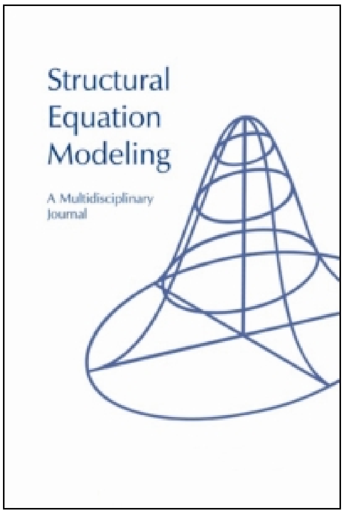
Small-sample robust estimators of noncentrality-based and incremental model fit

Boomsma, Anne; Herzog, W.

Published in:
Structural Equation Modeling

DOI:
10.1080/10705510802561279



Structural Equation Modeling: A Multidisciplinary Journal
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653699

Small-Sample Robust Estimators of Noncentrality-Based and Incremental Model Fit

Walter Herzog[a]; Anne Boomsma[b]
[a] University of St. Gallen, Switzerland [b] University of Groningen, The Netherlands

To cite this Article Herzog, Walter and Boomsma, Anne(2009) 'Small-Sample Robust Estimators of Noncentrality-Based and Incremental Model Fit', Structural Equation Modeling: A Multidisciplinary Journal, 16: 1, 1 — 27
To link to this Article: DOI: 10.1080/10705510802561279
URL: http://dx.doi.org/10.1080/10705510802561279

Structural Equation Modeling, 16:1–27, 2009

ISSN: 1070-5511 print/1532-8007 online
DOI: 10.1080/10705510802561279

Psychology Press
Taylor & Francis Group

Small–Sample Robust Estimators of Noncentrality–Based and Incremental Model Fit

Walter Herzog
University of St. Gallen, Switzerland

Anne Boomsma
University of Groningen, The Netherlands

Traditional estimators of fit measures based on the noncentral chi–square distribution (root mean square error of approximation [RMSEA], Steiger's γ, etc.) tend to overreject acceptable models when the sample size is small. To handle this problem, it is proposed to employ Bartlett's (1950), Yuan's (2005), or Swain's (1975) correction of the maximum likelihood chi–square statistic for the estimation of noncentrality–based fit measures. In a Monte Carlo study, it is shown that Swain's correction especially produces reliable estimates and confidence intervals for different degrees of model misspecification (RMSEA range: 0.000–0.096) and sample sizes (50, 75, 100, 150, 200). In the second part of the article, the study is extended to incremental fit indexes (Tucker–Lewis Index, Comparative Fit Index, etc.). For their small–sample robust estimation, use of Swain's correction is recommended only for the target model, not for the independence model. The Swain–corrected estimators only require a ratio of sample size to estimated parameters of about 2:1 (sometimes even less) and are thus strongly recommended for applied research. R software is provided for convenient use.

Despite the prominence of model fit statistics based on the noncentral chi–square distribution in applied research, there are surprisingly few investigations on their small–sample behavior. The studies by Curran et al. (2002), Curran et al. (2003), and Hu and Bentler (1999) indicate that sample size should be at least 200 to achieve robust inference based on the noncentral chi–square distribution. When sample size is smaller, the population noncentrality parameter is overestimated with the undesirable consequence that noncentrality–based statistics tend to overreject acceptable models. This finding was revalidated in a study by Olsson, Foss, and Breivik (2004). These authors also found that a sample size of 200 is not enough when model

Correspondence should be addressed to Walter Herzog, Institute of Marketing and Retailing, Dufourstrasse 40a, CH–9000 St. Gallen, Switzerland. E-mail: walter.herzog@unisg.ch

size increases, thereby implicitly highlighting the importance of considering sample size in relation to model size, as measured, for example, by the number of free parameters (e.g., Herzog, Boomsma, & Reinecke, 2007; Nevitt & Hancock, 2004).

Applied researchers, however, often deal with relatively small ratios of sample size to model size. In dyadic studies (Kenny, Kashy, & Cook, 2006), for example, it is very hard to get large data sets (e.g., Homburg & Fürst, 2005). Compared to observational studies, sample sizes in experimental studies are usually also small, which is one of the main reasons experimental researchers so rarely apply covariance structure methodology (Tomarken & Waller, 2005). The robustness against small sample size is still one of the most serious challenges in covariance structure modeling. Irrespective of model size, some applied researchers dare not even think of using this methodology when the number of observations is less than 200.

Our general objective in this study is to reduce problems associated with small sample sizes under varying degrees of model misspecification. The more specific aim of the investigation is twofold. First, we propose employing Bartlett's (1950), Swain's (1975), or Yuan's (2005) correction of the maximum likelihood chi–square statistic for the computation of noncentrality–based fit statistics. The primary interest here is the behavior of these corrected estimators compared to the traditional method when (a) sample size relative to model size decreases, and (b) the degree of misspecification increases. The goal is to select an estimator that performs best in approximating a noncentral chi–square distribution across a wide range of sample sizes and degrees of misspecification. Performance criteria are relative mean bias, relative standard deviation bias, and coverage rates of confidence intervals for the population noncentrality parameter. We also illustrate the practical relevance of the findings in the root mean square error of approximation (RMSEA) metric of the noncentrality parameter. Second, the study is extended to the independence model in the framework of incremental fit indexes, using the Tucker–Lewis Index (TLI) as an illustration. In summary, we want to provide applied researchers with small–sample robust estimators of both noncentrality–based and incremental population model fit indexes.

The article is structured as follows. First, covariance structure tests based on the central and the noncentral chi–square distribution are discussed formally, and relevant Monte Carlo studies are reviewed. Second, modified estimators of noncentrality–based model fit indexes are introduced with reference to Bartlett (1950), Yuan (2005), and Swain (1975), and expectations about the small–sample behavior of the statistics under study are formulated. Third, the design of our Monte Carlo investigation is described and results for noncentrality–based statistics are presented. In a supplementary analysis, the study is extended to the independence model and the small–sample robust estimation of incremental model fit indexes. Finally, the results are illustrated by correcting estimates of noncentrality–based and incremental fit measures reported in a recently published small–sample study.

CENTRAL CHI–SQUARE DISTRIBUTION

In the following, consider a vector of p random variables $\mathbf{z}$ $(p \times 1)$ with a corresponding empirical sample covariance matrix $\mathbf{S}$ $(p \times p)$ based on $N = n + 1$ independent observations, and a population model with covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ $(p \times p)$, where $\boldsymbol{\theta}$ $(t \times 1)$ is a vector of independent model parameters to be estimated. It is well known that, given the sample covariance matrix $\mathbf{S}$, the minimization of

$$F_{ML}[\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})] = \log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \mathrm{tr}[\mathbf{S}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}] - \log|\mathbf{S}| - p \tag{1}$$

yields the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$.

We define $\boldsymbol{\Sigma}_0$ as the true population covariance matrix generated by the population model M_0 of the p variables, and $\boldsymbol{\Sigma}(\boldsymbol{\theta}_j)$ as the population covariance matrix implied by a postulated model M_j. The null hypothesis $H_0 : \boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_j)$, that is, the hypothesis that the postulated model holds exactly in the population, can then be tested against the alternative that $\boldsymbol{\Sigma}_0$ is any positive definite covariance matrix, $\boldsymbol{\Omega}$ say; the alternative hypothesis can be expressed as $H_1 : \boldsymbol{\Sigma}_0 = \boldsymbol{\Omega}$. If $\boldsymbol{\Omega} \equiv \mathbf{S}$ and H_0 holds, the distribution of the likelihood ratio test statistic

$$T_{ML} = nF_{ML}[\mathbf{S}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j)] \tag{2}$$

converges to a central chi–square distribution with $d = p(p+1)/2 - t$ degrees of freedom as the sample size N increases (Wilks, 1938). The likelihood ratio test statistic can be used to test whether the proposed model should be rejected at a certain significance level: the so–called test of exact fit.

It should be noted that the test of exact fit is sensitive to violations of underlying assumptions. These are mainly independent observations, multivariate normality of the observed variables, and a large sample size (the functioning of asymptotic theory). If these assumptions are violated, empirical Type I error rates are typically too large and population models are rejected too frequently (e.g., Hu, Bentler, & Kano, 1992; Savalei, 2008).

NONCENTRAL CHI–SQUARE DISTRIBUTION

A rather implicit assumption underlying the derivations in the previous section is that a covariance structure model can in principle hold exactly in the population, as reflected by $H_0 : \boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_j)$. However, use of a statistical model implies that a researcher is willing to abstract from reality's complexity and therefore, every covariance structure model is by definition only an approximation of reality or, negatively formulated, "wrong" to some degree. Hence, it seems hardly appropriate to test whether a model holds exactly in the population. From the perspective that models never fit exactly, it is of far more practical interest to infer to what degree a model–implied covariance matrix differs from the population covariance matrix. This issue was raised by Steiger and Lind (1980), refined by Steiger, Shapiro, and Browne (1985), and elaborated on by Cudeck and Henly (1991) and Browne and Cudeck (1993).

Cudeck and Henly (1991) defined three types of discrepancies or errors (see also Browne & Cudeck, 1993). Let $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ denote the true population covariance matrix resulting from the population or "operating" model M_0 (Cudeck & Henly, 1991), which, as argued, never exactly corresponds to the estimated or postulated model M_j of the researcher. Moreover, let $\tilde{\boldsymbol{\Sigma}}_j = \boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}}_j)$ represent the best fit of the postulated model M_j to $\boldsymbol{\Sigma}_0$; that is, $\tilde{\boldsymbol{\theta}}_j =$ arg min $F[\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}(\boldsymbol{\theta}_j)]$, where F is a discrepancy function as defined by Browne (1984, p. 64). The so–called error of approximation refers to the lack of fit of model M_j to the population model M_0 by looking at differences between $\boldsymbol{\Sigma}_0$ and $\tilde{\boldsymbol{\Sigma}}_j$ in terms of F: $F_0 = F(\boldsymbol{\Sigma}_0, \tilde{\boldsymbol{\Sigma}}_j)$.

