Revisiting the Model Size Effect in Structural Equation Modeling

Dexin Shi, Taehun Lee & Robert A. Terry

Published online: 28 Sep 2017.

Submit your article to this journal ☐

Article views: 22
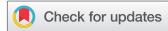
View related articles ☐

View Crossmark data ☐

Revisiting the Model Size Effect in Structural Equation Modeling

Dexin Shi,[1] Taehun Lee,[2] and Robert A. Terry[3]

[1]*University of South Carolina*
[2]*Chung-Ang University*
[3]*University of Oklahoma*

Fitting a large structural equation modeling (SEM) model with moderate to small sample sizes results in an inflated Type I error rate for the likelihood ratio test statistic under the chi-square reference distribution, known as the model size effect. In this article, we show that the number of observed variables (p) and the number of free parameters (q) have unique effects on the Type I error rate of the likelihood ratio test statistic. In addition, the effects of p and q cannot be fully explained using degrees of freedom (df). We also evaluated the performance of 4 correctional methods for the model size effect, including Bartlett's (1950), Swain's (1975), and Yuan's (2005) corrected statistics, and Yuan, Tian, and Yanagihara's (2015) empirically corrected statistic. We found that Yuan et al.'s (2015) empirically corrected statistic generally yields the best performance in controlling the Type I error rate when fitting large SEM models.

Keywords: correctional methods, likelihood ratio statistic, model size effect, structural equation modeling (SEM)

Structural equation modeling (SEM) allows researchers to study the relationships between observed and latent variables as well as among the latent variables (Bollen, 1989). In most applications of SEM, an important step is evaluating the proposed model's goodness of fit with the data. The likelihood ratio chi-square test under maximum likelihood (ML) estimation is the most commonly used statistical method for assessing overall goodness of fit (Jöreskog, 1969; Lawley & Maxwell, 1971). The basic framework of the chi-square test can be summarized as follows.

When fitting an SEM model, researchers obtain sample data with a number of observations N, a number of observed variables p, and a sample (observed) covariance matrix **S**. The corresponding population covariance matrix Σ is estimated by **S**. A model-implied covariance matrix $\hat{\Sigma}(\theta)$ is produced by the covariance structural model with q freely estimated parameters stored in a vector θ. The model estimation procedure yields a vector of estimated parameter values $\hat{\theta}$ that minimize the distance between the observed

and model-implied covariance matrices. This distance is measured by a discrepancy function $F(S, \hat{\Sigma}(\theta))$. Several discrepancy functions can be used to estimate SEM models (e.g., GLS [Browne, 1973], WLS [Browne, 1984]). Among the many existing fit functions, the most commonly used discrepancy function is based on ML estimation, with the following ML discrepancy function:

$$F_{ML} = \ln|\hat{\Sigma}(\theta)| - \ln|S| + \text{Tr}(S\hat{\Sigma}(\theta)^{-1}) - p.$$

Under standard ML estimation, the null hypothesis (H_0) that the model-implied covariance matrix is equal to the population covariance matrix can be evaluated statistically using the likelihood ratio statistic (Bollen, 1989; Jöreskog, 1969), defined as

$$T_{ML} = (N-1)F_{ML.} \tag{1}$$

Assuming that the null hypothesis holds, the test statistic asymptotically follows a central chi-square distribution with degrees of freedom (df) equal to $p(p + 1)/2 - q$. If the observed chi-square exceeds a critical value, which is determined by both df and the minimum acceptable Type I error rate

---

Correspondence should be addressed to Taehun Lee, Department of Psychology, Chung-Ang University, 84 Heukseok-Ro, Dongjak-Gu, Seoul 06974, Korea. E-mail: lee.taehun@gmail.com

(α, typically .05), the null hypothesis of having a correctly specified (or true) model is rejected. This chi-square test allows researchers to evaluate the fitness of a model using the null hypothesis significance test (NHST) approach. In this sense, the chi-square test is regarded as "the only substantive test of fit for SEM" (Barrett, 2007, p. 815).

