldsymbol{\theta})$ is the expected information matrix and can be considered as the function $\mathbf{I}(\ )$ evaluated at $\boldsymbol{\theta}$. That is,

$$\mathbf{I}(\boldsymbol{\theta}) = \mathrm{E}\big[-l_i''(\hat{\boldsymbol{\theta}})\big]_{\hat{\boldsymbol{\theta}}=\boldsymbol{\theta}},$$

where l_i'' refers to the second derivative of the log-likelihood l_i (for detailed discussion on MLE and the expected information matrix. see, e.g., Pawitan, 2001). However, the population parameter $\boldsymbol{\theta}$ is unknown, and thus $\mathbf{I}(\boldsymbol{\theta})$ is unknown and needs to be estimated. Because $\hat{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_n$ converges in probability to $\boldsymbol{\theta}$ as $n \to \infty$:

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}. \tag{4}$$

Under the commonly used assumption that the function $\mathbf{I}(\ )$ is differentiable, plugging $\hat{\boldsymbol{\theta}}_n$ into $\mathbf{I}(\ )$ leads to

$$\mathbf{I}(\hat{\boldsymbol{\theta}}_n) \xrightarrow{P} \mathbf{I}(\boldsymbol{\theta}). \tag{5}$$

The assumption that $\mathbf{I}(\ )$ is differentiable is a condition included in the so-called regularity conditions (e.g., Lehmann, 1998; Pawitan, 2001), which are commonly assumed in the literature when applying MLE.

If one takes the sum of the information over n individuals and calculates the mean, Equation 5 leads to

$$\frac{1}{n}\mathbf{I}_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{P} \mathbf{I}(\boldsymbol{\theta}), \tag{6}$$

which states that the estimated expected information matrix converges in probability to the expected information matrix as $n \to \infty$. Notice that the difference between Equations 5 and 6 is that $\mathbf{I}_n(\hat{\boldsymbol{\theta}}_n)$ is the sum of information over n individuals, that is,

$$\mathbf{I}_n(\hat{\boldsymbol{\theta}}_n) = \mathrm{E}\left[\sum_{i=1}^{n} -l_i''(\hat{\boldsymbol{\theta}})\right]_{\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}_n},$$

whereas $\mathbf{I}(\hat{\boldsymbol{\theta}}_n)$ is based on one observation, that is,

$$\mathbf{I}(\hat{\boldsymbol{\theta}}_n) = \mathrm{E}\big[-l_i''(\hat{\boldsymbol{\theta}})\big]_{\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}_n}.$$

Therefore, if one replaces $\mathbf{I}(\boldsymbol{\theta})^{-1}$ in Equation 3 with its estimate, at a given finite sample size n, $\hat{\boldsymbol{\theta}}_n$ approximately follows a multivariate normal distribution with covariance matrix $\mathbf{I}_n(\hat{\boldsymbol{\theta}}_n)^{-1}$:

$$\hat{\boldsymbol{\theta}}_n \approx N(\boldsymbol{\theta}, \hat{\mathbf{K}}), \tag{7}$$

where $\hat{\mathbf{K}} = [\hat{k}_{ij}] = \mathbf{I}_n(\hat{\boldsymbol{\theta}}_n)^{-1}$. Because $\mathbf{I}_n(\hat{\boldsymbol{\theta}}_n)^{-1} = \mathbf{I}(\hat{\boldsymbol{\theta}}_n)^{-1}/n$, Equation 7 is equivalent to

$$\hat{\boldsymbol{\theta}}_n \approx N(\boldsymbol{\theta}, \hat{\mathbf{H}}/n), \tag{8}$$

where $\hat{\mathbf{H}} = [\hat{h}_{ij}] = \mathbf{I}(\hat{\boldsymbol{\theta}}_n)^{-1}$.

After obtaining the approximate distribution of $\hat{\boldsymbol{\theta}}_n$, it is now possible to form CIs for model parameters of interest. Let θ_j, the jth element in $\boldsymbol{\theta}$, denote the model parameter of interest. Because multivariate normality implies univariate normality for all variables, $\hat{\boldsymbol{\theta}}_n$ following a multivariate normal distribution states that $\hat{\theta}_j$ at sample size n follows a normal distribution. Then the $(1 - \alpha)$ 100% CI for θ_j is

$$\mathrm{CI}_{(1-\alpha)} = \big[\hat{\theta}_j - z_{1-\alpha/2}\sqrt{\hat{k}_{jj}} \le \theta_j \le \hat{\theta}_j + z_{1-\alpha/2}\sqrt{\hat{k}_{jj}}\big], \tag{9}$$

or equivalently

$$\mathrm{CI}_{(1-\alpha)} = \left[\hat{\theta}_j - z_{1-\alpha/2}\sqrt{\frac{\hat{h}_{jj}}{n}} \le \theta_j \le \hat{\theta}_j + z_{1-\alpha/2}\sqrt{\frac{\hat{h}_{jj}}{n}}\right], \tag{10}$$

where $z_{1-\alpha/2}$ refers to the $(1 - \alpha/2)$th quantile of the standard normal distribution. In summary, the CI formation reviewed in this section is based on MLE with expected information matrix and serves as the basis of our sample size planning methods.

Accuracy in Parameter Estimation for SEM Model Parameters

Let w denote the full CI width obtained in a particular study. Let ω denote the desired CI width for the population parameter of interest, and it is specified by the researcher a priori according to the purpose of the study. The concept of "sufficiently narrow" (i.e., the value of ω) is loosely defined in the present article, because the specific value varies dramatically depending on at least (a) the goals of the study, (b) the measurement unit of the variable of interest, and (c) the conventions in the particular area of study. Therefore, the desired CI width needs to be specified by the researcher on a case-by-case basis. For example, if the study is more confirmatory in nature, the researcher may generally plan to obtain a smaller ω value, whereas the researcher may be satisfied with a larger ω value in a study that is more exploratory in nature. For the reasons discussed above, sample size planning methods, both in the literature and in this article, focus on how to achieve the goal $w \leq \omega$ instead of how to set the goal (i.e., the value of ω).

We need to emphasize that there is no such thing as the correct width for a study. Given the purposes of a study, a (continuous) range of values may be considered as suitable, and this range can sometimes be quite large. Instead of asking how to choose the desired CI width, sometimes it is easier for the researcher to start by precluding undesirable values. Because there is no "correct" ω value, any value is in fact reasonable as long as it does not make the CI too wide to be informative. We also discuss a practical method to facilitate the task of specifying ω value in a later section.

Planning Sample Size for the Expected CI Width to Be Sufficiently Narrow

Planning sample size from the AIPE perspective aims to find the minimum n such that $w \leq \omega$. Because the realized width w is a random variable, one approach is to aim at the expectation of w and obtain the minimum n so that $\mathrm{E}[w] \leq \omega$. Based on Equation 10, the full CI width obtained is

$$w = 2z_{1-\alpha/2}s_{\hat{\theta}_j} = 2z_{1-\alpha/2}\sqrt{\frac{\overline{h_{jj}}}{n}}, \tag{11}$$

where $s_{\hat{\theta}_j}$ is the standard error for $\hat{\theta}_j$, the point estimate of the model parameter of interest. Taking the expectation on both sides of Equation 11 gives

$$\begin{aligned}\mathrm{E}[w] &= \mathrm{E}[2z_{1-\alpha/2}s_{\hat{\theta}_j}] \\ &= 2z_{1-\alpha/2}\mathrm{E}[s_{\hat{\theta}_j}] \\ &= 2z_{1-\alpha/2}\sigma_{\hat{\theta}_j} \\ &= 2z_{1-\alpha/2}\sqrt{\frac{\overline{h_{jj}}}{n}}.\end{aligned} \tag{12}$$

Although strictly speaking $\mathrm{E}[s_{\hat{\theta}_j}] \neq \sigma_{\hat{\theta}_j}$, the difference between these two quantities is very trivial for typical sample sizes in the SEM context, and sample size planning can proceed using $\sigma_{\hat{\theta}_j}$ as the expectation of $s_{\hat{\theta}_j}$.[2]

The quantity h_{jj} in Equation 12 is an element of $\mathbf{H}$, which is equal to $\mathbf{I}(\boldsymbol{\theta})^{-1}$; to calculate $\mathbf{I}(\boldsymbol{\theta})^{-1}$, one can obtain $\mathbf{I}(\boldsymbol{\theta})$ first. If the proposed model is correct, that is, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$, the expected information matrix $\mathbf{I}(\boldsymbol{\theta})$ can be obtained by fitting the model to the population covariance matrix of manifest variables. More specifically, as discussed in the previous section of CI formation methods, $\mathbf{I}(\boldsymbol{\theta})$ is the resulting matrix of evaluating the function $\mathbf{I}(\,)$ at $\boldsymbol{\theta}$, and therefore, to obtain $\mathbf{I}(\boldsymbol{\theta})$, it requires $\mathbf{I}(\,)$ and $\boldsymbol{\theta}$. The function $\mathbf{I}(\,)$ can be obtained by taking the expectation of the second derivative of the log-likelihood function in Equation 1. Analytic derivations to obtain the second derivative of the log-likelihood function are available in Yuan and Hayashi (2006). The values of $\boldsymbol{\theta}$ can be obtained by substituting $\boldsymbol{\Sigma}$ for $\mathbf{S}$ in the log-likelihood function in Equation 1. After $\mathbf{I}(\boldsymbol{\theta})$ is obtained, because $\mathbf{H} = \mathbf{I}(\boldsymbol{\theta})^{-1}$, inverting $\mathbf{I}(\boldsymbol{\theta})$ gives $\mathbf{H}$, which is the asymptotic covariance matrix of the multivariate normal distribution in Equation 3, or equivalently the asymptotic counterpart of $\hat{\mathbf{H}}$ in Equation 8. Extracting the jjth component of $\mathbf{H}$ gives h_{jj}, which is the asymptotic variance of $\hat{\theta}_j$. Although the above steps to calculate h_{jj} from $\boldsymbol{\Sigma}$ might sound technical, this task can be easily performed by almost any mainstream SEM software. If one uses the population covariance matrix of manifest variables to fit the model, the "standard error" for $\hat{\theta}_j$ returned by the software is in fact the population standard deviation $\sqrt{k_{jj}}$. Then one can obtain h_{jj} given the relation between $\mathbf{K}$ and $\mathbf{H}$ discussed above.[3]

Therefore, based on Equation 12, given $\boldsymbol{\Sigma}$, which implies that h_{jj} is implicitly available, the CI width would depend only on the desired confidence level (i.e., $1 - \alpha$) and n. The desired confidence level (e.g., .95) is generally considered a fixed factor, yet n can vary. Correspondingly, solving Equation 12 for n leads to the smallest sample size that satisfies $\mathrm{E}[w] \leq \omega$:

$$n = \mathrm{ceiling}\left\{4(z_{1-\alpha/2})^2\frac{h_{jj}}{\omega^2}\right\}, \tag{13}$$

where ceiling$\{\cdot\}$ refers to the least integer greater than or equal to the value in the braces. Notice that the only quantity needed is h_{jj}, which is a population parameter, otherwise everything else is fixed and known a priori. We momentarily ignore the complication of obtaining reasonable estimates for h_{jj} and devote a section to this issue later in the article. For the moment, let us assume that the information about h_{jj} has been obtained and consider an example

[2] The expectation of sample standard deviation is $\mathrm{E}[s] = \sigma \times c(n)$, where

$$c(n) = \frac{\sqrt{n-1}\,\Gamma((n-1)/2)}{\sqrt{2}\,\Gamma(n/2)}$$

and $\Gamma(\cdot)$ is the gamma function (e.g., Casella & Berger, 2002, p. 364). The quantity $c(n)$ depends only on n and is a decreasing function of n. When $n = 100$, $c(n) = 1.0025$; when $n = 200$, $c(n) = 1.0013$. Thus the difference between $\mathrm{E}[s]$ and σ is negligible for typical sample sizes in the SEM context.

[3] Because mainstream SEM software generally returns only $\hat{\mathbf{K}}$ instead of $\hat{\mathbf{H}}$, and requires the sample size as necessary input when fitting a model, $\hat{\mathbf{H}}$ cannot be directly obtained from the software output. One solution is to give an arbitrary but usually large value (e.g., 100,000) first to n so that the software can return $\hat{\mathbf{K}}$. Then $\hat{\mathbf{H}}$ can be obtained based on $\hat{\mathbf{H}} = n\hat{\mathbf{K}}$.

about how to apply the method presented above to plan for sample size.

Suppose ϕ_{21}, the covariance of two exogenous latent variables in a certain SEM model, is of interest, and the asymptotic variance for $\hat{\phi}_{21}$ (i.e., h_{jj} in Equation 12) is estimated to be 0.20. If the researcher hopes the expected width for a 95% CI will be no wider than 0.10, the necessary sample size is, based on Equation 13, ceiling$\{4(1.96)^2(0.2)/(0.1)^2\}$, which is equal to 308. Thus, solving Equation 13 leads to the necessary sample size such that the expected CI width for ϕ_{21} in an SEM model is sufficiently narrow (i.e., $E[w] \leq \omega$). In summary, this subsection has developed the sample size planning method that ensures that the expected CI width for a model parameter of interest is no wider than desired.