Furthermore, let $\hat{\boldsymbol{\Sigma}}_j = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j)$ represent the best fit of model M_j to the sample covariance matrix $\mathbf{S}$; that is, $\hat{\boldsymbol{\theta}}_j = \arg\min F[\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_j)]$. The error of estimation refers to differences between $\tilde{\boldsymbol{\Sigma}}_j$ and $\hat{\boldsymbol{\Sigma}}_j$ in terms of F: $F_e = F(\tilde{\boldsymbol{\Sigma}}_j, \hat{\boldsymbol{\Sigma}}_j)$.

Finally, the overall error denotes the difference between $\boldsymbol{\Sigma}_0$ and $\hat{\boldsymbol{\Sigma}}_j$ in terms of F: $F_t = F(\boldsymbol{\Sigma}_0, \hat{\boldsymbol{\Sigma}}_j)$.

In the remainder of this article, we are only concerned with $F = F_{ML}$; see Equation 1. Hence, in the sequel, the maximum likelihood (ML) discrepancy function defines the error of approximation, the error of estimation, and the overall error.

Under the crucial assumption that the error of approximation is small relative to the error of estimation (in addition to the assumptions concerning the test of exact fit mentioned in the previous section), it has been shown that T_{ML} asymptotically follows a noncentral chi–square distribution with degrees of freedom d and noncentrality parameter $\lambda = nF_0$ (Steiger et al., 1985). A noncentral chi–square variate has an expectation of $d + \lambda$ and a standard deviation of $\sqrt{2d + 4\lambda}$.

The noncentral chi–square distribution explicitly incorporates the amount of substantial model error in terms of F_0 or λ. Since Steiger and Lind's (1980) seminal presentation to the Psychometric Society, attention has shifted from "statistical significance" of model error to "practical significance" of model error. The major drawback with the latter perspective, however, is that one cannot take refuge in mere significance testing of $F_0 = 0$, which is equivalent to the test of exact fit. Instead, one has to distinguish carefully between close– and not–close–fitting models in terms of the error of approximation, F_0. For that purpose, Steiger and Lind (1980) introduced the RMSEA as a metric of F_0 or λ, which is defined as the square rooted error of approximation per degree of freedom:

$$\text{RMSEA} = \sqrt{\frac{F_0}{d}} = \sqrt{\frac{\lambda}{dn}}. \tag{3}$$

The lower bound of RMSEA is zero because the population noncentrality parameter λ is nonnegative by definition, but the upper bound of RMSEA is not normed. The division by degrees of freedom d takes model complexity into account: In more complex models, more parameters t have to be estimated and, given a fixed number of observed variables p, the reduced degrees of freedom d result in larger values of RMSEA; that is, in less close model fit. As a consequence, conditional on the sample data, parsimonious models are favored over complicated models in RMSEA–based model evaluation. Further, the square root in Equation 3 prevents the occurrence of very small RMSEA values that might be difficult to interpret. In addition, the square root causes the second–order derivative of RMSEA with respect to F_0 or λ to be negative. Therefore, Equation 3 is more sensitive to changes in the error of approximation when F_0 is small than when it is large. As a result, RMSEA can discriminate between "good" and "rather good" models but it is less appropriate to discriminate between "bad" and "very bad" models.

Browne and Cudeck (1993, p. 144) suggested that an RMSEA value of less than 0.05 indicates "close" fit, a value of 0.05 up to 0.08 indicates "reasonable" fit, and that "one would not want to employ a model with an RMSEA greater than 0.10." Rules of thumb, however, are always subject to criticism and should be interpreted with caution and tolerance (e.g., Chen et al., 2008; Marsh, Hau, & Wen, 2004).

The noncentrality parameter λ is usually estimated by $\hat{\lambda} = \max(T_{ML} - d; 0)$. As a result, the population RMSEA is estimated by

$$\widehat{\text{RMSEA}} = \sqrt{\frac{\hat{\lambda}}{dn}}. \tag{4}$$

Because it is known that the sampling distribution of T_{ML} matches a noncentral chi–square distribution under conditions already mentioned, it is possible to estimate confidence limits for λ and RMSEA. In deriving confidence limits for λ, one usually employs the cumulative distribution function $G(T_{ML}|d, \lambda)$. If one is willing to accept a Type I error probability of α, then a lower confidence limit for λ can be calculated by solving for $\hat{\lambda}_L$ in the equation $G(T_{ML}|d, \hat{\lambda}_L) = 1 - \alpha/2$. This is a nonlinear numerical problem and any root–finding routine implemented in statistical software packages will be useful for this purpose. Similarly, it is possible to obtain the upper confidence limit for λ by solving for $\hat{\lambda}_U$ in the equation $G(T_{ML}|d, \hat{\lambda}_U) = \alpha/2$. It should be noted that the estimated confidence limits are in general not equidistant to the point estimate $\hat{\lambda}$. It follows that for every λ

$$P(\hat{\lambda}_L < \lambda < \hat{\lambda}_U) = 1 - \alpha, \tag{5}$$

and it is straightforward that for every RMSEA

$$P\left(\sqrt{\frac{\hat{\lambda}_L}{dn}} < \text{RMSEA} < \sqrt{\frac{\hat{\lambda}_U}{dn}}\right) = 1 - \alpha. \tag{6}$$

Note that Equations 5 and 6 also contain information about the test of exact fit, as the test of exact fit will be significant at error level $\alpha/2$ when $\hat{\lambda}_L > 0$ (Browne & Cudeck, 1993). For example, a positive $\hat{\lambda}_L$ based on a 90% confidence region means that the test of exact fit is significant at a 5% Type I error level.

The usefulness of the noncentral chi–square distribution for the evaluation of model fit is in most cases illustrated with reference to the RMSEA metric of λ, although there are many other interesting fit indexes that are functions of λ, for example Steiger's γ or McDonald's Centrality Index (for a discussion see Hu & Bentler, 1999). In principle, one can also estimate confidence intervals for corresponding population counterparts of these latter statistics. However, the RMSEA has clearly received the most attention in applied research, and it has also been recommended that this statistic be reported in combination with the standardized root mean square residual (SRMR; Hu & Bentler, 1999). Due to its practical relevance, we also present the results of our simulation study in the RMSEA metric of the noncentrality parameter. It should be noted, however, that our findings and recommendations hold for other noncentrality–based statistics as well (e.g., Steiger's γ, McDonald's Centrality Index, etc.).

As indicated, model evaluating inference in covariance structure modeling based on the central and the noncentral chi–square distribution is exact only asymptotically; that is, if sample size increases to infinity. Although there are a number of studies on the approximation of the central chi–square distribution with small sample sizes (e.g., Bentler & Yuan, 1999; Boomsma, 1983; Fouladi, 2000; Nevitt & Hancock, 2004), there is very limited empirical research on the small–sample behavior of noncentral chi–square statistics in covariance structure modeling.

The main message of these few studies is that noncentrality–based fit statistics, and especially $\widehat{\text{RMSEA}}$, tend to overreject acceptable models when sample size is small, say $N \leq 200$ (Curran et al., 2003; Hu & Bentler, 1999). This is because mean and variance of the noncentral chi–square distribution are overestimated when small samples are analyzed (Curran et al., 2002; Olsson et al., 2004).

Furthermore, the studies by Curran et al. (2003) and Olsson et al. (2004) confirm that the overestimation of mean and variance of the noncentral chi–square distribution leads to a decreased coverage rate when confidence intervals for λ and RMSEA are estimated using Equations 5 and 6, respectively.

It is also worth noting that Olsson et al. (2004) found that the approximation of the noncentral chi–square distribution breaks down with increasing model size. As a result, sample size recommendations relative to model size might be of advantage, for example, in terms of *N:t* ratios. As mentioned earlier, however, although rules of thumb, like *N:t* ratios, might provide sensible guidelines for applied researchers in many situations, they should always be employed with abundant vigilance and leniency (e.g., Brown, 2006; Jackson, 2001, 2003, 2007).

Large *N:t* ratios are rather rare in applied research. At the outset of this article, it was emphasized that in particular experimental and dyadic studies are often based on small sample sizes. In the next section, we introduce alternative estimators of noncentrality–based population model fit that are supposed to work even under very small *N:t* ratios, without losing the described asymptotic properties.

SMALL–SAMPLE CORRECTIONS

In the following, three corrective procedures are introduced that are known to improve the small–sample performance of T_{ML} under the assumption that H_0 holds; that is, they are designed to robustify the test of exact fit against small sample sizes. Although, to our knowledge, nothing is known about their performance under model misspecification, it is rather intuitive to employ these corrections in the noncentral case as well (see, e.g., Steiger et al., 1985). We first introduce Bartlett's (1950), Yuan's (2005), and Swain's (1975) correction for the test of exact fit. Expectations about the performance of the statistics under study in the noncentral case are formulated next.

Bartlett–Corrected Statistics

For the test of exact fit of exploratory factor models, Bartlett (1937, 1950, 1954) developed a small–sample correction of T_{ML}. Fouladi (2000) and Nevitt and Hancock (2004) proposed to employ Bartlett's correction for general covariance structure models as well. Bartlett suggested to multiply T_{ML} by

$$b = 1 - \frac{4k + 2p + 5}{6n} \tag{7}$$

which results in a new test statistic

$$T_{MLb} \equiv bT_{ML}. \quad (8)$$

Equation 7 is a function of the number of latent variables k, the number of observed variables p, and the sample size $N = n + 1$, and it was derived by expansion of a moment generating function (Bartlett, 1950, Equation 3). It follows from Equations 7 and 8 that asymptotically the sampling distribution of the Bartlett–corrected statistic matches that of T_{ML}. However, T_{MLb} more closely follows a chi–square distribution when N is small (Fouladi, 2000; Herzog et al., 2007; Nevitt & Hancock, 2004).