For the chi-square test to be valid, the behavior of the test statistic T_{ML} is subject to two major assumptions: (a) +the observed variables are multivariate normally distributed, and (b) the sample size N is sufficiently large. Multiple studies have shown that with nonnormal data, the test statistic under standard ML estimation could be either over- or underestimated, depending on the size of kurtosis relative to the normal distribution (see Curran, West, & Finch, 1996; Yuan, Bentler, & Zhang, 2005). In addition, it is generally understood that fitting a large SEM model (with many observed variables) with a small sample results in an inflated Type I error rate for the chi-square test under the chi-square reference distribution. This inflated rejection rate of the likelihood ratio chi-square test is known as the model size effect (Herzog, Boomsma, & Reinecke, 2007; Moshagen, 2012; Shi, Lee, & Terry, 2015).

Due to its important empirical implications, the model size effect in SEM has been investigated in many studies (e.g., Anderson & Gerbing, 1984; Herzog et al., 2007; Jackson, 2003; Kenny & McCoach, 2003). In these studies, the concept of "model size" has been defined using factors such as the number of observed variables (p), ratio of the observed variables to latent factors (p/f), the number of parameters to be estimated (q), or the degrees of freedom ($df = p(p + 1)/2 - q$). However, the operational definitions of "model size" used in the previous studies were somewhat ambiguous in that the ingredients tended to vary together. For example, a larger number of indicators (p) often resulted in more parameters being estimated (q), and thus also a larger df. Likewise, a large p typically implied a large p/f ratio. As a result, most former researchers, despite or due to these factors being changed together in their studies, simply attributed the inflated Type I error rates to one single component among all of the constituent factors (e.g., p, q, p/f, or df).[1]

In practice, however, given a fixed number of variables or items p, factor analysis models can vary tremendously in the number of their parameters (q) and latent factors (f),

depending on differences in model specifications. For 60 observed variables, fitting a single-factor parallel measurement model implies only two freely estimated parameters ($q = 2$), whereas a congeneric test model will yield about 120 free parameters ($q = 120$). In a 10-factor exploratory structural equation model (ESEM; Asparouhov & Muthén, 2009; Marsh et al., 2009) with only minimal constraints for model identification, the number of free parameters would be inflated to about 660.

Therefore, it is important to examine the effects of individual components of model size separately (cf. Moshagen, 2012). Disentangling the unique effects of different factors can be important for at least two reasons. First, understanding the unique effects of different factors helps empirical researchers to make better judgments (in terms of the risk of eliciting Type I errors) under different modeling situations. Suppose that a researcher fits a 20-factor model with 60 observed variables (i.e., $p = 60$, $f = 20$, $p/f = 3$) to a sample of small size. If the results of the study revealed that the number of observed variables (p) is the most influential factor for the model size effect, the researcher is advised to take the large p into account when evaluating model fit. However, if it is actually the ratio of the observed variables to latent factors (p/f) that is responsible for the model size effect, then the researcher might not need to be concerned as the p/f ratio is only three (assuming that the ratio of three turned out to exercise negligible effect). Second, information about the magnitude of effects of individual factors (e.g., p, q, f/p, df) will provide methodologists with an opportunity to develop correctional methods for the model size effect.

In fact, several correctional methods have been proposed for adjusting the inflated Type I error rates in large SEM models, including the Bartlett-corrected statistic (1950), the Swain-corrected statistic (1975), the Yuan-corrected statistic (2005), and Yuan, Tian, and Yanagihara's (2015) empirical corrected statistic. Herzog et al. (2007) compared the performance of the Bartlett-corrected statistic and the Swain-corrected statistic in evaluating large SEM models. Using a 5% empirical rejection rate and the Kolmogorov–Smirnov test as outcomes, they found that the ML ratio statistics with Swain correction yielded better performance in a variety of (simulated) modeling conditions. Yuan et al. (2015) also demonstrated that, compared to previous analytical corrections (i.e., the Bartlett-corrected statistic, the Swain-corrected statistic, and the Yuan-corrected statistic), their empirical correction performed better in terms of controlling for Type I error rates.