Ensuring a CI Width Is Sufficiently Narrow With a Specified Degree of High Assurance

As Equation 11 indicates, the value of w depends on $\hat{h}_{jj}$, which is in turn based on $\mathbf{I}(\hat{\boldsymbol{\theta}}_n)^{-1}$. Because $\mathbf{I}(\hat{\boldsymbol{\theta}}_n)^{-1}$ is obtained based on $\hat{\boldsymbol{\theta}}_n$, which is a random variable, $\mathbf{I}(\hat{\boldsymbol{\theta}}_n)^{-1}$ is a random matrix. As an element of $\mathbf{I}(\hat{\boldsymbol{\theta}}_n)^{-1}$, $\hat{h}_{jj}$ is itself a random variable. Consequently, the CI width obtained in a study, w, which is based on $\hat{h}_{jj}$, is also a random variable. When the estimate $\hat{h}_{jj}$ obtained is larger than h_{jj}, implying that the standard error is larger than the expected value, w will tend to be larger than ω. Conversely, when $\hat{h}_{jj}$ is smaller than h_{jj}, w will tend to be smaller than ω. Due to maximum likelihood properties, $E[w] \leq \omega$ implies that w is smaller than ω about 50% of the time.[4] Put another way, in the particular study that the researcher is going to conduct, there is approximately a .50 probability that the CI obtained will be narrower than desired, and thus there is about a .50 probability that the CI obtained will be wider than desired.

Although in some situations ensuring the *expected* width is sufficiently narrow is satisfactory, in other situations there is a desire that the width *observed* in a particular study will be sufficiently narrow. To have a specified (high) degree of assurance that w obtained in a particular study will be no wider than ω, sample size needs to be increased. As the sample size increases, the standard error for $\hat{\theta}_j$ becomes smaller and has less variability. Because the relationship between w and ω is affected by the relationship between $\hat{h}_{jj}$ and h_{jj}, we can manipulate the behavior of $\hat{h}_{jj}$ to increase the probability of $w \leq \omega$.

Let γ ($.5 < \gamma < 1$) denote the desired assurance that w is no larger than ω. The current task is to find the sample size such that $P(w \leq \omega) = \gamma$, with n^+ denoting this modified sample size. Because $\hat{\boldsymbol{\theta}}_{n^+}$ follows a multivariate normal distribution asymptotically (i.e., as $n \to \infty$), the covariance matrix of $\hat{\boldsymbol{\theta}}_{n^+}$ follows the Wishart distribution asymptotically (for a technical discussion on the Wishart distribution, see, e.g., Anderson, 2003). At any finite sample size, the sample covariance matrix of $\hat{\boldsymbol{\theta}}_{n^+}$ follows the Wishart distribution approximately:

$$\operatorname{cov}(\hat{\boldsymbol{\theta}}_{n^+}) \approx \text{Wishart}\left(\frac{\mathbf{I}(\boldsymbol{\theta})^{-1}}{n^+ - 1}, n^+ - 1\right). \tag{14}$$

A subset of the sample covariance matrix is still distributed according to the Wishart distribution, with the population covariance matrix being equal to the corresponding subset of the complete covariance matrix and the same degrees of freedom (e.g., Anderson, 2003). Because in the univariate case, the Wishart distribution reduces to a chi-square distribution, the distribution of the sample variance of $\hat{\theta}_j$ at sample size n^+ is approximately a constant times the chi-square with $(n^+ - 1)$ degrees of freedom:

$$\frac{(n^+ - 1)s^2_{n^+}}{\sigma^2_{n^+}} \approx \chi^2_{(n^+-1)}, \tag{15}$$

where $\sigma^2_{n^+} = \operatorname{var}(\hat{\theta}_j)_{n^+} = h_{jj}/n^+$ is the population variance of $\hat{\theta}_j$ based on a sample size of n^+, with $s^2_{n^+}$ being the corresponding sample variance.[5] Given that a function of the sample variance of $\hat{\theta}_j$ follows a chi-square distribution, and that the CI width is an increasing function of the sample standard error, it is possible to manipulate the value of $s^2_{n^+}$, which in turn helps to manipulate the CI width w.

Recall that the task is to find the value of n^+ such that $P(w \leq \omega) = \gamma$. Because the observed CI width, based on a sample of size n^+, is

$$w = 2z_{1-\alpha/2}s_{n^+}, \tag{16}$$

the probability statement $P(w \leq \omega) = \gamma$ becomes

$$P(2z_{1-\alpha/2}s_{n^+} \leq \omega) = \gamma. \tag{17}$$

As Equation 15 indicates, a function of $s^2_{n^+}$ follows the chi-square distribution. Because

$$P\left(\frac{(n^+ - 1)s^2_{n^+}}{\sigma^2_{n^+}} \leq \chi^2_{(n^+-1),\gamma}\right) = \gamma,$$

where $\chi^2_{(n^+-1),\gamma}$ refers to the γth quantile of $\chi^2_{(n^+-1)}$, a quantity can be found that $s^2_{n^+}$ does not exceed $\gamma 100\%$ of the time:

$$P\left(s^2_{n^+} \leq \frac{\sigma^2_{n^+}}{n^+ - 1}\chi^2_{(n^+-1),\gamma}\right) = \gamma. \tag{18}$$

Because the sample variance is always positive, squaring both sides of the inequality within the probability statement in Equation 17 does not change the probability and thus leads to

$$P(4(z_{1-\alpha/2})^2 s^2_{n^+} \leq \omega^2) = \gamma,$$

which can be rearranged into

$$P\left(s^2_{n^+} \leq \frac{\omega^2}{4(z_{1-\alpha/2})^2}\right) = \gamma. \tag{19}$$

Therefore, an identity can be established based on Equations 18 and 19:

[4] This can be seen empirically from the results of our Monte Carlo simulation study in a later section, where the median of the random ws is approximately equal to ω.

[5] More formal notations should include a j subscript in $\sigma^2_{n^+}$ and $s^2_{n^+}$ (e.g., $\sigma^2_{j_{n^+}}$ and $s^2_{j_{n^+}}$). Because the discussion of variances henceforth is in the context of $\hat{\theta}_j$ instead of $\hat{\boldsymbol{\theta}}$, we omit the j subscript so that the presentation is less cumbersome.

$$\frac{\sigma^2_{n^+}}{n^+ - 1}\chi^2_{(n^+-1),\gamma} = \frac{\omega^2}{4(z_{1-\alpha/2})^2}, \quad (20)$$

or equivalently

$$\frac{h_{jj}}{n^+(n^+ - 1)}\chi^2_{(n^+-1),\gamma} = \frac{\omega^2}{4(z_{1-\alpha/2})^2}. \quad (21)$$

The necessary sample size that ensures $P(w \leq \omega) = \gamma$ (i.e., the CI observed in a particular study will be sufficiently narrow with probability γ) can be obtained by solving Equation 21 for n^+:

$$n^+ = \text{ceiling}\left\{\frac{1}{2}\left[1 + \sqrt{1 + \frac{16h_{jj}\chi^2_{(n^+-1),\gamma}(z_{1-\alpha/2})^2}{\omega^2}}\right]\right\}. \quad (22)$$

Similar to Equation 13, Equation 22 indicates that all quantities except n^+ are fixed and known a priori. Because the value of n^+ also plays a role in the chi-square quantile, Equation 22 cannot be directly solved, and it needs to use an iterative procedure. This iterative procedure has been implemented into specialized R functions in the MBESS package, which we discuss in the Appendix.

Let us consider a brief example about how Equation 22 can be used to plan the sample size. Like the previous example, let us momentarily ignore the complication of obtaining reasonable estimates for h_{jj} and assume that the information about h_{jj} has been obtained. Suppose again ϕ_{21}, the covariance of two exogenous latent variables in a certain SEM model, is of interest, and the asymptotic variance for $\hat{\phi}_{21}$ (i.e., h_{jj} in Equation 22) is estimated to be 0.20. If the goal is that the 95% CI obtained in a particular study be no wider than 0.10 with 80% assurance, then the necessary sample size is, based on Equation 22, solved iteratively to be 328.[6] Recall that the necessary sample size was 308 in the previous example when the goal is the expected CI width for ϕ_{21} being no wider than 0.10. Thus, Equation 22 provides a way to plan sample size so that the CI for ϕ_{21} in an SEM model will be sufficiently narrow with 80% assurance, that is, $P(w \leq \omega) = .80$.

Specifying Proper Input Covariance Matrix

As is shown in a later section where we discuss our Monte Carlo simulation study, our sample size planning methods work well when the input parameters are correctly specified. As is the case for any other sample size planning methods, educated estimation of the population effect sizes is required to calculate the necessary sample size, from both the power perspective and the AIPE perspective (e.g., Cohen, 1988; Maxwell et al., 2008). At the end of this section, we show that our procedures are no more difficult than other sample size planning methods in the literature in the context of targeted effects in SEM, because our methods require no additional information as input. In particular, our methods require the following input information: (a) the proposed model, (b) the population covariance matrix of manifest variables, (c) desired CI width (i.e., ω), and (d) the confidence level (i.e., $1 - \alpha$). The first two pieces of information are also required by the power approach for targeted effects in SEM (see, e.g., Hancock, 2006). The extended method requires an additional assurance parameter (i.e., γ). Only the population covariance matrix of manifest variables is unknown a priori; other input parameters are specified according to the goals and the nature of the study, and they are relatively easy to specify. In this section we focus on the specification of the covariance matrix of manifest variables, which is the most difficult part in the sample size planning. Although sometimes it is also difficult to specify the model correctly, this task in practice is typically performed by the researcher based on substantive theories (i.e., not statistical methods) and known before planning the necessary sample size. Therefore, instead of discussing how to specify a model, we study how misspecifications of the model influence the sample size output in a later section.

We regard a systematic literature review as critically important, and it is the most fundamental source for specifying the input for sample size planning procedures. The task of specifying the covariance matrix can, nevertheless, be facilitated by several practical methods, five of which are discussed below.

Method 1

One way to estimate the input covariance matrix directly is based on the relationship between the covariance matrix and the correlation matrix:

$$\mathbf{\Sigma} = \mathbf{VPV}, \quad (23)$$

where $\mathbf{P}$ is the correlation matrix and $\mathbf{V}$ is a diagonal matrix containing the standard deviations of the corresponding variables. If one estimates the correlation coefficient of each pair of the variables first and arranges these coefficients into a matrix, the resulting matrix is $\mathbf{P}$. Generally, observable variables that load on the same latent factor have larger correlations among one another and smaller correlations with other observable variables (i.e., those that load on different latent factors).

The task of specifying the correlation matrix of the manifest variables is analogous to specifying the correlation matrix in the context of sample size planning for multiple regression (for discussions on specifying a correlation matrix for multiple regression sample size planning, see, e.g., S. B. Green, 1991; Kelley & Maxwell, 2003; Maxwell, 2000). A strategy based on the concept of *exchangeable correlation structure* has been proposed to simplify the specification of correlation matrix in the multiple regression context for sample size planning (Maxwell, 2000). An exchangeable correlation structure is one in which the correlations among the predictors are all the same and the correlations of the predictors with the response variable are all the same. Instead of estimating every element in the correlation matrix, the necessary sample size for multiple regression can be approximated based on a correlation matrix of exchangeable structure (i.e., the researcher needs to estimate only one correlation among predictors and one between the response and predictors), and the approximation is

[6] The right side of Equation 22 is

$$\text{ceiling}\left\{\frac{1}{2}\left[1 + \sqrt{1 + \frac{16(0.2)\chi^2_{(328-1),.80}(1.96)^2}{(0.1)^2}}\right]\right\}.$$

Notice that 328 is used in $\chi^2_{(328\ -\ 1),\ .80}$ to obtain the quantile. Finishing the calculations will give ceiling{327.6764}, which is 328 and equal to the left side of Equation 22 (i.e., n^+).

usually satisfactory (B. F. Green, 1977; Maxwell, 2000; Raju et al., 1999; Wainer, 1976). Reviewing the patterns of correlations in the behavioral data, B. F. Green (1977) concluded that, in the context of multiple regression, "many linear composites are barely different from using equal weights" (p. 274). In the SEM context, Rigdon (1998) suggested that an equal correlation model is best for the baseline model to evaluate goodness of fit because it "reflects the reality of a theoretical background correlation in non-experimental data sets" (p. 63). This phenomenon is also related to the *coefficient H* (Hancock & Mueller, 2001) on the measurement level of an SEM model, which is discussed below as another practical method of specifying input parameters.

To better understand the concept of exchangeable correlation structure, consider a simple example of a one-factor confirmatory factor analysis (CFA) model. Figure 1 indicates the path diagram and the standardized and unstandardized (in parentheses) population model parameters.[7] Table 1 indicates the population correlation matrix (below the diagonal) and a correlation matrix of exchangeable structure (above the diagonal) of the manifest variables, with corresponding variances and covariances in parentheses. The off-diagonal values are set to 0.3 in the exchangeable correlation matrix because this value is around the median of the original correlation coefficients.[8] Further suppose the desired (unstandardized) CI width is 0.3. Based on the population covariance matrix, the necessary sample size is 349, 540, 814, or 1,083, if the model parameter of interest is λ_1, λ_2, λ_3, or λ_4, respectively, calculated with the sample size planning methods developed in the previous section.[9] Note that if the model parameter of interest is different, the necessary sample size is usually different, everything else being the same, as is the case in this example. If more than one model parameter is of interest to the researcher, the largest sample size should be used for the study. With the covariance matrix calculated from the exchangeable correlation matrix, those sample sizes are 377, 589, 847, and 1,046 for λ_1, λ_2, λ_3, and λ_4, respectively, and they are thus quite similar to the "true" sample sizes.