Yuan–Corrected Statistics

As we have emphasized, the Bartlett–correction (Equation 8) is the appropriate small–sample correction for exploratory or unrestricted factor models only. For the test of general covariance structures, Wakaki, Eguchi, and Fujikoshi (1990) developed a Bartlett–like correction that seems to improve the performance of T_{ML} for arbitrary covariance structure models (Kensuke, Takahiro, & Kazuo, 2005). Unfortunately, this correction procedure is quite complicated, even for small models; therefore, Yuan (2005) recommended the "ad hoc correction"

$$T_{MLy} \equiv yT_{ML}, \quad (9)$$

with a modification of the Bartlett correction factor (Equation 7) to

$$y = 1 - \frac{2k + 2p + 7}{6n}. \quad (10)$$

Yuan (2005) argued that an exploratory factor model is identical to a confirmatory factor model when the number of latent variables k equals one. In this case, the Bartlett correction factor (Equation 7) is appropriate and $y = b$. In general for $k > 1$, the exploratory factor model has more factor loadings to be estimated compared to a "usual" covariance structure model with a factor complexity of one (i.e., every measured variable is related to only one factor). As a result, the Bartlett factor in Equation 7 is based on a too–large number of parameters to be estimated when a usual covariance structure model is tested. In Equation 7, k is the only variable that takes into account the number of parameters. Without giving a detailed derivation, Yuan (2005) proposed to employ the constant 2 instead of 4 as a multiplication factor of k in his ad hoc correction of T_{ML}. As a result, every additional latent variable does not decrease the correction factor by $\frac{\partial b}{\partial k} = -\frac{2}{3n}$, as for Bartlett's case, but by $\frac{\partial y}{\partial k} = -\frac{1}{3n}$, thereby taking into account that usually fewer parameters are estimated in "ordinary" covariance structure models compared to exploratory factor models.

In general, $b < y$, and thus $T_{MLb} < T_{MLy}$ as long as $k > 1$. However, from Equations 10 and 7, it is obvious that T_{MLb} and T_{MLy} should perform quite similarly.

Swain–Corrected Statistics

Swain (1975) derived four small–sample corrections of T_{ML} for general covariance structure models. In our study, only the most promising of these four will be considered; see also Browne (1982). Unlike Yuan (2005), who used the k–factor Bartlett correction as a starting point and argued that there are too many parameters t considered in Bartlett's original scaling factor when general covariance structure models are analyzed, Swain (1975) used the test of a fixed covariance structure as a starting point with $t = 0$ and $d = p(p + 1)/2$. For that situation, the appropriate scaling factor is also known (Bartlett, 1954, section IIIa), and Swain proposed several ways to consider that $t > 0$ in general covariance structure models. According to Swain (1975, pp. 78–82), the most promising small–sample correction factor of T_{ML} is defined as

$$s = 1 - \frac{p(2p^2 + 3p - 1) - q(2q^2 + 3q - 1)}{12dn}, \tag{11}$$

where

$$q = \frac{\sqrt{1 + 4p(p + 1) - 8d} - 1}{2}, \tag{12}$$

with p observed variables, d degrees of freedom, and $N = n + 1$ observations. The Swain correction of the test statistic T_{ML} is defined as

$$T_{MLs} \equiv sT_{ML}. \tag{13}$$

Noncentral Case

Remember that the three corrective procedures already presented are not designed for the case of model misspecification and therefore, it is not clear whether they also result in a better approximation of a noncentral chi–square distribution when $H_0 : \mathbf{\Sigma}_0 = \mathbf{\Sigma}(\boldsymbol{\theta}_j)$ does not hold. In earlier research (Steiger et al., 1985), however, Bartlett's statistic was applied for the evaluation of misspecified models without any comment in this matter.

To check whether the rather intuitive application of multiplicative corrections of T_{ML} is legitimate, the behavior of T_{ML}, T_{MLb}, T_{MLy}, and T_{MLs} is studied under different sample size conditions and different degrees of model misspecification in the remainder of this article. More specifically, empirical means and standard deviations of these model fit statistics are compared to the theoretical moments of the noncentral chi–square distribution. Coverage rates of estimated confidence intervals for F_0 or λ resulting from these four statistics are also investigated. Finally, the results are illustrated in the RMSEA metric of F_0 or λ. For these purposes we define

$$\hat{\lambda}_b = T_{MLb} - d, \qquad \hat{\lambda}_y = T_{MLy} - d, \qquad \hat{\lambda}_s = T_{MLs} - d, \tag{14}$$

and

$$\text{RM}\hat{\text{S}}\text{EA}_b = \sqrt{\frac{\hat{\lambda}_b}{dn}}, \qquad \text{RM}\hat{\text{S}}\text{EA}_y = \sqrt{\frac{\hat{\lambda}_y}{dn}}, \qquad \text{RM}\hat{\text{S}}\text{EA}_s = \sqrt{\frac{\hat{\lambda}_s}{dn}}. \tag{15}$$

Expectations

What are our expectations regarding the small–sample performance of the statistics under study? In accordance with the results of previous studies, we first expect for small sample sizes (in particular for $N < 200$) that both mean value and standard deviation of T_{ML} are too large, resulting in too liberal estimates of λ (i.e., λ being overestimated) and in coverage rates below nominal values (cf. Curran et al., 2003; Curran et al., 2002; Olsson et al., 2004).

Second, all corrective procedures are expected to give less liberal estimates of λ compared to the traditional procedure because all correction factors are smaller than one for finite sample sizes. In some cases, they might even be too conservative (i.e., λ being underestimated). It is therefore hard to predict whether the corrections under study are less biased compared to T_{ML}.

Third, as discussed earlier, it is clear that Bartlett's procedure is more conservative than Yuan's procedure as long as $k > 1$, but the difference between the two correction factors is quite small. From the results of earlier empirical studies, it can further be concluded that Swain's procedure is the least corrective and thus the least conservative statistic of all corrective procedures (Fouladi, 2000; Herzog et al., 2007). Hence, it is expected that $T_{MLb} < T_{MLy} < T_{MLs} < T_{ML}$, and that mean values and standard deviations are the smallest for T_{MLb} and the largest for T_{ML}. It is not clear though, which statistic will have the best approximation to a noncentral chi–square distribution. Furthermore, it is not straightforward which statistic produces the best estimates of λ, RMSEA, and the corresponding confidence intervals. The only hint in the literature is that T_{MLb} might be a bit too conservative, or to be more specific, the moments of the corresponding noncentral chi–square distribution might be underestimated when sample sizes are small and F_0 is rather large (Steiger et al., 1985). It is therefore possible that T_{MLy} and T_{MLs} perform better than T_{MLb} and T_{ML} for small $N{:}t$ ratios.

Finally, it is known that mean and standard deviation bias of T_{ML} slightly reduce with increasing F_0 (Curran et al., 2003; Curran et al., 2002). However, when F_0 increases much, the standard deviation bias inflates extremely but the mean bias seems unaffected (Curran et al., 2002; Olsson et al., 2004). These tendencies are expected for all statistics under study.

Unfortunately, to our knowledge, no more background information is available—neither from asymptotic theory nor from previous Monte Carlo studies—that could help to formulate more specific expectations. The following simulation work is designed to provide more insights into the small–sample behavior of the statistics under study.

MONTE CARLO DESIGN

Sample Size Conditions

Data sets with sample sizes of 50, 75, 100, 150, and 200 were generated to analyze the small–sample behavior of the statistics under investigation.

Population Model and Misspecifications

The population model M_0 is depicted in Figure 1. It is a confirmatory factor model with four latent factors and 24 observed variables. Solid arrows represent regular loadings of 0.70, dashed

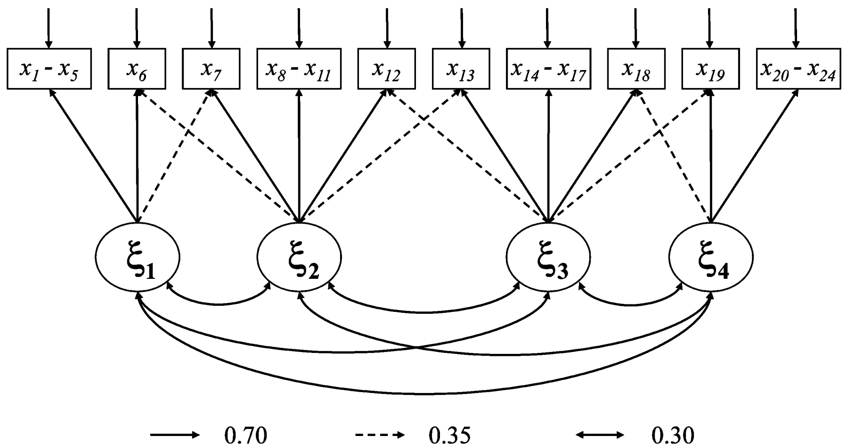


FIGURE 1 Population model M_0.

arrows denote cross–loadings of 0.35, and curved double–sided arrows stand for correlations among latent factors of 0.30. Variances of the latent factors were fixed to the value of 1. These population values were selected in accordance with a similar model generated by Curran, West, and Finch (1996) and Nevitt and Hancock (2001). As usual, error variances were determined so that all observed variable variances equal 1 (e.g., Olsson, Foss, Troye, & Howell, 2000). That is, the error variances were fixed to 0.2405 for $x_6, x_7, x_{12}, x_{13}, x_{18}$, and x_{19}, and to 0.5100 in all other cases.

For all sample size conditions, five models (M_0–M_4) were analyzed with an increasing degree of misspecification, as indicated by population values of RMSEA ranging from 0.000 up to 0.096 (for details see Table 1). This range was selected because it represents the variety of RMSEA values usually reported in applied studies. The value of 0.000 reflects perfect model fit, and a value close to 0.100 indicates very poor model fit that might be acceptable only under very specific conditions (e.g., as in the case of Homburg & Fürst, 2005).