In evaluating the performance of correctional methods, however, the modeling conditions considered in previous studies were somewhat limited. For example, in the Herzog et al. (2007) simulation, the number of observed variables (p) ranged from 12 to 48, whereas for each p, only one single value of q was considered. In psychological studies, many questionnaires can include a large number of items

---

[1] For example, in Kenny and McCoach (2003) simulation design, single-factor models were generated with manipulations of the number of observed variables (p). Consequently, across their simulation cases, the number of observed variables (p), the number of estimated parameters (q), the degrees of freedom (df), and the number of indicators per factor (p/f) were all changed together. However, in their conclusions, they attributed the model size effect solely to the number of observed variables. In Herzog et al. (2007) study, all four factors (i.e., p, q, df and p/f) differ across simulated conditions. In this sense, the inflated nominal Type I error rate (i.e., model size effect) could be attributed to any of the factors, yet the researchers simply interpreted the influences caused by changes in degrees of freedom (df).

(i.e., p). For example, the commonly used revised NEO Personality Inventory (NEO PI–R) has 240 items (Costa & MacCrae, 1992). In education-related studies, many comprehensive tests also include more than 100 items (e.g., ETS Major Field Tests; Ling, 2012). In this sense, the conclusions from the extant literature might have been derived based on a limited scope of confirmatory factor analysis (CFA) modeling scenarios.

The purpose of our study is to investigate putative influential factors for model size effect in Monte Carlo simulations designed to examine the effects of individual factors separately. Under such modeling conditions, we will compare the performance of four existing correctional methods with respect to their Type I error rates. The results obtained from the simulation will provide a basis for practical guidance for empirical researchers working with large models. Limitations and future research directions are also discussed.

THE CONCEPT OF MODEL SIZE

Previous studies have used different indexes to measure SEM model size, including the number of observed variables (p), the number of parameters to be estimated (q), the degrees of freedom ($df = p (p + 1)/2 - q$), and the ratio of the observed variables to latent factors (p/f).

Some researchers have used the number of observed variables (p) to define the size of an SEM model. For example, Kenny and McCoach (2003) found that when the sample size is small (e.g., less than 200), the empirical distribution of the likelihood ratio test statistics exhibits greater upward bias[2] as the number of observed variables (p) increases, even when the models are correctly specified. They also showed that the effect of a large number of observed variables was weaker when the sample size was large (i.e., $N = 1,000$). More recently, Moshagen (2012) studied the model size effect by manipulating p across a wider range (from 15–90). Moshagen's study clearly showed that with small samples ($N = 200$), a larger number of observed variables leads to an inflated Type I error rate of the likelihood ratio statistic and larger Kolmogorov–Smirnov distances between the empirical and theoretical chi-square distribution of the chi-square statistic.

Other researchers have focused on the effect of the number of free parameters to be estimated (q) on the empirical distribution of the likelihood ratio statistic. Jackson (2003) manipulated the number of model parameters (q) while keeping the number of observed variables (p) and sample size (N) fixed. Using "chi-square bias" (i.e., relative difference, which is taking the difference between the mean of the empirical chi-square and its expected value [df], then dividing by the expected value) as

the outcome variable, he found that the average values of the likelihood ratio statistic were closer to the expected values with fewer free parameters (i.e., a larger N:q ratio).

In other studies, researchers have attributed model size effects to degrees of freedom (df), which is a function of both p and q. Hoogland and Boomsma (1998) defined the size of the model as its df. Based on a review of previous literature, they concluded that the chi-square test suffered inflated Type I error rates mainly when fitting large models measured by large df to a small sample size. Additionally, using simulation studies, Herzog et al. (2007) also observed that inflated Type I error rates were associated with larger df.