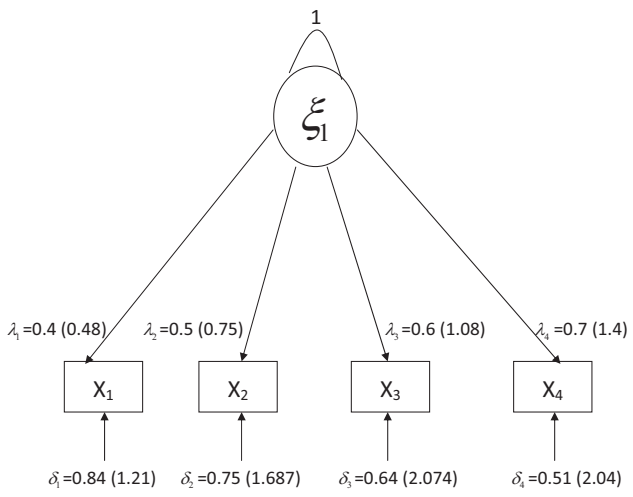

Figure 1. Path diagram for a one-factor confirmatory factor analysis model. Numbers shown are population standardized model parameters, with unstandardized values in parentheses. The variance of the latent factor is fixed at 1.

Table 1
The Population Correlation Matrix (Below the Diagonal) and a Correlation Matrix of Exchangeable Structure (Above the Diagonal) for the Model in Figure 1, With Corresponding Variances and Covariances in Parentheses

Variable	X_1	X_2	X_3	X_4
X_1	1 (1.2^2)	.30 (0.54)	.30 (0.648)	.30 (0.72)
X_2	.20 (0.36)	1 (1.5^2)	.30 (0.81)	.30 (0.90)
X_3	.24 (0.5184)	.30 (0.81)	1 (1.8^2)	.30 (1.08)
X_4	.28 (0.672)	.35 (1.05)	.42 (1.512)	1 (2.0^2)

Therefore, this example suggests that exchangeable correlation structure can be helpful for specifying the input covariance matrix for sample size planning, given a reasonable choice of the correlation values.

Conventions about correlation effect sizes are also useful resources. For example, Cohen (1988) defined "small," "medium," and "large" effects in various designs and analyses in the context of behavioral science studies, and his suggestions are supported in part by reviews of published studies in psychological journals (e.g., Sedlmeier & Gigerenzer, 1989). In the context of the correlation coefficient, Cohen's suggestions are 0.1, 0.3, and 0.5 for small, medium, and large effects, respectively.

Method 2

One practical way to estimate the standard deviation is to use the interquartile range (IQR), which is defined as the 75th percentile minus the 25th percentile. For normally distributed data, IQR is approximately equal to 1.349 standard deviation units, because the 75th and 25th quantiles of the standard normal distribution are 0.674 and −0.674, respectively.

The standard deviation can also be estimated based on the (full) range of the target population. Based on substantive theories, if the researcher is able to infer the maximum and minimum theoretically plausible values of a variable, the standard deviation is approximately equal to the range divided by 6, assuming that data are normally distributed. For example, many times a latent construct is measured by professional standardized tests, where the maximum and minimum scores of the test are known and test scores are approximately normally distributed. This method differs from that based on IQR in that the former requires stronger substantive theories; the maximum and/or minimum in prior data sets in the literature may not approximate the extreme values in the population, because there can be outliers (i.e., exceed the theoretical bounds due to errors in recording or measuring) or come from a restricted sample (i.e., fail to reach the theoretical limits). On the other hand, the IQR method is more robust to outliers, but it may

[7] Note that our sample size planning procedures do not require the values of population model parameters as input. We include those values in the path diagram for better illustration and completeness of information.

[8] The actual median is 0.29.

[9] These values were calculated with the R function ss.aipe.sem.path() in the MBESS package that implements the sample size planning methods presented previously. See the Appendix for detailed information.

rely more heavily on the existence of enough studies that are similar to the study in question.

Method 3

In addition to specifying the input covariance matrix directly, one can start from the model parameters (i.e., structural coefficients, path loadings, etc.). Based on the model input, if one specifies the plausible values of the model parameters (i.e., values in $\boldsymbol{\theta}$), the model-implied covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ can be obtained and can be used as the input $\boldsymbol{\Sigma}$, the covariance matrix of manifest variables. Actually this method is commonly used in power analysis on the likelihood ratio tests of both the model overall fit and model targeted effects (i.e., the method to estimate the chi-square noncentrality parameter developed by Satorra & Saris, 1985; see also Saris & Satorra, 1993). Hancock (2006) provided a detailed empirical example about how to specify proper values for model parameters, so as to obtain the population covariance matrix of manifest variables and calculate power for SEM.

Obtaining $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ from $\boldsymbol{\theta}$ can be accomplished directly with the specialized R function theta.2.Sigma.theta() in the MBESS package or indirectly with mainstream SEM software (for an empirical demonstration, see, e.g., Davey & Savla, 2010, Chapter 4).[10] Generally, it is easier to specify model parameter values in the standardized metric context, because it is more convenient to hypothesize the magnitude of the relationships in the model in terms of standardized coefficients. In particular, the variance of an exogenous variable is always 1, and all sources of variances of an endogenous variable should sum to 1. Therefore, error variances can be readily obtained after specifying path coefficients. For example, suppose the measurement equation of a manifest variable Y_1 is $Y_1 = \lambda_1\xi_1 + \delta_1$, where λ_1 is the path loading, ξ_1 is an exogenous latent variable, and δ_1 is the error. In the context of standardized metric, suppose, based on prior evidence, the researcher estimates the magnitude of λ_1 to be 0.6, then the error variance for Y_1 is 0.64 (i.e., $1 - \lambda_1^2$). Suppose another manifest variable Y_2 has measurement equation $Y_2 = \lambda_2\xi_1 + \lambda_3\xi_2 + \delta_2$, where λ_2 and λ_3 are path loadings, ξ_2 is another exogenous latent variable, δ_2 is the error, and the covariance between ξ_1 and ξ_2 is ϕ_{21}. In the standardized metric context, if the researcher estimates the magnitude of λ_2, λ_3, and ϕ_{21} to be 0.2, 0.6, and -0.3, respectively, then the error variance of Y_2 will be 0.672 (i.e., $1 - \lambda_2^2 - \lambda_3^2 - 2\lambda_2\lambda_3\phi_{21}$). Note that standardized model parameters imply a correlation matrix rather than a covariance matrix of the manifest variables. After obtaining a correlation matrix, the researcher can then apply Methods 1 and 2 discussed above to finally obtain a covariance matrix.

Method 4

On the measurement level, the reliability coefficient is helpful in estimating path loadings. In practice, many times the latent variables are measured with fully developed tests or scales, and the reliability of those tests are known. Therefore, specifying path loadings can be largely facilitated by the fact that, in the standardized coefficient context, the path loading is the square root of the reliability. This is the case because reliability can be defined as the variance due to the construct divided by the indicator's total variance, which is 1 in the standardized coefficient context. For example, suppose the measurement equation of a manifest variable Y_3 is $Y_3 = \lambda_4\xi_3 + \delta_3$, using the same notation as previously defined. Further suppose Y_3 is a certain standardized test and the reliability of this test is known to be 0.9. Because the reliability of this test can be defined as $\text{var}(\lambda_4\xi_3)/\text{var}(\lambda_4\xi_3 + \delta_3)$ (e.g., McDonald, 1999, Chapter 7), which is simplified as λ_4^2 in the standardized metric context, it easily follows that λ_4 is approximately 0.949 (i.e., $\sqrt{0.9}$). A detailed discussion of measurement reliability from the factor analytic perspective is available in McDonald (1999). In fact, such an interpretation of reliability is also useful if the researcher specifies model parameters in the unstandardized coefficient context, in that the error variance can still be obtained from the knowledge of the reliability and (raw scale) path loading.

Method 5

When the interest lies in the structural coefficients, which is often the case in studies in the behavioral and social sciences, specification of measurement models can be further simplified with an index called *coefficient H*, proposed by Hancock and Mueller (2001). This index measures the overall quality of a factor's measurement model; for a given factor with M indicators, it can be defined as

$$H = \frac{\sum_{m=1}^{M} \lambda_m^2/(1 - \lambda_m^2)}{1 + \sum_{m=1}^{M} \lambda_m^2/(1 - \lambda_m^2)}, \tag{24}$$

where λ_m is the *standardized* path loading of the mth ($m = 1, \ldots, M$) indicator. Based on Equation 24, H is (a) bounded between 0 and 1; (b) unaffected by loadings' signs; (c) a nondecreasing function of M; and (d) larger than or equal to $\max(\lambda_m^2)$, the reliability of the strongest indicator (Hancock & Mueller, 2001). This coefficient can be interpreted as the proportion of variance in the construct that is explained by its indicators, or the construct reliability associated with each factor's measurement model. Interpretations of this index are various (Fornell & Larcker, 1981; Hancock & Mueller, 2001), such as the *maximal reliability* in the context of scale construction (Raykov, 2004). Hancock and Kroopnick (2005) demonstrated that the specific value of each path loading does not affect the overall performance of the SEM model in terms of the noncentral chi-square distribution of the model chi-square statistic, as long as different combinations of path loadings give the same H value. Given the discussion on the coefficient H and exchangeable correlation structure, therefore, if the researcher can anticipate the overall construct reliability (or its achievable minimum) of a factor's measurement model, instead of specifying the path loading for each indicator, the task can be achieved by using either M indicators whose loadings are all equal to $\sqrt{H/(H + M - MH)}$ or one indicator with loading $\sqrt{H}$ as a

[10] See the Appendix for detailed information about the R function theta.2.Sigma.theta().

placeholder, although in some situations the latter may imply identification issues.[11] For a detailed empirical example of using the coefficient H to specify model parameters in the power analysis context, see Hancock (2006, pp. 90–93).

Discussion of Parameter Specification When Planning Sample Size

In summary, we have suggested five practical methods that are helpful for specifying the input covariance matrix, as a supplement to the literature review, and these methods generally are compatible with one another. The researcher can use these methods in a mix-and-match manner so that different methods are employed ultimately to specify different parts of the input covariance matrix, or can cross-validate the input by comparing the covariance matrices specified with different methods.

Any sample size planning method, from either the power-analytic or AIPE perspective, requires educated estimations of population effect sizes, and in this regard our methods require no additional information and are no more difficult than others. Actually, power analysis for targeted effects in SEM requires the knowledge of all model parameters. If a path coefficient in SEM is of interest, one can conduct hypothesis testing on the likelihood ratio of the nested models (i.e., the model with the path estimated and the model with the path fixed at zero), and infer if the null hypothesis that the path coefficient is zero can be rejected at a certain significance level. Accordingly, sample size planning can be performed so that the likelihood ratio test of the nested models has the desired power. The likelihood ratio follows a noncentral chi-square distribution, and in order to estimate the power of the likelihood ratio test, all parameters in the full model and all parameters in the restricted model are required to calculate the noncentrality parameter of the distribution (Saris & Satorra, 1993; Satorra & Saris, 1985). Therefore, our sample size planning methods are no more difficult to implement compared with the commonly used NHST approach to sample size planning in the literature, because they require a similar amount of input information.

It is also common for any sample size planning procedure that the output (i.e., the estimated necessary sample size) depend on the quality of the input parameters. For example, in the context of power analysis and sample size planning for SEM, MacCallum, Browne, and Cai (2006) acknowledged that the calculated power and sample size should be viewed as approximations, conditional on the input information and the validity of statistical assumptions (p. 34). It is the normal state of affairs that sample size planning procedures depend on quantities that are not yet fully known (i.e., population effect sizes), but those procedures can be highly useful in planning and evaluating research design (MacCallum et al., 2006, p. 34). The same reasoning applies equally to the sample size planning methods in the present article as well.

Specifying the Desired Width

The sample size planning methods presented above require the value of the desired CI width in the unstandardized context. Although this value is necessarily context specific and based on the goals of the study, there is a practical method to facilitate choosing appropriate values for the width. Because it is usually easier to conceptualize the desired CI width in the context of standardized model coefficients, if the relationship between the unstandardized and standardized CI widths can be found, the researcher can first specify the standardized width and transform it back to the unstandardized scale.

An intuitive but incorrect solution to the present issue is to fit the model with a correlation matrix: If one puts a correlation matrix into the covariance structure analysis procedure, the resulting model parameter estimates are standardized values, and the CI width would thus also be in a standardized metric. Thus it is tempting to replace the covariance matrix in the sample size planning methods with a correlation matrix so that one can specify the desired CI width in terms of standardized values. However, although analyzing a correlation matrix as if it were a covariance matrix generally provides the same point estimates of model parameters, the standard errors and discrepancy function statistic are incorrect (Cudeck, 1989; Lee, 1985; Shapiro & Browne, 1990). Therefore, the CIs formed based on a correlation matrix are incorrect. A correct approach to specifying the standardized CI width must be based on a covariance matrix.