TABLE 1
Overview of Models in the Monte Carlo Design

Model	*t*	*d*	F_0	*RMSEA*	*TLI*	*Model Error*
M_0	60	240	0.000	0.000	1.000	None
M_1	59	241	0.241	0.032	0.978	$\xi_4 \rightarrow x_{18}$ omitted
M_2	58	242	0.480	0.045	0.956	$\xi_4 \rightarrow x_{18}$ and $\xi_3 \rightarrow x_{19}$ omitted
M_3	54	246	1.349	0.074	0.877	All cross–loadings omitted
M_4	54	246	2.264	0.096	0.794	All cross–loadings omitted, but: ξ_1 loads on x_7, but ξ_2 does not; ξ_2 loads on x_{13}, but ξ_3 does not; ξ_3 loads on x_{19}, but ξ_4 does not
M_5	24	276	12.330	0.211	0.000	Independence model

Note. RMSEA = root mean square error of approximation; TLI = Tucker–Lewis Index.

Estimators of Model Fit

The third main factor in this study crossing the 25 sample size × model specification conditions is the model fit estimator with four levels: T_{ML}, T_{MLb}, T_{MLy}, and T_{MLs}. As a result, the experimental setup of the Monte Carlo study is a $5 \times 5 \times 4$ factorial design with 100 conditions. The response variables are described in the performance criteria section.

Data Generation and Model Estimation

Samples of size N were generated from a (24–variate) multinormal distribution with covariance structure $\mathbf{\Sigma}(\boldsymbol{\theta}_0)$ resulting from the population model M_0 (see Table 1 and Figure 1). Both generation of the sample data and estimation of the models was performed by using the M*plus* software program (Version 4.2; Muthén & Muthén, 2007). The seed values for the pseudo–random sample draws from the multivariate normal population distribution for the five sample size conditions were 71398311 ($N = 50$), 48781591 ($N = 75$), 39528719 ($N = 100$), 09187419 ($N = 150$), and 61846648 ($N = 200$).

The starting values for the model parameter estimates were set to their population values, and factor variances were fixed to 1.00 for reasons of identification. The factor models were estimated using the primary maximum likelihood estimation setting (ML) in M*plus*. For statistical analyses of the generated model estimates, R software (Version 2.4.1) was used; see, for example, Venables, Smith, and the R Development Core Team (2006).

Number of Replications

For each of the five sample size conditions, 10,000 data sets were generated from population model M_0 shown in Figure 1. Following Olsson et al. (2004), Curran et al. (2002), and Curran et al. (2003), additional data were generated for the rare cases in which improper solutions occurred (a maximum rate of 1.86% for model M_3 with $N = 50$). There were no problems with convergence during the iterative estimation process, which might be due to the large number of indicators per factor that has been shown to reduce problems with nonconvergence and improper solutions (Boomsma, 1985; Marsh et al., 1998).

Performance Criteria

A number of performance criteria are important when evaluating the finite–sample behavior of the model test statistics under study.

First, the relative mean bias (i.e., the mean of a statistic computed across all replications divided by its asymptotically predicted mean), the relative standard deviation bias (i.e., the standard deviation of a statistic computed across all replications divided by its asymptotically predicted standard deviation), and the coverage rates of the estimated confidence intervals for λ are studied. Across all replications, the coverage rate should be equal to $1 - \alpha$, and in our study we investigate the widely applied case where $1 - \alpha = 0.90$. Due to the large number of replications, we did not employ significance tests for the deviations of empirical means, standard deviations, and coverage rates from their asymptotically predicted values. From the derivations of Steiger et al. (1985), it is clear that for every $F_0 > 0$, no test statistic will exactly follow a noncentral chi–square distribution. This also holds for the finite–sample case: Asymptotic theory predicts that for $N < \infty$, test statistics will not follow a central or a noncentral chi–square distribution exactly. Thus, significance of deviations from theoretical values mainly depends on the number of replications. In this study, the primary objective therefore is to determine the practical significance of the deviations. The number of replications was set to 10,000 to get estimates of these deviations close to the population deviations. Concordant with Curran et al. (2002), acceptable values of relative mean bias and relative standard deviation bias are defined by the range [0.95, 1.05]. Similarly, acceptable coverage rates are defined by values between 0.88 and 0.92 for a 90% confidence interval (Curran et al., 2003). In the tables with the summarized results (see next section), acceptable relative mean biases, relative standard deviation biases, and coverage rates are printed in boldface.

Second, means and standard deviations of RMSEA estimators are reported to illustrate the practical relevance of our findings in this widely applied metric of the noncentrality parameter. We also report on the root mean squared error (RMSE) of the four RMSEA estimators (see Equations 4 and 15). The RMSE of an estimator equals the square rooted sum of its squared bias and its variance. For each cross–condition of sample size × degree of misspecification, the best performing RMSEA estimate (i.e., the one with the lowest RMSE) is highlighted by printing mean, standard deviation, and RMSE in boldface. When two or more RMSEA estimators have the lowest RMSE (equal to the third decimal place), mean, standard deviation, and RMSE of these estimators are printed in boldface.

RESULTS

In Tables 2 through 5, the results for T_{ML}, T_{MLb}, T_{MLy}, and T_{MLs} are reported in terms of the performance criteria described in the previous section.

Traditional Estimator

For T_{ML}, the relative mean bias reduces with increasing sample size and increasing degree of misspecification, as expected. Both findings cross–validate results reported by Curran et al. (2002). The mean of T_{ML} across the 10,000 replications is 30% larger compared to its expected value for RMSEA $= 0$ and $N = 50$; even for $N = 150$ and $N = 200$, the mean of T_{ML} is larger than its theoretical value. The relative standard deviation bias reduces with increasing sample size. It also decreases with increasing misspecification but this tendency is weaker and occasionally not monotone when N increases. The coverage rates associated with the traditional estimation method are not acceptable for any condition under study. In particular for RMSEA $= 0$ and $N = 50$, only 11% instead of 90% of the confidence intervals cover λ, which is quite an alarming result for applied researchers.

In RMSEA metric, the results show that the traditional method is too liberal, with the apparent consequence that acceptable models estimated with small sample sizes are rejected too frequently. The standard deviation of $\widehat{\text{RMSEA}}$ decreases monotone with increasing sample size and increasing misspecification. $\widehat{\text{RMSEA}}$ is among the best estimators compared to the other three estimators ($\widehat{\text{RMSEA}}_b$, $\widehat{\text{RMSEA}}_y$, and $\widehat{\text{RMSEA}}_s$) in terms of RMSE only for $N = 150$ and $N = 200$.

TABLE 2
Performance of T_{ML}

	Sample Size				
RMSEA	*N* = *50* *N:t* ≈ *0.9*	*N* = *75* *N:t* ≈ *1.3*	*N* = *100* *N:t* ≈ *1.7*	*N* = *150* *N:t* ≈ *2.6*	*N* = *200* *N:t* ≈ *3.4*
Relative mean bias[a]					
0.000	1.30	1.18	1.12	1.08	1.06
0.032	1.28	1.17	1.11	1.07	**1.05**
0.045	1.27	1.15	1.10	**1.05**	**1.03**
0.074	1.23	1.12	1.08	**1.04**	**1.02**
0.096	1.20	1.11	1.06	**1.03**	**1.02**
Relative standard deviation bias[b]					
0.000	1.32	1.17	1.12	1.07	**1.05**
0.032	1.29	1.16	1.10	**1.05**	**1.04**
0.045	1.27	1.15	1.09	**1.05**	**1.05**
0.074	1.22	1.12	1.08	**1.05**	**1.05**
0.096	1.21	1.14	1.11	1.09	1.09
Coverage rate for λ (90% confidence interval)[c]					
0.000	0.11	0.43	0.62	0.78	0.84
0.032	0.13	0.46	0.66	0.81	0.85
0.045	0.16	0.51	0.71	0.83	0.86
0.074	0.22	0.61	0.74	0.85	0.87
0.096	0.29	0.65	0.77	0.84	0.86
M (*SD*) [root mean squared error] of $\widehat{RMSEA}$[d]					
0.000	0.076	0.046	0.032	0.020	0.015
	(0.017)	(0.017)	(0.016)	(0.014)	(0.012)
	[0.078]	[0.049]	[0.036]	[0.024]	[0.019]
0.032	0.083	0.056	0.046	**0.037**	**0.035**
	(0.016)	(0.015)	(0.013)	(**0.011**)	(**0.009**)
	[0.053]	[0.029]	[0.019]	[**0.012**]	[**0.009**]
0.045	0.088	0.064	0.055	**0.049**	**0.047**
	(0.015)	(0.013)	(0.011)	(**0.009**)	(**0.007**)
	[0.046]	[0.024]	[0.016]	[**0.010**]	[**0.007**]
0.074	0.106	0.088	0.081	**0.077**	**0.075**
	(0.013)	(0.011)	(0.009)	(**0.007**)	(**0.006**)
	[0.035]	[0.017]	[0.011]	[**0.007**]	[**0.006**]
0.096	0.123	0.107	0.102	**0.098**	**0.097**
	(0.013)	(0.010)	(0.008)	(**0.006**)	(**0.005**)
	[0.030]	[0.015]	[0.010]	[**0.007**]	[**0.006**]

Note. RMSEA = root mean square error of approximation.

[a]Values in the range [0.95, 1.05] are defined as acceptable and printed in boldface. [b]Values in the range [0.95, 1.05] are defined as acceptable and printed in boldface. [c]Values in the range [0.88, 0.92] are defined as acceptable and printed in boldface. [d]For each RMSEA × sample size condition, mean values (M) are printed without parentheses, standard deviations (SD) in parentheses, and root mean squared errors in brackets. For each RMSEA × sample size condition, these three values are printed in boldface when no other estimator reported in Tables 3 through 5 has a smaller root mean squared error.