Finally, the number of indicators per factor (p/f) has also been examined concerning its role in influencing model size effects. Ding, Velicer, and Harlow (1995) demonstrated that as the p/f ratio increases from 2 to 6, the χ^2/df ratio increases as well, indicating a worsening model fit. Marsh, Hau, Balla, and Grayson (1998) also found that a larger ratio of observed variables to latent factors (p/f) is associated with a larger bias in terms of the χ^2/df ratio in likelihood ratio statistics.

To the best of our knowledge, Moshagen (2012) is the only simulation study exploring the unique effects among the potential influential factors. In his paper, Moshagen generated a series of models that varied in the number of observed variables (p) and the number of free parameters (q). After evaluating empirical rejection rates and Kolmogorov–Smirnov distance (between empirical and theoretical chi-square distributions) across a variety of combinations of p and q, Moshagen concluded that p was the "only" influential factor that needed to be considered, whereas q and df could safely be neglected (pp. 86, 96). Despite its merit for manipulating the potential influential factors separately, there are shortcomings in Moshagen's (2012) simulation design. First, the manipulations of q and df in the simulation varied within a narrow range, which means that the conclusions might not apply to a larger set of representative studies. Moreover, the manipulations varied q and df simultaneously with other possibly influential factors. For example, as Moshagen manipulated q, the number of latent factors was also manipulated, thus changing the p/f ratio.

CORRECTIONS FOR MODEL SIZE EFFECT

Methodologists have explored possible corrections for the likelihood ratio statistic to modify the effect of model size. Currently, four major corrections have been discussed in the model size literature: the Bartlett-corrected statistic (1950), the Swain-corrected statistic (1975), the Yuan-corrected statistic (2005), and the Yuan et al. (2015) empirical corrected statistic. The four corrections are as follows.

The Bartlett-corrected statistic (1950, 1954) was originally developed for correcting the chi-square statistic with a small sample size within the framework of explanatory factor analysis (EFA). However, researchers later suggested also using the Bartlett-corrected statistic for general SEM

---

[2] The empirical means of the chi-square statistics become larger than their expected values (= df).

models (Fouladi, 2000). Bartlett proposed a multiplier b for the likelihood ratio statistic:

$$T_{MLb} \equiv bT_{ML} \qquad (2)$$

where b is a function of the number of factors (f), the number of observed variables (p), and the sample size (N):

$$b = 1 - \frac{4f + 2p + 5}{6(N - 1)} \qquad (3)$$

Yuan (2005) modified the Bartlett-corrected statistic and proposed an "ad hoc" small-sample correction for general SEM models. The Yuan-corrected statistic is defined as

$$T_{MLy} \equiv yT_{ML} \qquad (4)$$

where y is a function of the number of factors (f), number of observed variables (p), and sample size (N):

$$y = 1 - \frac{2f + 2p + 7}{6(N - 1)} \qquad (5)$$

In the single-factor case, the Yuan-corrected statistic is exactly the same as the Bartlett-corrected statistic.

Instead of starting from the EFA, Swain (1975) proposed four corrections for a small sample size under the general SEM framework. In this article, for the purpose of comparison, we chose to study the Swain-corrected statistic that was examined in the previous studies (Fouladi, 2000; Herzog et al., 2007; Yuan et al., 2015), which is defined as

$$T_{MLs} \equiv sT_{ML} \qquad (6)$$

where

$$s = 1 - \frac{p(2p^2 + 3p - 1) - t(2t^2 + 3t - 1)}{12d(N - 1)} \qquad (7)$$

and

$$t = \frac{\sqrt{1 + 4p(p + 1) - 8d} - 1}{2} \qquad (8)$$

with p denoting the number of observed variables, d indicating the degree of freedom, and N representing the sample size.

Yuan et al. (2015) suggested an empirical correction specifically for model size effect. The empirically corrected statistic is developed based on simulations across 342 different conditions that vary the sample size (N), the number of observed variables (p), the number of latent factors (f), the number of freely estimated parameters (q), and the degrees of freedom (df). Yuan et al. (2015) formulated their recommended correction as follows:

$$T_{MLe} \equiv eT_{ML} \qquad (9)$$

where

$$e = \frac{N - (2.381 + 0.361p + 0.006q)}{N - 1} \qquad (10)$$

with all letter notations defined as previously listed.