To find the relationship between unstandardized and standardized CI widths, one first needs to understand the relationship between unstandardized and standardized model parameters. In particular, a certain standardized path coefficient from a variable V to a variable U can be defined as (e.g., Bollen, 1989, p. 349)

$$b_{UV}^{(s)} = b_{UV}\frac{\sigma_V}{\sigma_U} \equiv g(b_{UV}), \tag{25}$$

where $b_{UV}^{(s)}$ and b_{UV} are the population standardized and unstandardized path coefficients, respectively; σ_U and σ_V are the population standard deviations of variables U and V, respectively; and $g(b_{UV})$ refers to a function of b_{UV}. If U or V is a manifest variable, σ_U or σ_V is directly available from the population covariance matrix of manifest variables; if U or V is a latent variable, its standard deviation can be calculated based on population model parameters. In accordance with the previous discussion on the asymptotic distribution of model parameter estimates, the MLE of b_{UV} is distributed asymptotically as

$$\sqrt{n}(\hat{b}_{UV} - b_{UV}) \xrightarrow{L} N(0, \sigma_{b_{UV}}^2), \tag{26}$$

where $\sigma_{b_{UV}}^2$ is the corresponding element in the inverse expected information matrix of model parameters as defined in Equation 3. Under the delta method (e.g., Pawitan, 2001, p. 89), the asymptotic distribution of a function of $\hat{b}_{UV}$ is

$$\sqrt{n}[\, g(\hat{b}_{UV}) - g(b_{UV})] \xrightarrow{L} N(0, \sigma_{b_{UV}}^2[g'(b_{UV})]^2), \tag{27}$$

where $g'(b_{UV}) = \sigma_V/\sigma_U$ is the first derivative of $g(\,)$ evaluated at b_{UV}. Equation 27 can be rewritten as

[11] Because all path loadings are equal, Equation 24 reduces to

$$H = \frac{M[\lambda^2/(1-\lambda^2)]}{1 + M[\lambda^2/(1-\lambda^2)]},$$

where λ is the path loading. Solving this equation for λ gives

$$\lambda = \sqrt{H/(H + M - MH)}.$$

$$\sqrt{n}(\hat{b}_{UV}^{(s)} - b_{UV}^{(s)}) \xrightarrow{L} N\left(0, \sigma_{b_{UV}}^2 \frac{\sigma_V^2}{\sigma_U^2}\right), \quad (28)$$

so that the asymptotic distribution of $\hat{b}_{UV}^{(s)}$ is derived. Therefore, the expected CI width for $b_{UV}^{(s)}$ at sample size n is

$$\omega^{(s)} = 2z_{1-\alpha/2}\frac{\sigma_{b_{UV}}(\sigma_V/\sigma_U)}{\sqrt{n}}. \quad (29)$$

Based on Equation 26, the expected CI width for b_{UV} at sample size n is

$$\omega = 2z_{1-\alpha/2}\frac{\sigma_{b_{UV}}}{\sqrt{n}}. \quad (30)$$

Therefore, based on Equations 29 and 30, at the same sample size, the relationship between ω and $\omega^{(s)}$ can be established:

$$\omega = \omega^{(s)}\frac{\sigma_U}{\sigma_V}. \quad (31)$$

Note that although Equation 29 provides a method to calculate the standardized CI width, itself alone cannot be used to plan for sample size, because $\sigma_{b_{UV}}$, σ_V, and σ_U are in an unstandardized metric and need to be necessarily obtained from a covariance matrix.

Given the above discussion of standardized and unstandardized CI widths, the task of specifying desired CI width can be facilitated with Equation 31. When it is easier to conceptualize the desired CI width in terms of a standardized metric, the researcher can first determine a reasonable value for $\omega^{(s)}$ and then calculate ω based on Equation 31. Because using a correlation matrix to fit a model generally leads to incorrect solutions, the sample size planning methods in the present article are framed in the context of unstandardized model parameters and always require a covariance matrix as input.

Monte Carlo Simulation Study

Although the distribution of $\hat{\boldsymbol{\theta}}_n$ converges to multivariate normal as n approaches infinity, at any finite sample size the distribution of $\hat{\boldsymbol{\theta}}_n$ is only approximately multivariate normal, which implies that the sample variance of $\hat{\theta}_j$ does not exactly follow a chi-square distribution. Therefore, it is necessary to ensure the effectiveness of the sample size planning procedures in realistic situations where the methods will be applied, especially when the procedure-implied sample size is relatively small. If the SEM standard error estimation theories are exact, the necessary sample size returned by the sample size planning methods will be exact. Because our sample size planning procedures are necessarily dependent on the MLE, they cannot be expected to work well when maximum likelihood itself does not work well. Extensive Monte Carlo simulations were conducted to evaluate the performance of the procedures we developed.

The simulation study was conducted in the context of three models representative of applied research where SEM is often used: (a) a CFA model (Model 1) based on Holzinger and Swineford (1939); (b) an autoregressive model (Model 2) based on Curran, Bollen, Chen, Paxton, and Kirby (2003) and Paxton, Curran, Bollen, Kirby, and Chen (2001); and (c) a complex SEM model (Model 3) with mediation among latent variables, based on Maruyama and McGarvey (1980). Some modifications of the original model and/or parameters were made so that these models are more generally applicable. Path diagrams and model parameters used in the Monte Carlo simulation are provided in Figures 2–4.

Given the models, we first specified the unstandardized model parameters, whose values are indicated in Figures 2–4. After specifying the model and model parameters, the model-implied population covariance matrix, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, can be obtained, and it is used as the population covariance matrix of the manifest variables, that is, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$. The specifications of the present simulation study should not be confused with the input of our sample size planning methods, which do not include the model parameters. The specified model parameters of interest include factor loadings (e.g., λ_1 in Model 1, λ_6 in Model 2), structural path coefficients (e.g., β_{21} in Model 2, λ_{11} in Model 3), and covariances between latent variables (e.g., ϕ_{21} in Model 1). The targeted effects are meant to represent a variety and magnitude of possible effects in an SEM context. We planned the sample size with regard to the expected CI width (i.e., with the goal that $E[w] \leq \omega$), as well as with the goal to achieve assurance of .80 (i.e., the extended method with γ set to .80). The value of .80 is reasonable for our simulation study because (a) the CI observed in a study is ensured to be sufficiently narrow with relatively high assurance; (b) not all the random widths are narrower than desired, so that the sample size returned by the planning procedure can be ascertained to be the minimum necessary. This is the case because if γ is set to a very large value (e.g., .99) and all the random widths are narrower than desired, it is unknown whether it is because the planning procedure gave a correct sample size, or because a larger than necessary (i.e., not correct) sample size was used. Given (a) the particular model, (b) the population covariance matrix $\boldsymbol{\Sigma}$, (c) the model parameter of interest θ_j, and (d) the desired width ω, the necessary sample size N is returned from the sample size planning procedures ($N = n$ as defined in Equation 13, or $N = n^+$ as defined in Equation 22). A random sample of size N was generated from a multivariate normal distribution with the MASS package (Venables & Ripley, 2002) in R with the population covariance being equal to $\boldsymbol{\Sigma}$.[12] Then we fit the model with the sample covariance matrix using maximum likelihood and calculated the standard error of $\hat{\theta}_j$ using the sem package (Fox, 2006) in R. A 95% CI for θ_j was formed. Each condition was replicated 5,000 times. There is no guideline regarding the appropriate number of replications, and currently the typical number in the SEM context is 1,000. Generally, the larger the number is, the more reliable the simulation results are, other things being the same. We used 5,000 replications so that the results are more reliable than if each condition was repeated 1,000 times.

Results

Simulation results are reported in detail in Tables 2–11, where (a) the mean (M_w) and median (Mdn_w) of the 5,000 random CI widths are given; (b) P_{80}, P_{75}, and P_{70} refer to the subscripted percentiles of the 5,000 random widths; (c) γ_{emp} is the empirical

[12] MASS was originally an acronym for Modern Applied Statistics with S, but is used as a stand-alone package title now.

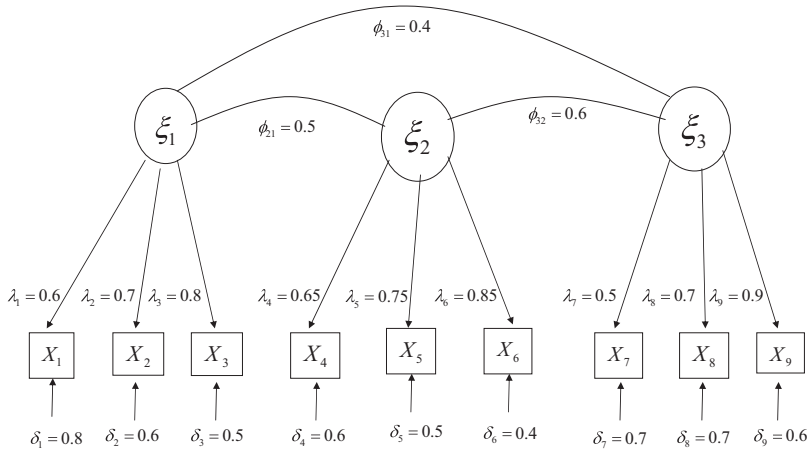

Figure 2. Path diagram for the population Model 1 in our Monte Carlo simulation study. This is a confirmatory factor analysis model, where each of the three correlated latent factors is measured by three manifest variables. The variances for the latent factors are fixed at 1. Numbers shown are unstandardized population model parameters.

assurance, which refers to the percentage of the random widths no larger than the desired value; and (d) α_{emp} refers to the empirical Type I error rate. Each table consists of two parts; the upper half (indicated by $\gamma = E[w]$) presents the performance of the standard sample size planning method with the goal $E[w] \leq \omega$, and the lower half (indicated by $\gamma = .80$) presents the performance of the extended method (i.e., the one with the assurance parameter).

If our extended sample size planning method is satisfactory, γ_{emp} should be close to the specified assurance, which is .80 in the present context. That is, $\gamma_{emp} \geq .80$ implies $P_{80} \leq \omega$, and $\gamma_{emp} \leq .80$ implies $P_{80} \geq \omega$. In cases where γ_{emp} falls below .80, we are interested in studying whether the difference between P_{80} and ω is of any substantive meaning. One way to achieve this task is to evaluate the difference between the percentiles and the desired value directly, with indices P_{80}/ω and P_{70}/ω. A second way is to calculate $(P_{80} - \omega)/\theta_j$ and $(P_{70} - \omega)/\theta_j$, and evaluate whether the excess of P_{80} (or P_{70}) on ω is negligible compared with the magnitude of θ_j, the population effect size of interest. For example, in Table 2 the parameter of interest is γ_1, and therefore θ_j takes the value of 0.60 (see Figure 2) in calculating those two indices in Table 2. To help evaluate the performance of the extended sample size planning method, we have also included the above four indices in the tables.

The Method Based on Expected CI Width

The CFA model. The standard method is effective in the context of Model 1, and it produced sample sizes so that the means of the random CI widths are equal to the desired values, even for relatively small sample sizes. For example, Table 2 shows that $M_w = 0.402$ and $Mdn_w = 0.401$ when $\omega = 0.40$ ($N = 133$) and λ_1 is of interest. For another, $M_w = 0.2996$ and $Mdn_w = 0.2997$ when $\omega = 0.30$ ($N = 210$) and ϕ_{21} is of interest (see Table 4). Except for the three cases of the smallest sample sizes (i.e., $N = 133$ in Table

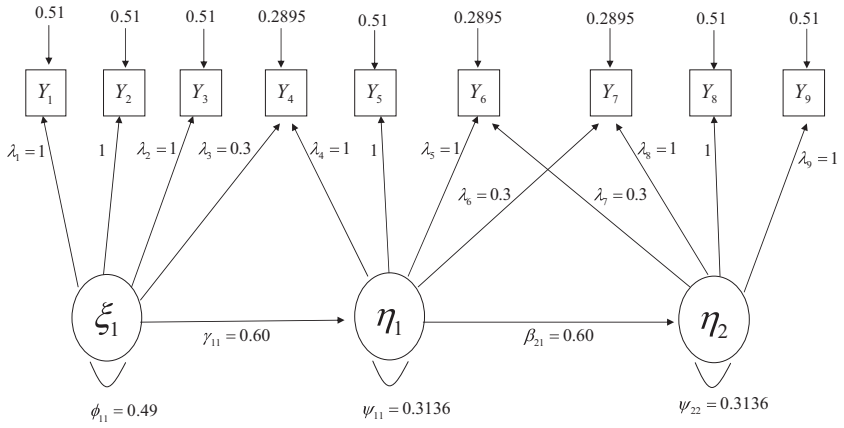

Figure 3. Path diagram for the population Model 2 in our Monte Carlo simulation study. This structural equation model contains nine manifest variables, one exogenous latent variable, and two endogenous latent variables. Numbers shown are unstandardized population model parameters.