TABLE 3
Performance of T_{MLb}

	Sample Size				
RMSEA	*N = 50* *N:t ≈ 0.9*	*N = 75* *N:t ≈ 1.3*	*N = 100* *N:t ≈ 1.7*	*N = 150* *N:t ≈ 2.6*	*N = 200* *N:t ≈ 3.4*
Relative mean bias[a]					
0.000	**0.99**	**0.99**	**0.99**	**0.99**	**1.00**
0.032	**0.98**	**0.98**	**0.98**	**0.98**	**0.99**
0.045	**0.97**	**0.97**	**0.97**	**0.97**	**0.97**
0.074	0.94	**0.95**	**0.95**	**0.96**	**0.96**
0.096	0.92	0.93	0.94	**0.95**	**0.96**
Relative standard deviation bias[b]					
0.000	**1.00**	**0.99**	**0.99**	**0.99**	**0.99**
0.032	**0.99**	**0.98**	**0.97**	**0.97**	**0.98**
0.045	**0.97**	**0.97**	**0.96**	**0.97**	**0.99**
0.074	0.94	0.94	**0.96**	**0.97**	**1.00**
0.096	0.92	**0.96**	**0.98**	**1.00**	**1.02**
Coverage rate for λ (90% confidence interval)[c]					
0.000	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**
0.032	**0.89**	**0.90**	**0.90**	**0.90**	**0.90**
0.045	**0.88**	**0.89**	**0.89**	**0.89**	**0.89**
0.074	0.84	0.84	0.85	0.86	0.86
0.096	0.76	0.79	0.80	0.82	0.83
M (*SD*) [root mean squared error] of $\widehat{\text{RMSEA}}_b$[d]					
0.000	**0.016**	**0.013**	**0.011**	**0.009**	**0.008**
	(**0.020**)	(**0.017**)	(**0.014**)	(**0.012**)	(**0.010**)
	[**0.026**]	[**0.021**]	[**0.018**]	[**0.015**]	[**0.013**]
0.032	**0.024**	0.025	0.025	0.027	0.029
	(**0.023**)	(0.019)	(0.017)	(0.013)	(0.010)
	[**0.024**]	[0.020]	[0.018]	[0.014]	[0.011]
0.045	0.032	0.036	0.038	0.041	0.042
	(0.024)	(0.019)	(0.015)	(0.010)	(0.008)
	[0.027]	[0.021]	[0.016]	[0.011]	[0.008]
0.074	0.060	0.066	0.068	0.070	**0.071**
	(0.020)	(0.012)	(0.009)	(0.007)	(**0.006**)
	[0.024]	[0.015]	[0.011]	[0.008]	[**0.006**]
0.096	0.081	0.087	0.089	0.092	**0.093**
	(0.015)	(0.010)	(0.008)	(0.006)	(**0.005**)
	[0.021]	[0.014]	[0.011]	[0.008]	[**0.006**]

Note. RMSEA = root mean square error of approximation.

[a]Values in the range [0.95, 1.05] are defined as acceptable and printed in boldface. [b]Values in the range [0.95, 1.05] are defined as acceptable and printed in boldface. [c]Values in the range [0.88, 0.92] are defined as acceptable and printed in boldface. [d]For each RMSEA × sample size condition, mean values (M) are printed without parentheses, standard deviations (SD) in parentheses, and root mean squared errors in brackets. For each RMSEA × sample size condition, these three values are printed in boldface when no other estimator reported in Tables 2, 4, and 5 has a smaller root mean squared error.

TABLE 4
Performance of T_{MLy}

	Sample Size				
RMSEA	*N = 50* *N:t ≈ 0.9*	*N = 75* *N:t ≈ 1.3*	*N = 100* *N:t ≈ 1.7*	*N = 150* *N:t ≈ 2.6*	*N = 200* *N:t ≈ 3.4*
Relative mean bias[a]					
0.000	**1.02**	**1.01**	**1.00**	**1.00**	**1.00**
0.032	**1.01**	**1.00**	**0.99**	**0.99**	**0.99**
0.045	**1.00**	**0.99**	**0.98**	**0.98**	**0.98**
0.074	**0.97**	**0.96**	**0.96**	**0.97**	**0.97**
0.096	0.94	**0.95**	**0.95**	**0.96**	**0.97**
Relative standard deviation bias[b]					
0.000	**1.03**	**1.01**	**1.01**	**1.01**	**0.99**
0.032	**1.01**	**0.99**	**0.98**	**0.98**	**0.99**
0.045	**1.00**	**0.99**	**0.97**	**0.98**	**0.99**
0.074	**0.96**	**0.96**	**0.97**	**0.98**	**1.00**
0.096	**0.95**	**0.98**	**0.99**	**1.01**	**1.03**
Coverage rate for λ (90% confidence interval)[c]					
0.000	**0.89**	**0.90**	**0.90**	**0.90**	**0.90**
0.032	**0.89**	**0.90**	**0.90**	**0.89**	**0.90**
0.045	**0.90**	**0.90**	**0.90**	**0.89**	**0.89**
0.074	**0.89**	0.87	0.87	0.87	0.87
0.096	0.83	0.83	0.83	0.84	0.85
M (*SD*) [root mean squared error] of $\widehat{RMSEA}_y$[d]					
0.000	0.022	0.016	0.013	0.010	0.009
	(0.022)	(0.018)	(0.015)	(0.012)	(0.010)
	[0.031]	[0.024]	[0.020]	[0.016]	[0.014]
0.032	**0.031**	0.028	**0.027**	0.028	0.029
	(**0.024**)	(0.019)	(**0.017**)	(0.013)	(0.010)
	[**0.024**]	[0.020]	[**0.017**]	[0.013]	[0.010]
0.045	0.039	0.039	0.040	**0.041**	0.042
	(0.024)	(0.018)	(0.015)	(**0.010**)	(0.007)
	[0.024]	[0.019]	[0.015]	[**0.010**]	[0.008]
0.074	0.065	0.068	**0.069**	0.071	**0.072**
	(0.018)	(0.012)	(**0.009**)	(0.007)	(**0.006**)
	[0.020]	[0.013]	[**0.010**]	[0.008]	[**0.006**]
0.096	0.086	0.089	0.091	**0.092**	**0.093**
	(0.014)	(0.010)	(0.008)	(**0.006**)	(**0.005**)
	[0.018]	[0.012]	[0.010]	[**0.007**]	[**0.006**]

Note. RMSEA = root mean square error of approximation.

[a]Values in the range [0.95, 1.05] are defined as acceptable and printed in boldface. [b]Values in the range [0.95, 1.05] are defined as acceptable and printed in boldface. [c]Values in the range [0.88, 0.92] are defined as acceptable and printed in boldface. [d]For each RMSEA × sample size condition, mean values (M) are printed without parentheses, standard deviations (SD) in parentheses, and root mean squared errors in brackets. For each RMSEA × sample size condition, these three values are printed in boldface when no other estimator reported in Tables 2, 3, and 5 has a smaller root mean squared error.

TABLE 5
Performance of T_{MLs}

	Sample Size				
RMSEA	*N = 50* *N:t ≈ 0.9*	*N = 75* *N:t ≈ 1.3*	*N = 100* *N:t ≈ 1.7*	*N = 150* *N:t ≈ 2.6*	*N = 200* *N:t ≈ 3.4*
Relative mean bias[a]					
0.000	1.06	**1.03**	**1.02**	**1.01**	**1.01**
0.032	**1.05**	**1.01**	**1.01**	**1.00**	**1.00**
0.045	**1.03**	**1.00**	**1.00**	**0.99**	**0.99**
0.074	**1.00**	**0.98**	**0.98**	**0.98**	**0.98**
0.096	**0.98**	**0.97**	**0.97**	**0.97**	**0.97**
Relative standard deviation bias[b]					
0.000	1.07	**1.03**	**1.02**	**1.01**	**1.00**
0.032	**1.05**	**1.01**	**1.00**	**0.99**	**0.99**
0.045	**1.04**	**1.01**	**0.99**	**0.99**	**1.00**
0.074	**1.00**	**0.98**	**0.98**	**0.99**	**1.01**
0.096	**0.99**	**1.00**	**1.01**	**1.02**	**1.04**
Coverage rate for λ (90% confidence interval)[c]					
0.000	0.83	**0.88**	**0.89**	**0.90**	**0.90**
0.032	0.86	**0.90**	**0.90**	**0.90**	**0.90**
0.045	**0.88**	**0.90**	**0.90**	**0.90**	**0.90**
0.074	**0.90**	**0.90**	**0.89**	**0.89**	**0.88**
0.096	**0.89**	**0.88**	**0.88**	**0.88**	**0.88**
M (*SD*) [root mean squared error] of $\text{RM}\hat{\text{S}}\text{EA}_s$[d]					
0.000	0.031	0.020	0.015	0.011	0.010
	(0.024)	(0.019)	(0.016)	(0.012)	(0.011)
	[0.039]	[0.027]	[0.022]	[0.017]	[0.014]
0.032	0.040	**0.032**	**0.030**	0.029	0.030
	(0.024)	**(0.019)**	**(0.016)**	(0.012)	(0.010)
	[0.025]	**[0.019]**	**[0.017]**	[0.013]	[0.010]
0.045	**0.048**	**0.043**	**0.042**	**0.042**	0.043
	(0.023)	**(0.018)**	**(0.014)**	**(0.010)**	(0.007)
	[0.023]	**[0.018]**	**[0.014]**	**[0.010]**	[0.008]
0.074	**0.073**	**0.071**	**0.071**	**0.072**	**0.072**
	(0.017)	**(0.012)**	**(0.009)**	**(0.007)**	**(0.006)**
	[0.017]	**[0.012]**	**[0.010]**	**[0.007]**	**[0.006]**
0.096	**0.092**	**0.092**	**0.092**	**0.093**	**0.094**
	(0.014)	**(0.010)**	**(0.008)**	**(0.006)**	**(0.005)**
	[0.014]	**[0.011]**	**[0.009]**	**[0.007]**	**[0.006]**

Note. RMSEA = root mean square error of approximation.