STUDY I

In Study I, we aimed to clarify the concept of "model size" by exploring the unique effect of four potential factors, including the number of observed variables (p), the number of parameters to be estimated (q), the degrees of freedom ($df = p(p + 1)/2 - q$), and the ratio of the observed variables to latent factors (p/f). The effects of these factors on the likelihood ratio statistic were examined using a series of Monte Carlo simulations.

First, we generated data based on CFA models in which the number of observed variables (p) and the number of latent factors (f) were varied. For simplicity, we adopted a CFA measurement model with continuous outcomes as the population model. We set the population values for all factor loadings to 0.7, and set the population residual variances to 0.51. The factor variances were fixed to one. For each simulation condition, we generated 1,000 data sets. To be consistent with Moshagen (2012), the sample sizes were fixed at 200. All simulation studies were conducted using M*plus* 7.11 (Muthén & Muthén, 1998).

Next, we fitted a series of models to the simulated data based on certain combinations of p, q, and p/f (conditions). To avoid model misspecification, for each modeling condition, the number of observed variables (p) and the number of observed variables per factor (p/f) were predetermined to be identical with the population model. However, to manipulate the number of free parameters (q) to a greater degree, we either specified more parameters to be estimated (e.g., correlated residuals)[3] or constrained the parameters to their true population values (to decrease q).

To examine the model size effect, we compared the two distributions: the empirical distributions of the chi-square statistics and the associated theoretical chi-square

---

[3] We recognize that to manipulate the number of free parameters (q) to a greater degree, in some fitted models, many parameters (e.g., correlations among residuals) are freely estimated; even their population values are zeros. Strictly speaking, such overparameterized models are considered misspecified (cf. Hu & Bentler, 1998; Saris, Satorra, & Van Der Veld, 2009). However, both overparameterized models and correctly specified models are free of model errors in the population (or the approximation discrepancy equals zero; see Cudeck & Henly, 1991), thus they produce the same theoretical chi-square distributions. As a result, for overparameterized models, we should expect the empirical Type I errors to be close to nominal rates as well.

distributions. To evaluate the goodness of approximation to the chi-square distributions, first, we compared the averages of the empirical distributions with the expected values of the theoretical chi-square distributions for all modeling conditions. The absolute difference (Δ_{mean}) was defined as $\Delta_{mean} = \frac{\sum \chi^2_{emp}}{1000} - df$, where $\frac{\sum \chi^2_{emp}}{1000}$ indicates the mean of the empirical distribution of the chi-square statistic across 1,000 replications and df represents the degrees of freedom, or the expected value of the theoretical chi-square distribution. In addition, we evaluated the empirical rejection rates of the chi-square test statistics for nominal alpha levels of 10%, 5%, and 1%, which are of special interest in practice. The following subsections summarize the detailed conditions and results for each of the potential factors.

Effect of p/f

To examine the effect of the ratio of the observed variables to latent factors (p/f), we considered two different numbers of observed variables ($p = 30$ or $p = 60$), and for each value of p, we included three values for q (the number of free parameters: small-sized q, medium-sized q, and large-sized q). For each fixed combination of p and q, we varied the number of observed variables per factor (p/f) to a large degree. For example, for $p = 60$ and $q = 120$, we manipulated six levels for the p/f ratio, ranging from 3 ($f = 20$) to 60 ($f = 1$). In total, we investigated 36 ($2 \times 3 \times 6$) different conditions. A complete list of the modeling scenarios is provided in Table 1.

Table 1 shows that, given fixed numbers of observed variables (p) and free parameters (q), all three outcome variables exhibited very small changes across six different conditions for p and p/f. This implies that the model size effect in the likelihood ratio statistic is not influenced by the number of latent factors (f) or the ratio of the observed variables to latent factors (p/f) above and beyond p, q, and df. For example, with $p = 30$ and $q = 60$, all six models (Model IDs 7–12) varied considerably from 3 to 30 in p/f but yielded very similar results for all outcome variables. A similar pattern can also be observed for other combinations of p and q.