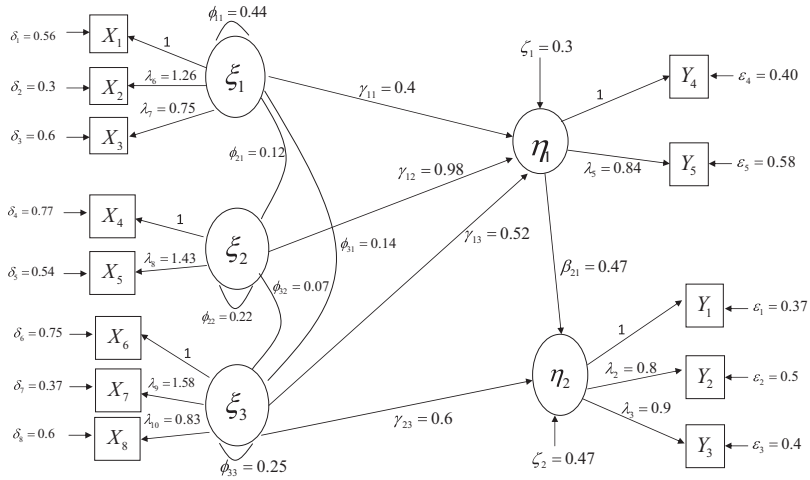

Figure 4. Path diagram for the population Model 3 in our Monte Carlo simulation study. This complex structural equation model contains three correlated exogenous latent variables and two endogenous latent variables, with a mediation effect among the latent variables. Numbers shown are unstandardized population model parameters.

2, $N = 131$ in Table 3, and $N = 154$ in Table 4), in all cases M_w and Mdn_w are smaller than or equal to ω to the first three decimals (see Tables 2–4).

SEM models. The standard method is also effective in the context of SEM models; similar patterns to those in the CFA context have been observed in Tables 5–11. When sample sizes are small, M_w and Mdn_w are close to ω or exceed it by only a small amount. For example, CIs for β_{21} in Model 2 have $M_w = 0.46$ and $Mdn_w = 0.45$ when $\omega = 0.45$ ($N = 179$; see Table 5). The CIs for γ_{11} in Model 2 have $M_w = 0.407$ and $Mdn_w = 0.3996$ when $\omega = 0.40$ ($N = 222$; see Table 6). The worst case is in Table 11 when $M_w = 0.736$ and $Mdn_w = 0.7109$ while $\omega = 0.70$ ($N = 283$). However, as the sample size becomes larger, the discrepancy between M_w (or Mdn_w) and ω decreases quickly. Consider Table 11 again: When the sample size is 554, $M_w = 0.5099$ and $Mdn_w = 0.5017$, very close to the desired value $\omega = 0.50$. Because the CI formation is based on asymptotics, at any finite sample size the model parameters follow a normal distribution only approximately. Instead of the sample size planning method itself failing at smaller sample sizes, the reason why $M_w > \omega$ in some cases of small sample sizes is that some model parameters converge to the normal distribution less

Table 2
Empirical Distributions of Confidence Interval Widths for λ_1 in Model 1 ($\theta_j = \lambda_1 = 0.6$)

				$\gamma = E[w]$				
ω	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
N	8,487	2,122	943	531	340	236	174	133
Mw	0.04997	0.09997	0.15003	0.20007	0.25010	0.30054	0.35071	0.40210
Mdnw	0.04997	0.09999	0.14996	0.20001	0.24999	0.30021	0.34993	0.40085
				$\gamma = .80$				
ω	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
N	8,596	2,176	979	558	362	254	189	146
M_w	0.04966	0.09873	0.14723	0.19506	0.24249	0.28996	0.33642	0.38309
Mdn_w	0.04965	0.09871	0.14722	0.19487	0.24208	0.28923	0.33581	0.38199
P_{80}	0.05001	0.10002	0.15018	0.20023	0.25047	0.30138	0.35183	0.40385
P_{75}	0.04994	0.09977	0.14962	0.19908	0.24872	0.29910	0.34850	0.39867
P_{70}	0.04988	0.09952	0.14906	0.19824	0.24722	0.29691	0.34546	0.39448
P_{80}/ω	1.00015	1.00020	1.00121	1.00117	1.00187	1.00461	1.00524	1.00962
P_{70}/ω	0.99758	0.99517	0.99373	0.99121	0.98886	0.98972	0.98702	0.98621
$(P_{80} - \omega)/\theta_j$	0.00001	0.00003	0.00030	0.00039	0.00078	0.00231	0.00305	0.00641
$(P_{70} - \omega)/\theta_j$	−0.00020	−0.00080	−0.00157	−0.00293	−0.00464	−0.00514	−0.00757	−0.00919
γ_{emp}	0.7936	0.7942	0.7846	0.7904	0.7860	0.7728	0.7704	0.7620
α_{emp}	0.0484	0.0454	0.0486	0.0494	0.0496	0.0488	0.0526	0.0534

Note. M_w and Mdn_w are the mean and median of the 5,000 random confidence interval widths. P_{80}, P_{75}, and P_{70} refer to the respective percentiles of the 5,000 random widths. γ_{emp} is the percentage of the random widths no larger than the desired value. α_{emp} is the empirical Type I error rate.

Table 3
Empirical Distributions of Confidence Interval Widths for λ_6 in Model 1 ($\theta_j = \lambda_6 = 0.85$)

			$\gamma = E[w]$			
ω	0.10	0.15	0.20	0.25	0.30	0.35
N	1,599	711	400	256	178	131
M_w	0.10006	0.14998	0.20022	0.25056	0.30063	0.35132
Mdn_w	0.10003	0.14997	0.20010	0.25021	0.30027	0.35046
			$\gamma = .80$			
ω	0.10	0.15	0.20	0.25	0.30	0.35
N	1,646	742	423	275	194	144
M_w	0.09859	0.14687	0.19475	0.24162	0.28789	0.33505
Mdn_w	0.09859	0.14675	0.19472	0.24141	0.28760	0.33454
P_{80}	0.09994	0.14985	0.20000	0.24960	0.29940	0.35013
P_{75}	0.09968	0.14921	0.19895	0.24797	0.29702	0.34727
P_{70}	0.09943	0.14866	0.19802	0.24647	0.29493	0.34445
P_{80}/ω	0.99943	0.99903	0.99998	0.99842	0.99800	1.00036
P_{70}/ω	0.99433	0.99105	0.99009	0.98589	0.98311	0.98414
$(P_{80} - \omega)/\theta_j$	−0.00006	−0.00017	0.00000	−0.00047	−0.00071	0.00015
$(P_{70} - \omega)/\theta_j$	−0.00067	0.05724	0.11531	0.17232	0.22933	0.28759
γ_{emp}	0.8116	0.8080	0.8000	0.8124	0.8128	0.7984
α_{emp}	0.0476	0.0476	0.0530	0.0520	0.0534	0.0546

Note. M_w and Mdn_w are the mean and median of the 5,000 random confidence interval widths. P_{80}, P_{75}, and P_{70} refer to the respective percentiles of the 5,000 random widths. γ_{emp} is the percentage of the random widths no larger than the desired value. α_{emp} is the empirical Type I error rate.

quickly than others. Nevertheless, at large sample sizes, M_w and Mdn_w are very close to ω in all cases, and the procedure for planning sample size thus works well.

The Extended Method

The CFA model. The extended method is effective; it tends to ensure the proportion of $w \leq \omega$ equal to the specified assurance value .80. Tables 2–3 show that all γ_{emp} values are very close to .80 (typically .80 ± .02), indicating that if λ_1 or λ_6 in Model 1 is of interest, the sample size planning procedure will ensure that the CI obtained in a particular study is sufficiently narrow with 80% assurance. For example, Table 3 shows that even when N is as small as 194 ($\theta_j = \lambda_6$, $\omega = 0.30$), γ_{emp} is already as high as .8128.

Although γ_{emp} values in Table 4 are typically only around .70 rather than .80, in each case the 80th percentile of those random

Table 4
Empirical Distributions of Confidence Interval Widths for ϕ_{21} in Model 1 ($\theta_j = \phi_{21} = 0.5$)

			$\gamma = E[w]$			
ω	0.10	0.15	0.20	0.25	0.30	0.35
N	1,882	837	471	302	210	154
M_w	0.10006	0.15011	0.20021	0.25019	0.29964	0.35151
Mdn_w	0.10005	0.15013	0.20033	0.25059	0.29974	0.35184
			$\gamma = .80$			
ω	0.10	0.15	0.20	0.25	0.30	0.35
N	1,933	871	496	322	227	168
M_w	0.09879	0.14719	0.19512	0.24225	0.28804	0.33554
Mdn_w	0.09875	0.14731	0.19510	0.24237	0.28754	0.33517
P_{80}	0.10094	0.15199	0.20341	0.25499	0.30668	0.36102
P_{75}	0.10049	0.15098	0.20177	0.25239	0.30285	0.35592
P_{70}	0.10010	0.15018	0.20026	0.25009	0.29950	0.35097
P_{80}/ω	1.00938	1.01324	1.01704	1.01996	1.02226	1.03149
P_{70}/ω	1.00102	1.00117	1.00128	1.00038	0.99832	1.00278
$(P_{80} - \omega)/\theta_j$	0.00188	0.00397	0.00682	0.00998	0.01336	0.02204
$(P_{70} - \omega)/\theta_j$	0.00020	0.00035	0.00051	0.00019	−0.00101	0.00195
γ_{emp}	0.6874	0.6902	0.6890	0.6978	0.7072	0.6906
α_{emp}	0.0502	0.0540	0.0510	0.0534	0.0570	0.0592

Note. M_w and Mdn_w are the mean and median of the 5,000 random confidence interval widths. P_{80}, P_{75}, and P_{70} refer to the respective percentiles of the 5,000 random widths. γ_{emp} is the percentage of the random widths no larger than the desired value. α_{emp} is the empirical Type I error rate.

Table 5
Empirical Distributions of Confidence Interval Widths for β_{21} in Model 2 ($\theta_j = \beta_{21} = 0.6$)

				$\gamma = E[w]$				
ω	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
N	3,608	1,604	902	578	401	295	226	179
M_w	0.10014	0.15060	0.20112	0.25124	0.30315	0.35355	0.40761	0.46027
Mdn_w	0.10006	0.15023	0.20016	0.24942	0.30068	0.34853	0.40085	0.45023
				$\gamma = .80$				
ω	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
N	3,679	1,651	937	606	425	315	244	195
M_w	0.09918	0.14833	0.19720	0.24577	0.29452	0.34218	0.39163	0.43905
Mdn_w	0.09909	0.14796	0.19647	0.24444	0.29169	0.33836	0.38431	0.43079
P_{80}	0.10196	0.15445	0.20781	0.26236	0.31790	0.37430	0.43329	0.48951
P_{75}	0.10139	0.15324	0.20544	0.25880	0.31209	0.36668	0.42361	0.47802
P_{70}	0.10088	0.15201	0.20351	0.25553	0.30770	0.35959	0.41300	0.46670
P_{80}/ω	1.01959	1.02968	1.03905	1.04944	1.05965	1.06943	1.08322	1.08781
P_{70}/ω	1.00875	1.01337	1.01757	1.02213	1.02567	1.02741	1.03250	1.03712
$(P_{80} - \omega)/\theta_j$	0.00326	0.00742	0.01302	0.02060	0.02983	0.04050	0.05548	0.06586
$(P_{70} - \omega)/\theta_j$	0.00146	0.00334	0.00586	0.00922	0.01284	0.01599	0.02167	0.02784
γ_{emp}	0.6098	0.6072	0.6002	0.6036	0.611522	0.616263	0.616466	0.617202
α_{emp}	0.0446	0.0490	0.0506	0.0500	0.048810	0.052273	0.050201	0.060569

Note. M_w and Mdn_w are the mean and median of the 5,000 random confidence interval widths. P_{80}, P_{75}, and P_{70} refer to the respective percentiles of the 5,000 random widths. γ_{emp} is the percentage of the random widths no larger than the desired value. α_{emp} is the empirical Type I error rate.

widths exceed ω by only a trivial amount. For example, in the worst case in Table 4 where N is as small as 168 and $\omega = 0.35$, $P_{80} = 0.361$, exceeding ω by only .011 or 3% ($P_{80}/\omega = 1.03$). As the sample size increases, the discrepancy between P_{80} and ω decreases quickly: When $N = 496$ and $\omega = 0.20$, $P_{80} = 0.2034$, exceeding ω by only .0034 (see Table 4). Therefore, such a small discrepancy between P_{80} and ω does not have any substantive impact and is quite negligible. One explanation for why $\gamma_{emp} < \gamma$ in some cases can be found in the approximate nature of the chi-square distribution of MLE standard errors. Because the normal distribution of maximum likelihood $\hat{\boldsymbol{\theta}}_n$ refers to the asymptotic property and at any finite sample size the normal distribution is approximate, the chi-square distribution for the sample variance for $\hat{\theta}_j$ is also approximate at finite sample sizes. In the present CFA context, ϕ_{21} in Model 3 converges less quickly than the model parameters in Tables 2–3. Because we developed sample size planning procedures based on the conventional MLE and CI formation methods in the literature, the success of sample size plan-

Table 6
Empirical Distributions of Confidence Interval Widths for γ_{11} in Model 2 ($\theta_j = \gamma_{11} = 0.6$)

				$\gamma = E[w]$			
ω	0.10	0.15	0.20	0.25	0.30	0.35	0.40
N	3,540	1,573	885	567	394	289	222
M_w	0.10003	0.15061	0.20073	0.25116	0.30321	0.35588	0.40684
Mdn_w	0.09989	0.15036	0.19965	0.24926	0.30038	0.35053	0.39962
				$\gamma = .80$			
ω	0.10	0.15	0.20	0.25	0.30	0.35	0.40
N	3,610	1,620	920	595	417	309	239
M_w	0.09912	0.14814	0.19725	0.24549	0.29474	0.34330	0.39106
Mdn_w	0.09899	0.14771	0.19666	0.24381	0.29231	0.33895	0.38388
P_{80}	0.10217	0.15479	0.20918	0.26363	0.31918	0.37749	0.43542
P_{75}	0.10149	0.15337	0.20654	0.25966	0.31361	0.36912	0.42459
P_{70}	0.10091	0.15217	0.20422	0.25601	0.30837	0.36241	0.41449
P_{80}/ω	1.02171	1.03190	1.04588	1.05453	1.06392	1.07855	1.08855
P_{70}/ω	1.00909	1.01445	1.02108	1.02404	1.02790	1.03545	1.03622
$(P_{80} - \omega)/\theta_j$	0.00362	0.00798	0.01529	0.02272	0.03196	0.04582	0.05903
$(P_{70} - \omega)/\theta_j$	0.00152	0.00361	0.00703	0.01002	0.01395	0.02068	0.02414
γ_{emp}	0.6108	0.6066	0.5986	0.608922	0.6072	0.600320	0.608766
α_{emp}	0.0530	0.0502	0.0506	0.051410	0.0504	0.049459	0.054885

Note. M_w and Mdn_w are the mean and median of the 5,000 random confidence interval widths. P_{80}, P_{75}, and P_{70} refer to the respective percentiles of the 5,000 random widths. γ_{emp} is the percentage of the random widths no larger than the desired value. α_{emp} is the empirical Type I error rate.