[a]Values in the range [0.95, 1.05] are defined as acceptable and printed in boldface. [b]Values in the range [0.95, 1.05] are defined as acceptable and printed in boldface. [c]Values in the range [0.88, 0.92] are defined as acceptable and printed in boldface. [d]For each RMSEA × sample size condition, mean values (M) are printed without parentheses, standard deviations (SD) in parentheses, and root mean squared errors in brackets. For each RMSEA × sample size condition, these three values are printed in boldface when no other estimator reported in Tables 2 through 4 has a smaller root mean squared error.

In summary, the results for T_{ML} revalidate findings of Curran et al. (2002), Curran et al. (2003), Olsson et al. (2004), and Hu and Bentler (1999): Traditional estimators of noncentrality–based population model fit like $\hat{\text{RMSEA}}$ should not be applied when the sample size is small ($N \leq 200$).

Bartlett's Estimator

In contrast to T_{ML}, the relative mean bias of T_{MLb} seems to be quite stable with increasing sample size. However, it decreases with the degree of misspecification. More specifically, the mean of T_{MLb} is lower compared to its population counterpart when misspecification increases, and especially with small sample sizes. This result is consistent with findings of Steiger et al. (1985). T_{MLb} performs much better compared to T_{ML} in terms of relative mean bias, especially for the population model M_0, where RMSEA $= 0$. The relative standard deviation bias is rather stable when sample size increases. Like for the mean, the standard deviation of T_{MLb} underestimates its population value with increasing misspecification, particularly for small sample sizes. The coverage rates are much better compared to those based on T_{ML}, but they clearly decrease below the nominal 90% coverage of λ when misspecification increases.

The estimator $\hat{\text{RMSEA}}_b$ has the smallest RMSE when RMSEA $= 0$. This result, however, should not be interpreted in favor of T_{MLb}: A correction factor of 0 would have been the best option for RMSEA $= 0$, because the lower bound of λ is constrained to zero. T_{MLb} is the most conservative correction and therefore T_{MLb} is by definition the best statistic when RMSEA $= 0$. For RMSEA > 0, however, $\hat{\text{RMSEA}}_b$ clearly underestimates the population value when N is small; that is, under realistic conditions ($F_0 > 0$) Bartlett's procedure is too conservative. This undesirable tendency increases with the degree of misspecification and therefore, the power of $\hat{\text{RMSEA}}_b$ to reject an unacceptable model is clearly lower compared to that of $\hat{\text{RMSEA}}$.

In summary, the performance of T_{MLb} is better compared to T_{ML}, but this advantage is clearly at the cost of decreased power to reject unacceptable models, especially when the sample size is small. This result is in accordance with findings of decreased power of T_{MLb} in Nevitt and Hancock's (2004) study. It is further consistent with results reported by Herzog et al. (2007), who observed that T_{MLb} underestimates Type I error rates when $N{:}t$ decreases. We consider this property of T_{MLb} as problematic, hence we hesitate to recommend this statistic for applied research.

Yuan's Estimator

As expected, the behavior of T_{MLy} is very similar to that of T_{MLb}. The relative mean bias is quite stable with increasing sample size. Although the performance of T_{MLy} is somewhat better than that of T_{MLb}, with increasing misspecification there is still a trend to underestimate the population mean when N is small. The standard deviation of T_{MLy} is somewhat closer to the expected value compared to T_{MLb}. However, as for T_{MLb}, the coverage rates for T_{MLy} are acceptable for minor misspecifications only and become inadequate with increasing model misspecification.

As for $\hat{\text{RMSEA}}_b$, $\hat{\text{RMSEA}}_y$ has the drawback that population values are underestimated for small sample sizes and increasing misspecification. As a result, its power to reject an unacceptable model is lower compared to the traditional estimator $\hat{\text{RMSEA}}$ when the sample size is small. $\text{RM}\hat{\text{S}}\text{EA}_y$ is the estimator with the smallest RMSE among the four estimators under study in 7 out of 25 conditions.

In summary, the performance of T_{MLy} is somewhat better compared to that of T_{MLb}, but there is not too much of a difference. We are of the opinion that more empirical work on this estimator is required before applied researchers should be advised to use it.

Swain's Estimator

Compared to T_{ML}, the relative mean and standard deviation bias of T_{MLs} is quite stable across all sample size conditions. A main result is that, compared to T_{MLb} and T_{MLy}, T_{MLs} gives a good approximation of the corresponding noncentral chi–square distributions in terms of mean and standard deviation, even when the degree of misspecification increases. This result is reflected in nearly perfect coverage rates as long as N is at least 75.

The estimates $\text{RM}\hat{\text{S}}\text{EA}_s$ are close to the corresponding population values across all sample size and model specification conditions. In particular, $\text{RM}\hat{\text{S}}\text{EA}_s$ does not underestimate the population values severely when N is small and thus does not suffer from reduced power to reject unacceptable models in such situations. As a result, $\text{RM}\hat{\text{S}}\text{EA}_s$ is the estimator with the smallest RMSE in 16 out of 25 conditions in our Monte Carlo design.

Conclusion

T_{ML} is much too liberal when N is small and therefore, acceptable models are rejected too frequently. In contrast, T_{MLb} and T_{MLy} follow noncentral chi–square distributions more closely for minor model misspecifications, but they tend to suffer from decreased power to reject misspecified models. T_{MLs}, on the other hand, does not reject models too frequently when small samples are analyzed (unlike T_{ML}), and it has enough power to reject misspecified models (unlike T_{MLb} and T_{MLy}). T_{MLs} is clearly the most stable statistic across all sample size conditions and degrees of specification error under study. Its use is therefore recommended for applied research when model inference is based on the noncentral chi–square distribution and functions of the noncentrality parameter are used to evaluate model fit. It is emphasized again that our findings and recommendations not only hold for RMSEA estimation, but also for the estimation of other noncentrality–based fit indexes (e.g., Steiger's γ, McDonald's Centrality Index, etc.).

SUPPLEMENTARY ANALYSIS: INCREMENTAL FIT INDEXES

The results presented so far lead to the recommendation to use the Swain–corrected statistic T_{MLs} instead of T_{ML} for inferences based on the noncentral chi–square distribution when sample size is small. Our findings, however, are not informative regarding incremental fit indexes—another family of chi–square–dependent fit indexes, most often reported in combination with noncentrality–based fit statistics. Because incremental fit indexes are functions of chi–square measures of model fit, the question under study is whether they also need corrections when sample size is small.

Definitions

Incremental fit indexes were developed to quantify the increment of fit of a hypothesized model M_j relative to a baseline model. The choice of the baseline model is not indisputable (see e.g., Sobel & Bohrnstedt, 1985). Most, if not all, software packages, however, take the so–called independence model, M_i, as the baseline, a model where only variances of the (supposedly uncorrelated) observed variables are estimated. Given that the chi–square value of the target model M_j should be Swain–corrected (see previous section), it is unclear now whether one should also use Swain's correction factor for the chi–square value of the independence model M_i.

The Tucker–Lewis Index (TLI), for example, is defined as (Gerbing & Anderson, 1993, p. 56; McDonald & Marsh, 1990, Equations 7 and 18)

$$\text{TLI} = \frac{F_i/d_i - F_j/d_j}{F_i/d_i}, \tag{16}$$

where F_j and F_i are the errors of approximation of the target model and the independence model, respectively, and d_j and d_i are the corresponding degrees of freedom. The value of TLI equals zero when the target model fits as bad as the independence model. For $F_j = 0$, TLI equals one, but sample fluctuations may allow estimates of TLI to be larger than one in some cases. TLI can be estimated by

$$\hat{\text{TLI}} = \frac{T_i/d_i - T_j/d_j}{T_i/d_i - 1}, \tag{17}$$

where T_i and T_j are the likelihood ratio test statistics for the independence model M_i and the target model M_j, respectively, with corresponding degrees of freedom d_i and d_j.

We extend our study presented so far to the independence model to investigate whether Swain's correction also improves the small–sample behavior of Equation 17. Two options are considered. The first option is to correct only the target model fit statistic leading to the estimator

$$\hat{\text{TLI}}_{s1} = \frac{T_i/d_i - s_j T_j/d_j}{T_i/d_i - 1}, \tag{18}$$

with s_j being Swain's correction factor for the target model M_j. The second option is to correct both the target and the independence model fit statistic, resulting in the estimator

$$\hat{\text{TLI}}_{s2} = \frac{s_i T_i/d_i - s_j T_j/d_j}{s_i T_i/d_i - 1}, \tag{19}$$

where s_i is Swain's correction factor for the independence model M_i.

Expectations

The specification of the independence model in the simulation study was presented earlier in Table 1 (see model M_5). The study by Curran et al. (2002) revealed that the mean of T_i corresponds closely to the expected value of a noncentral chi–square distribution, but the variance of T_i is inflated severely, so that confidence intervals for the noncentrality parameter λ are in general not trustworthy. Based on this finding, we expect that it is better not to compute the Swain correction of T_i because this would result in an empirical mean smaller than the expected value of the corresponding noncentral chi–square distribution. The standard deviation of T_i and coverage rates based on T_i are expected to be unreliable with or without Swain's correction.

Results and Conclusions

The results for the independence model are summarized in Table 6. They are in accordance with our expectations.

Our findings are further illustrated in TLI metric (Tables 7 through 9). As before, for every cross–condition of sample size × model specification, mean values, standard deviations, and RMSEs are printed in boldface for the estimator having the smallest RMSE among the three estimators under study ($\hat{TLI}$, $\hat{TLI}_{s1}$, and $\hat{TLI}_{s2}$). When two or more TLI estimators have the smallest RMSE (equal to the third decimal place), mean, standard deviation, and RMSE of these estimators are printed in boldface.

The traditional estimator $\hat{TLI}$ performs best in terms of RMSE in only 1 out of 25 conditions. It underestimates the population value for small sample sizes. Both $\hat{TLI}_{s1}$ and $\hat{TLI}_{s2}$ perform much better, but the advantage in terms of RMSE is clearly on the side of $\hat{TLI}_{s1}$, as expected. Based on these results, we recommend correcting only the target model fit statistic using Swain's multiplier when population values of incremental fit indexes are estimated—the fit statistic of the independence model should not be corrected. Hence, $\hat{TLI}_{s1}$ should be applied in practice. Again, it should be noted that our recommendations not only hold for the estimation of TLI, but also for the estimation of other incremental fit indexes like CFI.

TABLE 6
Results for Independence Model

	Sample Size				
Statistic	*N = 50* *N:t ≈ 2.1*	*N = 75* *N:t ≈ 3.1*	*N = 100* *N:t ≈ 4.2*	*N = 150* *N:t ≈ 6.3*	*N = 200* *N:t ≈ 8.3*
Relative mean bias[a]					
T_{ML}	1.10	**1.04**	**1.03**	**1.02**	**1.01**
T_{MLs}	0.91	0.92	0.94	**0.95**	**0.96**
Relative standard deviation bias[b]					
T_{ML}	2.26	2.32	2.35	2.33	2.36
T_{MLs}	1.54	1.82	1.96	2.07	2.16
Coverage rate for λ (90% confidence interval)[c]					
T_{ML}	0.51	0.67	0.70	0.71	0.71
T_{MLs}	0.53	0.55	0.57	0.60	0.63

[a]Values in the range [0.95, 1.05] are defined as acceptable and printed in boldface. [b]Values in the range [0.95, 1.05] are defined as acceptable and printed in boldface. [c]Values in the range [0.88, 0.92] are defined as acceptable and printed in boldface.

TABLE 7
Performance of $\hat{TLI}$

	Sample Size				
TLI	*N = 50* *N:t ≈ 0.9*	*N = 75* *N:t ≈ 1.3*	*N = 100* *N:t ≈ 1.7*	*N = 150* *N:t ≈ 2.6*	*N = 200* *N:t ≈ 3.4*
M (*SD*) [root mean squared error]					
1.000	0.881	0.950	0.973	0.989	0.994
	(0.046)	(0.030)	(0.022)	(0.014)	(0.011)
	[0.128]	[0.059]	[0.035]	[0.018]	[0.012]
0.978	0.862	0.928	0.951	0.967	0.972
	(0.047)	(0.031)	(0.024)	(0.016)	(0.012)
	[0.125]	[0.058]	[0.035]	[0.019]	[0.014]
0.956	0.844	0.909	0.931	0.946	0.951
	(0.047)	(0.033)	(0.025)	(0.017)	(0.014)
	[0.122]	[0.057]	[0.035]	[0.020]	[0.015]
0.877	0.776	0.835	0.855	0.868	**0.873**
	(0.050)	(0.036)	(0.029)	(0.021)	(**0.018**)
	[0.113]	[0.056]	[0.036]	[0.023]	[**0.018**]
0.794	0.701	0.754	0.773	0.785	0.789
	(0.054)	(0.042)	(0.034)	(0.026)	(0.022)
	[0.107]	[0.058]	[0.041]	[0.028]	[0.023]

Note. TLI = Tucker–Lewis Index. For each TLI × sample size condition, mean values (M) are printed without parentheses, standard deviations (SD) in parentheses, and root mean squared errors in brackets. For each TLI × sample size condition, these three values are printed in boldface when no other estimator reported in Tables 8 and 9 has a smaller root mean squared error.

ILLUSTRATION

To illustrate the relevance of our findings for applied research under conditions of small sample sizes, we corrected noncentrality–based fit statistics and incremental fit indexes of a covariance structure model analyzed by Obermiller, Spangenberg, and MacLachlan (2005, Figure 2). The model was specified with $p = 25$ and $d = 265$, and it was estimated with an extremely small sample size of $N = 54$. In Table 3, the authors further reported that $T_{ML} = 398.291$, $\hat{RMSEA} = 0.097$, and $\hat{CFI} = 0.838$ (for a definition of the CFI, see Hu & Bentler, 1999, Table 1). The authors concluded that their "model does not fit the data well, which merely suggests that the model is underspecified" (p. 14). The question now is whether the authors would have reached another conclusion if they had applied our Swain–corrected estimators of noncentrality–based and incremental model fit measures.

Given the described model specification and the sample size, Swain's scaling factor equals 0.819, hence test statistic $T_{MLs} = 326.240$. The resulting Swain–corrected point estimate $\hat{RMSEA}_s = 0.066$ with a 90% confidence interval [0.037, 0.089] is substantially smaller compared to the traditional estimate $\hat{RMSEA} = 0.097$ with a confidence interval [0.077, 0.117].

TABLE 8
Performance of $\hat{TLI}_{s1}$

	Sample Size				
TLI	*N = 50* *N:t ≈ 0.9*	*N = 75* *N:t ≈ 1.3*	*N = 100* *N:t ≈ 1.7*	*N = 150* *N:t ≈ 2.6*	*N = 200* *N:t ≈ 3.4*
M (*SD*) [root mean squared error]					
1.000	**0.978**	**0.992**	**0.996**	**0.999**	**0.999**
	(**0.039**)	(**0.027**)	(**0.021**)	(**0.014**)	(**0.010**)
	[**0.045**]	[**0.028**]	[**0.021**]	[**0.014**]	[**0.010**]
0.978	**0.963**	**0.973**	**0.976**	**0.978**	**0.978**
	(**0.039**)	(**0.028**)	(**0.022**)	(**0.015**)	(**0.012**)
	[**0.042**]	[**0.028**]	[**0.022**]	[**0.015**]	[**0.012**]
0.956	**0.948**	**0.955**	**0.958**	**0.959**	**0.958**
	(**0.040**)	(**0.029**)	(**0.023**)	(**0.016**)	(**0.013**)
	[**0.041**]	[**0.029**]	[**0.023**]	[**0.017**]	[**0.013**]
0.877	**0.892**	**0.890**	**0.889**	**0.886**	**0.884**
	(**0.041**)	(**0.032**)	(**0.026**)	(**0.020**)	(**0.017**)
	[**0.044**]	[**0.034**]	[**0.029**]	[**0.022**]	[**0.018**]
0.794	**0.831**	0.820	0.814	0.807	0.804
	(**0.044**)	(0.037)	(0.031)	(0.025)	(0.021)
	[**0.057**]	[0.045]	[0.037]	[0.028]	[0.023]

Note. TLI = Tucker–Lewis Index. For each TLI × sample size condition, mean values (M) are printed without parentheses, standard deviations (SD) in parentheses, and root mean squared errors in brackets. For each TLI × sample size condition, these three values are printed in boldface when no other estimator reported in Tables 7 and 9 has a smaller root mean squared error.

As pointed out earlier, our corrections are not limited to RMSEA estimation and can be applied for all noncentrality–based fit statistics. Given the reported values for T_{ML} and the number of observed variables p, the traditional estimate of Steiger's γ (cf. Hu & Bentler, 1999), for example, equals 0.833 with a 90% confidence interval [0.776, 0.888], but the Swain–corrected estimate for this statistic equals 0.915 with a confidence interval [0.856, 0.971], suggesting a much better model fit compared to the traditional estimate.

For the calculation of $\hat{TLI}$, which was not reported by Obermiller et al. (2005), we derived the test statistic for the independence model, $T_i = 1122.784$, from their reported estimate $\hat{CFI} = 0.838$. It follows that $\hat{TLI} = 0.817$, but the Swain–corrected estimate $\hat{TLI}_{s1} = 0.916$ hints at a much better population model fit. A similar conclusion is reached when comparing the traditional estimate $\hat{CFI} = 0.838$ with the Swain–corrected estimate $\hat{CFI}_{s1} = 0.926$.

The model size of Obermiller et al. (2005) is comparable to the simulated population model size in our study and their sample size is close to our smallest sample size condition. From the performance of the Swain–corrected estimators in Tables 5 and 8, the corrected estimates are expected to be much closer to the population values than the traditional estimates reported by

TABLE 9
Performance of $\hat{\text{TLI}}_{s2}$

	Sample Size				
TLI	*N* = *50* *N:t* ≈ *0.9*	*N* = *75* *N:t* ≈ *1.3*	*N* = *100* *N:t* ≈ *1.7*	*N* = *150* *N:t* ≈ *2.6*	*N* = *200* *N:t* ≈ *3.4*
M (*SD*) [root mean squared error]					
1.000	0.971	0.990	0.996	0.998	0.999
	(0.052)	(0.032)	(0.023)	(0.015)	(0.011)
	[0.060]	[0.034]	[0.024]	[0.015]	[0.011]
0.978	0.951	0.968	0.973	0.976	**0.977**
	(0.052)	(0.033)	(0.024)	(0.016)	(**0.012**)
	[0.059]	[0.035]	[0.025]	[0.016]	[**0.012**]
0.956	0.931	0.947	0.953	0.955	0.956
	(0.053)	(0.034)	(0.026)	(0.018)	(0.014)
	[0.058]	[0.036]	[0.026]	[0.018]	[0.014]
0.877	0.856	0.870	0.875	**0.877**	**0.878**
	(0.054)	(0.038)	(0.030)	(**0.022**)	(**0.018**)
	[0.060]	[0.038]	[0.030]	[**0.022**]	[**0.018**]
0.794	0.772	**0.786**	**0.790**	**0.793**	**0.793**
	(0.059)	(**0.043**)	(**0.035**)	(**0.027**)	(**0.022**)
	[0.063]	[**0.044**]	[**0.035**]	[**0.027**]	[**0.022**]

Note. TLI = Tucker–Lewis Index. For each TLI × sample size condition, mean values (M) are printed without parentheses, standard deviations (SD) in parentheses, and root mean squared errors in brackets. For each TLI × sample size condition, these three values are printed in boldface when no other estimator reported in Tables 7 and 8 has a smaller root mean squared error.

Obermiller et al. (2005). The Swain–corrected estimates of the discussed fit indexes suggest a rather acceptable (albeit not very good) model fit.

For a recent application of our proposed Swain–corrected fit measures see Morhart, Herzog, and Tomczak (in press).