Table 7
Empirical Distributions of Confidence Interval Widths for λ_6 in Model 2 ($\theta_j = \lambda_6 = 0.3$)

			$\gamma = E[w]$			
ω	0.15	0.20	0.25	0.30	0.35	0.40
N	2,641	1,486	951	661	486	372
M_w	0.15072	0.20235	0.25444	0.30748	0.36142	0.41898
Mdn_w	0.15001	0.20041	0.25145	0.30139	0.35004	0.40377
			$\gamma = .80$			
ω	0.15	0.20	0.25	0.30	0.35	0.40
N	2,702	1,532	987	691	512	395
M_w	0.14903	0.19839	0.24912	0.29967	0.35068	0.40589
Mdn_w	0.14836	0.19704	0.24587	0.29419	0.34303	0.39301
P_{80}	0.15666	0.21132	0.26949	0.32863	0.38944	0.45563
P_{75}	0.15492	0.20842	0.26476	0.32117	0.37901	0.44115
P_{70}	0.15336	0.20562	0.26055	0.31458	0.37019	0.42963
P_{80}/ω	1.04439	1.05659	1.07794	1.09545	1.11267	1.13906
P_{70}/ω	1.02239	1.02810	1.04221	1.04860	1.05770	1.07407
$(P_{80} - \omega)/\theta_j$	0.02219	0.03773	0.06495	0.09545	0.13145	0.18542
$(P_{70} - \omega)/\theta_j$	0.01119	0.01873	0.03518	0.04860	0.06731	0.09876
γ_{emp}	0.565	0.580	0.5626	0.5650	0.5542	0.542651
α_{emp}	0.044	0.052	0.0458	0.0478	0.0450	0.041850

Note. M_w and Mdn_w are the mean and median of the 5,000 random confidence interval widths. P_{80}, P_{75}, and P_{70} refer to the respective percentiles of the 5,000 random widths. γ_{emp} is the percentage of the random widths no larger than the desired value. α_{emp} is the empirical Type I error rate.

ning procedures depends on the effectiveness of MLE. In cases where MLE is less effective, the sample size planning procedures will be less satisfactory, because the statistic of interest is not converging to the asymptotic distribution quickly enough at the current sample size, and using the asymptotic distribution to approximate the distribution at the current sample size involves nonnegligible errors. Nevertheless, given the trivial discrepancy between P_{80} and ω in Table 4, it is still justified to conclude that the extended method is effective in ensuring that the CI obtained in a particular study is sufficiently narrow with assurance close to the desired value.

SEM models. Tables 5–11 summarize the performance of the extended method in the SEM context. The method is less effective in the SEM models we studied than in the CFA model, in that γ_{emp} are all smaller than the specified assurance value. However, simply looking at γ_{emp} and γ can be misleading, and it would be unfair to conclude that the extended method is not effective in the SEM context without further inspection on the results.

Table 8
Empirical Distributions of Confidence Interval Widths for β_{21} in Model 3 ($\theta_j = \beta_{21} = 0.47$)

			$\gamma = E[w]$			
ω	0.10	0.15	0.20	0.25	0.30	0.35
N	3,450	1,533	863	552	384	282
M_w	0.10019	0.15065	0.20110	0.25348	0.30533	0.35853
Mdn_w	0.10009	0.15031	0.20022	0.25173	0.30256	0.35309
			$\gamma = .80$			
ω	0.10	0.15	0.20	0.25	0.30	0.35
N	3,520	1,579	898	580	407	302
M_w	0.09920	0.14828	0.19742	0.24722	0.29595	0.34657
Mdn_w	0.09908	0.14790	0.19656	0.24490	0.29258	0.34068
P_{80}	0.10206	0.15489	0.20875	0.26544	0.32169	0.38122
P_{75}	0.10146	0.15350	0.20613	0.26131	0.31561	0.37259
P_{70}	0.10098	0.15216	0.20405	0.25723	0.31076	0.36515
P_{80}/ω	1.02058	1.03258	1.04376	1.06176	1.07229	1.08919
P_{70}/ω	1.00982	1.01440	1.02026	1.02890	1.03586	1.04328
$(P_{80} - \omega)/\theta_j$	0.00438	0.01040	0.01862	0.03285	0.04614	0.06642
$(P_{70} - \omega)/\theta_j$	0.00209	0.00460	0.00862	0.01537	0.02289	0.03223
γ_{emp}	0.60580	0.60700	0.59620	0.58492	0.58991	0.58389
α_{emp}	0.04840	0.04660	0.04705	0.04382	0.05022	0.04893

Note. M_w and Mdn_w are the mean and median of the 5,000 random confidence interval widths. P_{80}, P_{75}, and P_{70} refer to the respective percentiles of the 5,000 random widths. γ_{emp} is the percentage of the random widths no larger than the desired value. α_{emp} is the empirical Type I error rate.

Table 9
Empirical Distributions of Confidence Interval Widths for λ_{10} in Model 3 ($\theta_j = \lambda_{10} = 0.83$)

				$\gamma = E[w]$			
ω	0.15	0.20	0.25	0.30	0.35	0.40	0.45
N	4,435	2,495	1,597	1,109	815	624	493
M_w	0.15041	0.20081	0.25190	0.30307	0.35532	0.40627	0.46136
Mdn_w	0.15019	0.19989	0.25012	0.30004	0.35099	0.39921	0.45079
				$\gamma = .80$			
ω	0.15	0.20	0.25	0.30	0.35	0.40	0.45
N	4,514	2,554	1,644	1,148	849	653	519
M_w	0.14896	0.19858	0.24802	0.29761	0.34744	0.39713	0.44877
Mdn_w	0.14857	0.19769	0.24692	0.29490	0.34261	0.39026	0.43861
P_{80}	0.15532	0.20983	0.26570	0.32255	0.38076	0.43934	0.50546
P_{75}	0.15389	0.20741	0.26164	0.31671	0.37314	0.42934	0.49074
P_{70}	0.15273	0.20508	0.25839	0.31160	0.36619	0.42022	0.47791
P_{80}/ω	1.03548	1.04917	1.06278	1.07517	1.08788	1.09834	1.12325
P_{70}/ω	1.01823	1.02538	1.03356	1.03865	1.04627	1.05056	1.06203
$(P_{80} - \omega)/\theta_j$	0.00641	0.01185	0.01891	0.02717	0.03706	0.04739	0.06682
$(P_{70} - \omega)/\theta_j$	0.00329	0.00612	0.01011	0.01397	0.01951	0.02437	0.03363
γ_{emp}	0.56720	0.56960	0.55760	0.56514	0.56814	0.56842	0.56464
α_{emp}	0.05280	0.04920	0.05520	0.05283	0.04990	0.04896	0.04692

Note. M_w and Mdn_w are the mean and median of the 5,000 random confidence interval widths. P_{80}, P_{75}, and P_{70} refer to the respective percentiles of the 5,000 random widths. γ_{emp} is the percentage of the random widths no larger than the desired value. α_{emp} is the empirical Type I error rate.

In most cases, at smaller sample sizes, the 80th percentiles of the random widths exceed the desired value by only a negligible amount. For example, P_{80} in Table 5 is 0.374 when $\omega = 0.35$ ($N = 315$) and 0.318 when $\omega = 0.30$ ($N = 425$); the differences are only .024 and .018, respectively. Similar patterns are observed in most of the situations, such as Table 6, where $P_{80} = 0.377$ and $\omega = 0.35$ when $N = 309$; Table 8, where $P_{80} = 0.322$ and $\omega = 0.30$ when $N = 407$; and Table 10, where $P_{80} = 0.107$ and $\omega = 0.10$ when $N = 570$.

There are some situations where the results at smaller sample sizes are less satisfactory. For example, when the desired width is 0.40, $P_{80} = 0.456$ ($N = 395$; see Table 7), exceeding ω by almost 14%. However, in this case the empirical Type I error rate is only .042, less than the nominal rate .05. Instead of forming intervals at the 95% confidence level, the MLE generally gives 96% CIs in this situation, which are wider than 95% ones. Therefore, at least part of the reason for the random CI widths being wider than ω by a nontrivial amount is because the MLE is not yet effective enough

Table 10
Empirical Distributions of Confidence Interval Widths for ϕ_{31} in Model 3 ($\theta_j = \phi_{31} = 0.14$)

				$\gamma = E[w]$				
ω	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
N	6,025	3,389	2,169	1,507	1,107	848	670	543
M_w	0.02999	0.04000	0.05011	0.06016	0.06999	0.08003	0.09017	0.10020
Mdn_w	0.02998	0.04000	0.05009	0.06010	0.06989	0.07993	0.08992	0.10009
				$\gamma = .80$				
ω	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
N	6,117	3,458	2,224	1,553	1,146	882	700	570
M_w	0.02977	0.03959	0.04939	0.05917	0.06880	0.07864	0.08816	0.09772
Mdn_w	0.02979	0.03958	0.04934	0.05915	0.06873	0.07858	0.08817	0.09745
P_{80}	0.03056	0.04099	0.05167	0.06230	0.07310	0.08400	0.09523	0.10658
P_{75}	0.03040	0.04073	0.05124	0.06165	0.07225	0.08292	0.09371	0.10476
P_{70}	0.03027	0.04049	0.05082	0.06111	0.07146	0.08188	0.09239	0.10326
P_{80}/ω	1.01859	1.02484	1.03343	1.03840	1.04435	1.05000	1.05806	1.06583
P_{70}/ω	1.00911	1.01236	1.01644	1.01842	1.02086	1.02348	1.02654	1.03257
$(P_{80} - \omega)/\theta_j$	0.00398	0.00710	0.01194	0.01646	0.02218	0.02857	0.03733	0.04702
$(P_{70} - \omega)/\theta_j$	0.00195	0.00353	0.00587	0.00790	0.01043	0.01341	0.01706	0.02326
γ_{emp}	0.5912	0.6016	0.5940	0.5860	0.5930	0.5890	0.5850	0.5970
α_{emp}	0.0448	0.0528	0.0546	0.0542	0.0528	0.0509	0.0564	0.0605

Note. M_w and Mdn_w are the mean and median of the 5,000 random confidence interval widths. P_{80}, P_{75}, and P_{70} refer to the respective percentiles of the 5,000 random widths. γ_{emp} is the percentage of the random widths no larger than the desired value. α_{emp} is the empirical Type I error rate.