SOFTWARE

For convenient calculation of the proposed corrections in applied research, the `R` function `swain` is provided at `http://www.gmw.rug.nl/~boomsma` along with documentation (Boomsma & Herzog, 2007). The function uses the chi–square statistic of the target model T_{ML}, the chi–square statistic of the independence model T_i, sample size N, degrees of freedom d, and the number of variables p as input variables (these values are obtained from the output of any standard software package like Amos, EQS, LISREL, or M*plus*) and calculates both uncorrected and the proposed Swain–corrected estimates of noncentrality–based statistics (including confidence intervals) and incremental fit indexes.

GENERAL DISCUSSION

Summary and Recommendations

"Structural equation modeling software does not work with small sample sizes." In our opinion, this is a quite broadly established conviction among applied scientists. The results of this study, however, should encourage applied scientists to use covariance structure methodology even when rather small sample sizes are available and appropriate corrections of model fit estimators are applied.

For misspecified models, Swain's (1975) correction of T_{ML} closely follows a noncentral chi–square distribution for realistically sized models (in this study, 24 variables) and a sample size of $N = 75$. This corresponds to an *N:t* ratio of about 1.3:1, which is very promising for researchers dealing with small sample sizes because ratios of 5:1 have been recommended in earlier studies (Bentler & Chou, 1987). The study by Herzog et al. (2007), however, reveals that ratios of less than 2:1 should not be applied for larger models (even when Swain's correction is used). In summary, applied researchers dealing with *N:t* ratios close to 2:1 are on the safe side although smaller *N:t* ratios might be used for smaller models. Notice, once more, that this recommendation should be interpreted with great caution because the behavior of parameter and standard error estimators has not been investigated in this study. Although it is known that estimates of parameters and standard errors are robust for small to medium sample sizes under conditions of multivariate normality (e.g., Gerbing & Anderson, 1985), research about possible corrections of parameter and standard error estimators is needed before one should go beyond the "2:1 border."

Our supplementary analysis reveals that only the chi–square value of the target model, not that of the independence model, should be corrected by Swain's multiplier in estimators of population incremental fit indexes. Applied researchers with *N:t* ratios of 2:1 are on the safe side when Swain–corrected estimators of incremental fit indexes are applied.

Finally, it should be noted that Swain's multiplier converges to 1 when the sample size increases. Therefore, it should do no harm to use Swain's multiplier in general (independent of sample size). To do so, the `R` function `swain` is available for the convenient calculation of the proposed Swain–corrected estimators in applied research (Boomsma & Herzog, 2007).

Limitations and Research Opportunities

A few prospects for further research can be enumerated. First, nonnormality has an inflating effect on chi–square model fit statistics (cf. Boomsma, 1983), which also affects functionally related fit measures like RMSEA or TLI estimators. It would be interesting to investigate whether certain corrective procedures, such as Satorra and Bentler's (1994) scaling correction, in combination with Swain's multiplier would robustify T_{ML} against both small sample size and nonnormality for varying degrees of misspecification.

Second, this study was restricted to confirmatory factor models. Generalization of our findings to a variety of common model structures would be desirable. However, confirmatory factor models are used extensively in applied research, and it should be noted that no relevant differences compared to full covariance structure models were observed in earlier investigations of chi–square statistics (e.g., Nevitt & Hancock, 2004).

Third, our analysis did not deal with the small–sample robust point and interval estimation of parameters. It is known, however, that chi–square differences based on T_{ML} can be used to estimate confidence intervals for parameters and test the significance of parameter estimates (Cheung, 2007; Neale & Miller, 1997). It might be promising to investigate whether confidence intervals and significance tests based on Swain–corrected chi–square differences perform well under conditions of small sample sizes.

Finally, noncentrality–based and incremental fit indexes are frequently used to test for measurement invariance in multiple–group analyses (e.g., Chen, 2007). It might be interesting to generalize our proposed estimators to multiple–group structural equation modeling and to compare their behavior with that of traditional estimators when sample size is small.

ACKNOWLEDGMENTS

We would like to thank the editor and three anonymous reviewers for their efforts in the review process. We are indebted to James H. Steiger (Vanderbilt University) and Kenneth A. Bollen (University of North Carolina, Chapel Hill) for their valuable comments on our project. This article is an elaboration of a paper presented at the 71st Annual Meeting of the Psychometric Society, June 14, 2006, HEC Montréal, Canada.

REFERENCES

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London, Series A, 160,* 268–282.

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology (Statistical Section), 3,* 77–85.

Bartlett, M. S. (1954). A note on the multiplying factors for various χ^2 approximations. *Journal of the Royal Statistical Society, Series B, 16,* 296–298.

Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural equation modeling. *Sociological Methods & Research, 16,* 78–117.

Bentler, P. M., & Yuan, K.-H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research, 34,* 181–197.

Boomsma, A. (1983). *On the robustness of LISREL (maximum likelihood estimation) against small sample size and non–normality.* Unpublished doctoral dissertation, University of Groningen, The Netherlands.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika, 50,* 229–242.

Boomsma, A., & Herzog, W. (2007). `R` function swain: Correcting structural equation model fit statistics and indexes under small–sample and/or large–model conditions. Retrieved January 21, 2008, from `http://www.gmw.rug.nl/~boomsma`

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* New York: Guilford.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–142). Cambridge, UK: Cambridge University Press.

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

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14,* 464–504.

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research, 36,* 462–494.

Cheung, M. W. L. (2007). Comparison of approaches to constructing confidence intervals for mediating effects using structural equation models. *Structural Equation Modeling, 14,* 227–246.

Cudeck, R., & Henly, S. J. (1991). Model selection in covariance structures analysis and the "problem" of sample size: A clarification. *Psychological Bulletin, 109,* 512–519.

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

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

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1,* 16–29.

Fouladi, R. T. (2000). Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate nonnormality. *Structural Equation Modeling, 7,* 356–410.

Gerbing, D. W., & Anderson, J. C. (1985). The effects of sampling error and model characteristics on parameter estimation for maximum likelihood confirmatory factor analysis. *Multivariate Behavioral Research, 20,* 255–271.

Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness–of–fit indices for structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 40–65). Newbury Park, CA: Sage.

Herzog, W., Boomsma, A., & Reinecke, S. (2007). The model–size effect on traditional and modified tests of covariance structures. *Structural Equation Modeling, 14,* 361–390.

Homburg, C., & Fürst, A. (2005). How organizational complaint handling drives customer loyalty: An analysis of the mechanistic and the organic approach. *Journal of Marketing, 69,* 95–114.

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

Hu, L. T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112,* 351–362.

Jackson, D. L. (2001). Sample size and number of parameter estimates in maximum likelihood confirmatory factor analysis: A Monte Carlo investigation. *Structural Equation Modeling, 8,* 205–223.

Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling, 10,* 128–141.

Jackson, D. L. (2007). The effect of the number of observations per parameter in misspecified confirmatory factor analytic models. *Structural Equation Modeling, 14,* 48–76.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis.* New York: Guilford.

Kensuke, O., Takahiro, H., & Kazuo, S. (2005, July). *Bartlett correction of likelihood ratio statistics in structural equation modeling.* Paper presented at the annual meeting of the Psychometric Society, Tilburg, The Netherlands.

Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33,* 181–220.

Marsh, H. W., Hau, K.-T., & Wen, Z. L. (2004). In search of golden rules: Comment on hypothesis–testing approaches to setting cutoff values for fit indexes and danger in overgeneralizing Hu and Bentler's 1999 findings. *Structural Equation Modeling, 11,* 320–341.

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

Morhart, F. M., Herzog, W., & Tomczak, I. (in press). Brand–specific leadership: Turning employees into brand champions. *Journal of Marketing.*

Muthén, L. K., & Muthén, B. O. (2007). M*plus user's guide* (4th ed.). Los Angeles: Muthén & Muthén.

Neale, M. C., & Miller, M. B. (1997). The use of likelihood–based confidence intervals in genetic models. *Behavior Genetics, 27,* 113–120.

Nevitt, J., & Hancock, G. R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling, 8,* 353–377.

Nevitt, J., & Hancock, G. R. (2004). Evaluating small sample approaches for model test statistics in structural equation modeling. *Multivariate Behavioral Research, 39,* 439–478.

Obermiller, C., Spangenberg, E., & MacLachlan, D. L. (2005). Ad skepticism: The consequences of disbelief. *Journal of Advertising, 34,* 7–17.

Olsson, U. H., Foss, T., & Breivik, E. (2004). Two equivalent discrepancy functions for maximum likelihood estimation: Do their test statistics follow a non–central chi–square distribution under model misspecification? *Sociological Methods & Research, 32,* 453–500.

Olsson, U. H., Foss, T., Troye, S. V., & Howell, R. D. (2000). The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural Equation Modeling, 7,* 557–595.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Savalei, V. (2008). Is the ML chi–square ever robust to nonnormality? A cautionary note with missing data. *Structural Equation Modeling, 15,* 1–22.

Sobel, M. E., & Bohrnstedt, G. W. (1985). Use of null models in evaluating the fit of covariance structure models. In N. B. Tuba (Ed.), *Sociological methodology* (pp. 152–178). San Francisco: Jossey–Bass.

Steiger, J. H., & Lind, J. C. (1980, May). *Statistically–based tests for the number of 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,* 253–264.

Swain, A. J. (1975). *Analysis of parametric structures for variance matrices.* Unpublished doctoral dissertation, Department of Statistics, University of Adelaide, Australia.

Tomarken, A. J., & Waller, N. G. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology, 1,* 31–65.

Venables, W. N., Smith, D. M., & The R Development Core Team. (2006). *An introduction to* R (Version 2.4.1). Retrieved June 14, 2007, from http://www.r-project.org/

Wakaki, H., Eguchi, S., & Fujikoshi, Y. (1990). A class of tests for a general covariance structure. *Journal of Multivariate Analysis, 32,* 313–325.

Wilks, S. S. (1938). The large–sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics, 9,* 60–62.

Yuan, K.-H. (2005). Fit indices versus test statistics. *Multivariate Behavioral Research, 40,* 115–148.