Table 11
Empirical Distributions of Confidence Interval Widths for γ_{23} in Model 3 ($\theta_j = \gamma_{23} = 0.6$)

				γ = E[w]				
ω	0.15	0.20	0.25	0.30	0.40	0.50	0.60	0.70
N	6,147	3,458	2,213	1,537	865	554	385	283
M_w	0.15036	0.20088	0.25107	0.30225	0.40638	0.50994	0.62067	0.73600
Mdn_w	0.15000	0.20027	0.24993	0.30033	0.40173	0.50165	0.60517	0.71091
				γ = .80				
ω	0.15	0.20	0.25	0.30	0.40	0.50	0.60	0.70
N	6,240	3,528	2,269	1,583	900	582	408	303
M_w	0.14921	0.19897	0.24831	0.29806	0.39759	0.49956	0.60272	0.71296
Mdn_w	0.14896	0.19833	0.24748	0.29665	0.39311	0.49074	0.58775	0.68600
P_{80}	0.15437	0.20783	0.26228	0.31774	0.43254	0.55298	0.67729	0.81377
P_{75}	0.15326	0.20589	0.25916	0.31350	0.42426	0.54004	0.65950	0.78712
P_{70}	0.15234	0.20415	0.25631	0.30973	0.41788	0.52951	0.64174	0.76097
P_{80}/ω	1.02914	1.03913	1.04914	1.05912	1.08136	1.10595	1.12882	1.16253
P_{70}/ω	1.01560	1.02075	1.02524	1.03244	1.04470	1.05901	1.06957	1.08710
$(P_{80} - \omega)/\theta_j$	0.00728	0.01304	0.02047	0.02956	0.05424	0.08829	0.12882	0.18962
$(P_{70} - \omega)/\theta_j$	0.00390	0.00692	0.01052	0.01622	0.02980	0.04918	0.06957	0.10162
γ_{emp}	0.5610	0.5620	0.5580	0.5530	0.5620	0.5520	0.5490	0.5440
α_{emp}	0.0528	0.0468	0.0518	0.0426	0.0487	0.0442	0.0510	0.0461

Note. M_w and Mdn_w are the mean and median of the 5,000 random confidence interval widths. P_{80}, P_{75}, and P_{70} refer to the respective percentiles of the 5,000 random widths. γ_{emp} is the percentage of the random widths no larger than the desired value. α_{emp} is the empirical Type I error rate.

at the current sample size and gives larger standard error estimates than it should. Similar patterns can be found in Tables 9 and 11, where for smaller sample sizes P_{80} exceeds ω by a nontrivial amount (e.g., $P_{80}/\omega > 1.10$). Nevertheless, in these cases (i.e., Tables 7, 9, and 11) the 70th percentiles exceed ω by only a negligible amount. Consider $N = 395$ and $\omega = 0.40$ in Table 7 again. In this case $P_{70} = 0.4296$, exceeding ω by .0296, or about 7%, and the discrepancy can be considered as trivial. For another, Table 9 shows that $P_{70} = 0.4779$ when $N = 519$, and exceeds $\omega = 0.45$ by .0279, or about 6%. Given the above comparisons of the 70th and 80th quantiles to the desired width value, it can be seen that at smaller sample sizes where MLE is not effective enough and overestimates the standard errors, the extended method tends to underestimate the necessary sample. Within the simulation study in the present article, in worst cases the sample size returned by the extended method ensures that a CI obtained is sufficiently narrow about 70% of the time when the desired assurance is set to .80.

When the sample size is relatively large, in all cases the extended procedure returns the necessary sample sizes such that a CI obtained is sufficiently narrow with close to 80% assurance. Although γ_{emp} values for the SEM models are less than the specified value .80 in all cases, the 80th quantiles of those random CI widths are very close to the corresponding desired values, and the discrepancy between the two does not have substantive impact. For example, $P_{80} = 0.1022$ in Table 6 when $N = 3{,}610$ and $\omega = 0.10$, exceeding ω by 2.1%; $P_{80} = 0.102$ in Table 8 when $N = 3{,}520$ and $\omega = 0.10$, exceeding ω by 2%; $P_{80} = 0.154$ in Table 11 when $N = 6{,}240$ and $\omega = 0.15$, exceeding ω by 2.9%.

In summary, the extensive Monte Carlo simulation study indicates that both the standard and extended sample size planning procedures are effective in estimating the necessary sample size according to the specified goals. The standard method is quite effective in both the CFA and SEM contexts, and the sample size returned indeed ensures that the expected CI width is no wider than desired. The extended method is most effective in the CFA context, and in most of the cases the empirical proportion of sufficiently narrow random CIs is close to the specified .80 level. Although in the SEM context, the extended method tends to underestimate the necessary sample size in cases where MLE is less effective, the 80th percentile of the random CI widths usually exceeds the desired width by only a negligible amount that has no substantive impact, and the discrepancy between the 80th percentile and ω decreases as the sample size increases.

Issues in Statistical Theory

We discuss two issues regarding the assumptions of our sample size planning methods: (a) the model is correctly specified, and (b) data are normally distributed. Neither of these two assumptions is unique to our sample size planning methods; in fact, these assumptions are commonly made not only in sample size planning but in the SEM literature as a whole. For example, the power analysis on likelihood ratio tests of nested models is based on the assumption that the full model (i.e., the one with the path coefficients of interest estimated) is correctly specified (see, e.g., Hancock, 2006; Saris & Sattora, 1993; Satorra & Saris, 1985). In addition, the asymptotic chi-square distribution of the model chi-square statistic is based on the assumption that data follow a multivariate normal distribution (see, e.g., MacCallum et al., 2006; Steiger, Shapiro, & Browne, 1985). As discussed below, the implications of the violations of these two assumptions are still open to study.

Model Misspecifications

The methods developed in the present article are based on the assumption that the model is correctly specified, but in practice it may be difficult to develop an exactly correct model. However,

sample size planning can proceed as long as the proposed model is reasonable (i.e., close to the true model), under the additional assumption called *population drift* (or parameter drift), which in essence states that the proposed model is not badly misspecified and that lack of fit due to model misspecification is approximately of the same magnitude as lack of fit due to sampling error (Steiger et al., 1985; see also MacCallum et al., 2006). Actually the population drift assumption was used to derive the asymptotic chi-square distribution of the chi-square model fit statistic in Steiger et al. (1985), and thus is commonly recognized as a reasonable assumption in the SEM literature. For a more formal statement and discussion of this assumption, see Steiger et al. (1985; see also Browne & Cudeck, 1992; MacCallum et al., 2006).

Nevertheless, it is not clear when a "badly specified" model is bad enough, and in practice it is generally not possible to verify the population drift assumption. The issue of how misspecifications of models influence the empirical distribution of CI width and sample size planning methods is open to study, but it is not the focus of the present article. The effects of misspecifications of a model can be studied empirically with a priori Monte Carlo simulations in a manner similar to that discussed in Muthén and Muthén (2002) and Paxton et al. (2001). A random sample is generated based on the true model-implied covariance matrix, whereas the sample size is planned based on the proposed model (i.e., on the [incorrect] assumption that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$). Various properties of the random CI width can be studied based on a large number of replications of fitting the proposed model to the random data. Specialized R functions have been included in the MBESS package to implement such an a priori Monte Carlo simulation study, so that researchers can try with different models and study the effects of the input model on the returned sample size when planning the sample size.[13]

Nonnormal Distributions

It is still an issue open to study in the literature regarding the effects of nonnormal data on estimating various effect sizes in SEM (e.g., model fit indices, targeted effects), especially in the sample size planning stage, although some solutions to analyze nonnormal data have been developed (e.g., Browne, 1984; Kano, Berkane, & Bentler, 1990; Yuan & Bentler, 1997, 1998a, 1998b; Yuan, Bentler, & Zhang 2005; Yuan, Chan, & Bentler, 2000; Yuan & Hayashi, 2006). Generally, robust methods give larger standard error estimations than do standard methods (e.g., the Wald CI we reviewed in a previous section), which implies that the CI returned by robust methods tend to be wider than those returned by standard methods, other things being the same. Thus one should note that if the researcher plans the sample size based on our methods and uses robust methods to analyze the data, the CI obtained might not be sufficiently narrow, if the normality assumption is seriously violated. Nevertheless, we do recommend using robust methods to *analyze* data, because in reality the normality assumption is usually violated to different extents. We based our sample size planning methods on the normality assumption mainly because (a) robust methods, both analytical treatments and bootstrap applications, require raw data, which are unavailable in the research design stage; (b) the problem of nonnormal data can be alleviated by better or more careful data collecting methods and/or data transformation; and (c) in order to study the behavior of the CI width obtained in a study, which is a random variable dependent on $\hat{\boldsymbol{\theta}}$ and the covariance matrix of $\hat{\boldsymbol{\theta}}$, both of which are also random variables, certain assumptions must be made regarding the distribution of the data.

Discussion

When a researcher designs a study, there are at least three ways to decide the sample size: (a) collect as many participants as resources allow, (b) follow rules of thumb, or (c) apply formal sample size planning methods that attempt to accomplish a specified goal. Recruiting fewer participants than necessary can be a waste of resources because the study fails to achieve its goals. Using more than necessary participants is also a waste of resources in the sense that the extra data do not necessarily increase the understanding of the topic under study. Rules of thumb in the SEM literature are usually based on simulation studies. Those simulations were conducted in different contexts, and conclusions drawn regarding the "appropriate" sample size are based on various criteria. However, MacCallum, Widaman, Zhang, and Hong (1999), in the CFA context, demonstrated that rules of thumb are generally invalid because the performance of estimation is highly dependent on the model characteristics. Given that SEM is more complex than CFA, the rules of thumb should perform no better in SEM. In fact, by reviewing recent psychological studies that apply SEM, MacCallum and Austin (2000) recommended *against* the use of rules of thumb. The only reasonable option left is formal sample size planning methods. The necessary sample size can be planned with rigorously derived statistical methods so that the desired goals can be accomplished. Although sample size planning methods are based on assumptions that tend not to hold perfectly in reality, they are based on rigorous statistical reasoning and should perform reasonably well in most cases, as long as the assumptions are not seriously violated. Thus, the option of formal sample size planning methods is preferred.

The sample size for an SEM study can be planned from at least the power-analytic perspective or the AIPE perspective. When targeted effects in SEM are of interest, often of interest are not only their directions but their magnitude. The dichotomous result obtained by NHST (i.e., if the population effect is larger than zero or smaller than zero, without loss of generality) does not necessarily answer the question about the population effect size. On the other hand, a CI provides a range of plausible values for the population parameter of interest and thus helps better understand the magnitude of the population effect. With the goal of obtaining narrow CIs for the targeted effects of interest in SEM, sample size planning methods are developed from the AIPE perspective.

Sample size planning methods from AIPE require educated estimation of the population effect sizes, as does the power perspective (for power analysis in SEM, see, e.g., Hancock, 2006; Hancock & Freeman, 2001; Kim, 2005; MacCallum et al., 2006; MacCallum, Browne, & Sugawara, 1996; Satorra & Saris, 1985; for power analysis in general, see also Cohen, 1988; Murphy & Myors, 2004). The task of specifying the input effect sizes needs to be accomplished by consulting the literature. It is natural in the course of the accumulation

[13] This simulation study can be performed with the specialized R function ss.aipe.sem.path.sensitiv() in MBESS. See the Appendix for detailed information.

of knowledge that the specification of some not fully known information as input needs to be finished before conducting the investigation: Ongoing research updates the current knowledge based on early looks at the data from previous studies (Coffey & Muller, 2003; Proschan, 2005). In fact, Senn (2002) pointed out:

> This is common to all science. An astronomer does not know the magnitude of new stars until he has found them, but the magnitude of star he is looking for determines how much he has to spend on a telescope. (p. 1304)

In the spirit of the increasing urge that applied studies report effect sizes and CIs, we believe that sample size planning from the AIPE perspective should be considered when targeted effects are of interest in an SEM context. If an SEM model parameter of interest has a wide CI, its value is unclear. Thus a narrow CI is desirable because it excludes more implausible values and further pins down the population parameter. We believe the AIPE approach to planning sample sizes for SEM studies when targeted effects are of interest will help to better understand the magnitude of the effects in the population and evaluate substantive theories.

References

American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher, 35,* 33–40. doi:10.3102/0013189X035006033

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis* (3rd ed.). New York, NY: Wiley.

Arminger, G., & Schoenberg, R. J. (1989). Pseudo maximum likelihood estimation and a test for misspecification in mean and covariance structure models. *Psychometrika, 54,* 409–425. doi:10.1007/BF02294626

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606. doi:10.1037/0033-2909.88.3.588

Bentler, P. M., & Dudgeon, P. (1996). Covariance structure analysis: Statistical practice, theory, directions. *Annual Review of Psychology, 47,* 563–592. doi:10.1146/annurev.psych.47.1.563

Beran, R., & Srivastava, M. S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *Annuals of Statistics, 13,* 95–115. doi:10.1214/aos/1176346579

Bollen, K. A. (1989). *Structural equations with latent variables.* New York, NY: Wiley.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53,* 605–634. doi:10.1146/annurev.psych.53.100901.135239

Bollen, K. A., & Stine, R. (1993). Bootstrapping goodness-of-fit measures in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 111–135). Newbury Park, CA: Sage.

Browne, M. W. (1984). Asymptotic distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37,* 62–83.

Browne, M. W., & Arminger, G. (1995). Specification and estimation of mean- and covariance-structure models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 185–249). New York, NY: Plenum.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21,* 230–258. doi:10.1177/0049124192021002005

Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Belmont, CA: Duxbury Press.

Coffey, C. S., & Muller, K. E. (2003). Properties of internal pilots with the univariate approach to repeated measures. *Statistics in Medicine, 22,* 2469–2485. doi:10.1002/sim.1466

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49,* 997–1003. doi:10.1037/0003-066X.49.12.997

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105,* 317–327. doi:10.1037/0033-2909.105.2.317

Curran, P. J., Bollen, K. A., Chen, F., Paxton, P., & Kirby, J. B. (2003). Finite sampling properties of the point estimates and confidence intervals of the RMSEA. *Sociological Methods & Research, 32,* 208–252. doi:10.1177/0049124103256130

Curran-Everett, D., & Benos, D. J. (2004). Guidelines for reporting statistics in journals published by the American Physiological Society. *Journal of Neurophysiology, 92,* 669–671. doi:10.1152/jn.00534.2004

Davey, A., & Savla, J. (2010). *Statistical power analysis with missing data: A structural equation modeling approach.* New York, NY: Taylor & Francis.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* New York, NY: Chapman & Hall/CRC.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18,* 39–50. doi:10.2307/3151312

Fox, J. (2006). Structural equation modeling with the sem package in R. *Structural Equation Modeling, 13,* 465–486. doi:10.1207/s15328007sem1303_7

Green, B. F. (1977). Parameter sensitivity in multivariate methods. *Multivariate Behavioral Research, 12,* 263–287. doi:10.1207/s15327906mbr1203_1

Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research, 26,* 499–510. doi:10.1207/s15327906mbr2603_7

Guenther, W. C. (1965). *Concepts of statistical inference.* New York, NY: McGraw-Hill.

Hancock, G. R. (2006). Power analysis in covariance structure modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 69–115). Greenwich, CT: Information Age.

Hancock, G. R., & Freeman, M. J. (2001). Power and sample size for the root mean square error of approximation test of not close fit in structural equation modeling. *Educational and Psychological Measurement, 61,* 741–758. doi:10.1177/00131640121971491

Hancock, G. R., & Kroopnick, M. H. (2005, April). *A simplified sample size determination for tests of latent covariance structure parameters.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. H. C. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Past and present. A Festschrift in honor of Karl G. Jöreskog* (pp. 195–216). Chicago, IL: Scientific Software International.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.

Holzinger, K. J., & Swineford, F. (1939). *A study in factor analysis: The stability of a bi-factor solution.* Chicago, IL: University of Chicago.

International Committee of Medical Journal Editors. (2004). *Uniform*

requirements for manuscripts submitted to biomedical journals: Writing and editing for biomedical publication. Philadelphia, PA: Author.

Kano, Y., Berkane, M., & Bentler, P. M. (1990). Covariance structure analysis with heterogeneous kurtosis parameters. *Biometrika, 77,* 575–585. doi:10.1093/biomet/77.3.575

Kelley, K. (2007a). Confidence intervals for standardized effect sizes: Theory, application, and implementation. *Journal of Statistical Software, 20,* 1–24.

Kelley, K. (2007b). Methods for the behavioral, educational, and social sciences: An R package. *Behavior Research Methods, 39,* 979–984.

Kelley, K. (2007c). Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach. *Behavior Research Methods, 39,* 755–766.

Kelley, K. (2008). Sample size planning for the squared multiple correlation coefficient: Accuracy in parameter estimation via narrow confidence intervals. *Multivariate Behavioral Research, 43,* 524–555. doi:10.1080/00273170802490632

Kelley, K., & Lai, K. (2010). *MBESS* [Computer software manual]. Retrieved from http://www.cran.r-project.org

Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods, 8,* 305–321. doi:10.1037/1082-989X.8.3.305

Kelley, K., Maxwell, S. E., & Rausch, J. R. (2003). Obtaining power or obtaining precision: Delineating methods of sample-size planning. *Evaluation and the Health Professions, 26,* 258–287. doi:10.1177/0163278703255242

Kelley, K., & Rausch, J. R. (2006). Sample size planning for the standardized mean difference: Accuracy in parameter estimation via narrow confidence intervals. *Psychological Methods, 11,* 363–385. doi:10.1037/1082-989X.11.4.363

Kim, K. H. (2005). The relation among fit indexes, power, and sample size in structural equation modeling. *Structural Equation Modeling, 12,* 368–390. doi:10.1207/s15328007sem1203_2

Lee, S.-Y. (1985). Analysis of covariance and correlation structures. *Computational Statistics & Data Analysis, 2,* 279–295. doi:10.1016/0167-9473(85)90002-7

Lehmann, E. L. (1998). *Elements of large-sample theory.* New York, NY: Springer.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51,* 201–226. doi:10.1146/annurev.psych.51.1.201

MacCallum, R. C., Browne, M. W., & Cai, L. (2006). Testing differences between nested covariance structure models: Power analysis and null hypotheses. *Psychological Methods, 11,* 19–35. doi:10.1037/1082-989X.11.1.19

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. doi:10.1037/1082-989X.1.2.130

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4,* 84–99. doi:10.1037/1082-989X.4.1.84

Mace, A. E. (1964). *Sample-size determination.* New York, NY: Reinhold.

Maruyama, G., & McGarvey, B. (1980). Evaluating causal models: An application of maximum-likelihood analysis of structural equations. *Psychological Bulletin, 87,* 502–512. doi:10.1037/0033-2909.87.3.502

Maxwell, S. E. (2000). Sample size and multiple regression analysis. *Psychological Methods, 5,* 434–458. doi:10.1037/1082-989X.5.4.434

Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology, 59,* 537–563. doi:10.1146/annurev.psych.59.103006.093735

McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the reticular action model for moment structures. *British Journal of Mathematical and Statistical Psychology, 37,* 234–251.

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Erlbaum.

Meehl, P. E. (1997). The problem in epistemology, not statistics: Replacing significance tests by confidence intervals and quantify accuracy of risky numerical predictions. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 393–426). Mahwah, NJ: Erlbaum.

Mulaik, S. A. (2009). *Linear causal modeling with structural equations.* Boca Raton, FL: Chapman & Hall.

Murphy, K. R., & Myors, B. (2004). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests* (2nd ed.). Mahwah, NJ: Erlbaum.

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. doi:10.1207/S15328007SEM0904_8

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5,* 241–301. doi:10.1037/1082-989X.5.2.241

Pawitan, Y. (2001). *In all likelihood: Statistical modelling and inference using likelihood.* New York, NY: Oxford University Press.

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling, 8,* 287–312. doi:10.1207/S15328007SEM0802_7

Proschan, M. A. (2005). Two-stage sample size re-estimation based on a nuisance parameter: A review. *Journal of Biopharmaceutical Statistics, 15,* 559–574. doi:10.1081/BIP-200062852

Raju, N. S., Bilgic, R., Edwards, J. E., & Fleer, P. F. (1999). Accuracy of population validity and cross-validity estimation: An empirical comparison of formula-based, traditional empirical, and equal weights procedures. *Applied Psychological Measurement, 23,* 99–115. doi:10.1177/01466219922031220

Raykov, T. (2004). Estimation of maximal reliability: A note on a covariance structure modelling approach. *British Journal of Mathematical and Statistical Psychology, 57,* 21–27. doi:10.1348/000711004849295

R Development Core Team. (2010). *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria: R Foundation for Statistical Computing.

Rigdon, E. E. (1998). The equal correlation baseline model for comparative fit assessment in structural equation modeling. *Structural Equation Modeling, 5,* 63–77. doi:10.1080/10705519809540089

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). Newbury Park, CA: Sage.

Satorra, A., & Saris, W. E. (1985). The power of the likelihood ratio test in covariance structure analysis. *Psychometrika, 50,* 83–90. doi:10.1007/BF02294150

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods, 1,* 115–129. doi:10.1037/1082-989X.1.2.115

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105,* 309–316. doi:10.1037/0033-2909.105.2.309

Senn, S. J. (2002). Power is indeed irrelevant in interpreting completed studies. *British Medical Journal, 325,* 1304. doi:10.1136/bmj.325.7375.1304

Shapiro, A. (1983). Asymptotic distribution theory in the analysis of covariance structures (a unified approach). *South African Statistical Journal, 17,* 33–81.

Shapiro, A., & Browne, M. W. (1990). On the treatment of correlation structures as covariance structures. *Linear Algebra and Its Applications, 127,* 567–587. doi:10.1016/0024-3795(90)90362-G

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. doi:10.1007/BF02294104

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York, NY: Springer.

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin, 83,* 213–217. doi:10.1037/0033-2909.83.2.213

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica, 50,* 1–25. doi:10.2307/1912526

Yuan, K.-H., & Bentler, P. M. (1997). Improving parameter tests in covariance structure analysis. *Computational Statistics & Data Analysis, 26,* 177–198. doi:10.1016/S0167-9473(97)00025-X

Yuan, K.-H., & Bentler, P. M. (1998a). Robust mean and covariance structure analysis. *British Journal of Mathematical and Statistical Psychology, 51,* 63–88.

Yuan, K.-H., & Bentler, P. M. (1998b). Structural equation modeling with robust covariances. *Sociological Methodology, 28,* 363–396. doi:10.1111/0081-1750.00052

Yuan, K.-H., Bentler, P. M., & Zhang, W. (2005). The effect of skewness and kurtosis on mean and covariance structure analysis: The univariate case and its multivariate implication. *Sociological Methods & Research, 34,* 240–258. doi:10.1177/0049124105280200

Yuan, K.-H., Chan, W., & Bentler, P. M. (2000). Robust transformation with applications to structural equation modelling. *British Journal of Mathematical and Statistical Psychology, 53,* 31–50. doi:10.1348/000711000159169

Yuan, K.-H., & Hayashi, K. (2006). Standard errors in covariance structure models: Asymptotics versus bootstrap. *British Journal of Mathematical and Statistical Psychology, 59,* 397–417. doi:10.1348/000711005X85896

Appendix

Implementing the Methods Discussed With MBESS

All the methods discussed in the present article have been implemented in R functions (R Development Core Team, 2010) in the MBESS package (Kelley, 2007a, 2007c; Kelley & Lai, 2010). This appendix briefly discusses how to use those functions to plan for sample size; more detailed documentation is available in the help files in the MBESS package. Functions in the appendix call the sem package (Fox, 2006) in R when they need to fit structural equation models (SEMs). Models are specified in the RAM notation (see, e.g., McArdle & McDonald, 1984) in the same manner as the sem R package, which uses a one-headed arrow to represent the regression effect and a two-headed arrow to represent the variance or covariance. A successfully specified model object should be of class "mod" in R. Readers interested in how to specify models in R using the RAM notation are referred to Fox (2006) or the help files of the sem and MBESS packages. This appendix also includes a brief example about SEM model specifications in the RAM notation.

To plan for sample size, the user first needs to specify the model, which can be performed with the function specify.model() in the sem package. For example, to specify the one-factor confirmatory factor analysis (CFA) model in Figure 1, the following syntax can be used:

```
> library(MBESS)
> library(sem)
> model.cfa <- specify.model()
xi1 -> X1, lam1, 0.5
xi1 -> X2, lam2, 0.5
xi1 -> X3, lam3, 0.5
xi1 -> X4, lam4, 0.5
xi1 <-> xi1, NA, 1
X1 <-> X1, del1, 0.5
X2 <-> X2, del2, 0.5
X3 <-> X3, del3, 0.5
X4 <-> X4, del4, 0.5
```

In each line of the model specification above, there are three parts, separated by commas. The first part indicates the path, the second part indicates the name of this particular path coefficient, and the third part indicates the starting value for this path coefficient in estimation. If the path name is NA, then the path is fixed at the value in the third part. For example, the variance of xi1 in this model is fixed at 1.

The function theta.2.Sigma.theta() in MBESS creates the model-implied covariance matrix given a model and model parameter values. This function is not used in sample size planning but is convenient for constructing a covariance matrix as input for sample size planning or conducting simulation studies. For example, given the population standardized model parameters in Figure 1, the model-implied correlation matrix can be obtained as follows:

(*Appendix continues*)

```
> theta.cfa <- c(.4, .5, .6, .7, .84, .75, .64, .51)
> names(theta.cfa) <- c("lam1", "lam2", "lam3", "lam4",
  "del1", "del2", "del3", "del4")
> res <- theta.2.Sigma.theta(model=model.cfa, theta=
  theta.cfa, latent.vars="xi1")
> cor.mat <- res$Sigma.theta
```

The argument model is an object of "mod" class containing the specifications of an SEM model, theta is a vector of model parameters, and latent.vars is a vector indicating the names of the latent variables in the model.

The function ss.aipe.sem.path() plans for sample size for SEM with the goal of obtaining a sufficiently narrow confidence interval (CI) for a model parameter of interest. Suppose λ_1 in the above one-factor CFA model is of interest, and it is desired to have a confidence interval for λ_1 no wider than 0.3. To replicate the example that we used to illustrate the concept of exchangeable correlation structure in the text, we specify the standard deviations of X_1, X_2, X_3, and X_4 to be 1.2, 1.5, 1.8, and 2, respectively. Then the following syntax can be used to calculate the necessary sample size:

```
> SD <- diag(c(1.2, 1.5, 1.8, 2))
> rownames(SD)<-colnames(SD)<- c("X1","X2","X3","X4")
> Sigma.cfa <- SD %*% cor.mat %*% SD
> ss.aipe.sem.path(model=model.cfa, Sigma=Sigma.cfa,
  desired.width=0.3, which.path="lam1")
```

The sample size returned is 349. The argument Sigma refers to the input covariance matrix of the manifest variables, desired.width refers to the desired CI width, and which.path is name of the model parameter of interest. If one desires to have .80 assurance that the CI obtained will be no wider than desired, the argument assurance can be added to the function:

```
> ss.aipe.sem.path(model=model.cfa, Sigma=Sigma.cfa,
  desired.width=0.3, which.path="lam1", assurance=.80)
```

The function ss.aipe.sem.path.sensitiv() can be used to conduct a priori Monte Carlo simulations to study how misspecifications of model and/or the covariance matrix of manifest variables affect the effectiveness of sample size planning methods. An example call to this function is

```
> ss.aipe.sem.path.sensitiv(model, est.Sigma, true.Sigma=
  est.Sigma, which.path, desired.width, G=1000),
```

where random data are generated based on the true covariance matrix of manifest variables (true.Sigma) but fit to the proposed model (model), whereas the sample size is planned based on the proposed model and the estimated covariance matrix of manifest variables (est.Sigma). Various properties of the random CI widths can be studied empirically by replicating the process a large number of times (i.e., the argument G).

Received May 14, 2009
Revision received September 2, 2010
Accepted September 7, 2010 ■