Effect of p

We considered eight different values of the number of observed variables (p), ranging from 10 to 120. To examine the unique effects of p, we kept the p/f ratio at 10. In addition, the number of free parameters (q) was also fixed as a constant. Two values of q were considered, which implies 16 (8×2) different scenarios. Table 2 summarizes the results for the effects of p.

As Table 2 shows, a larger value of p was associated with larger differences in the expected value and higher empirical rejection rates. For example, when q is 1, as p increased

from 10 (Model ID 1) to 90 (Model ID 7), the empirical rejection rates for 1% nominal significance level increased from 1% to 100%. This suggests that researchers will always reject the correctly specified model according to the chi-square test. In addition, the absolute mean difference increased from 0.48 to 842.86, suggesting a very large discrepancy between the empirical and expected chi-square distributions.

Effect of q

To examine the unique effects of the free parameters (q), for each fixed value of p, we manipulated q at six different levels. As mentioned earlier, we manipulated the values of q to a greater degree than in previous studies. For example, in Moshagen (2012), only two levels of q ($q = 123$ or $q = 310$) were examined when $p = 60$, whereas in our simulation, q ranged from 1 to 1,030 for the same value of p. We included 48 (8×6) different conditions in our analyses; Table 3 summarizes the results.

Table 3 shows that the number of parameters (q) had a unique effect on the likelihood ratio statistics in that increasing the number of estimated parameters reduced bias in the likelihood ratio statistics. For example, under the condition $p = 60$, as q increased from 1 to 1,030 (Model IDs 31–36), empirical rejection rates decreased from 88.8% to 25.9% for a 1% nominal alpha level. The absolute mean differences also suggest a better approximation to the theoretical chi-square distribution as q increases. It can be seen in Table 3 that the effects of q held across different levels of p.

Effect of df

As the preceding analyses demonstrate, both p and q are influential factors for model size effects, and their effects act in the opposite direction. Thus, we hypothesized that the effects of p and q can be explained parsimoniously by the degrees of freedom ($df = p (p + 1)/2 - q$). To test this hypothesis, we manipulated both p and q so that pairs of p and q produce the same degrees of freedom ($df = p (p + 1)/2 - q$). Table 4 shows the pairs of p and q used to produce the six levels of df.

The results presented in Table 4 show that p and q had unique effects above and beyond the degrees of freedom. For example, it can be seen that the two cases in P3, with the same df (464), produced noticeably different outcomes; the 5% empirical rejection rates increased from 23.4% ($p = 30$, $q = 1$, and $df = 464$) to 37.6% ($p = 40$, $q = 365$, and $df = 464$) as p increased from 30 to 40. Similar patterns of results also occurred across all other conditions.

STUDY II

In Study II, we aimed to use our comprehensive design to investigate the relative performance of proposed

TABLE 1
Effect of p/f on the Likelihood Ratio Test Statistic

Model ID	p	q	df	f	p/f	Δ_{Mean}	Empirical Rejection Rate %		
							1%	5%	10%
1	30	1	464	1	30	29.04	8.6%	23.8%	34.5%
2	30	1	464	2	15	29.06	8.7%	24.1%	34.6%
3	30	1	464	3	10	29.05	8.6%	23.5%	34.5%
4	30	1	464	5	6	29.05	8.6%	23.5%	34.5%
5	30	1	464	6	5	29.04	8.6%	23.9%	34.7%
6	30	1	464	10	3	29.07	8.6%	24.2%	35.0%
7	30	60	405	1	30	27.99	8.6%	24.2%	36.8%
8	30	60	405	2	15	28.13	9.1%	24.3%	36.4%
9	30	60	405	3	10	28.27	8.6%	24.7%	36.8%
10	30	60	405	5	6	28.07	8.4%	23.6%	37.0%
11	30	60	405	6	5	27.79	9.1%	24.3%	36.3%
12	30	60	405	10	3	27.97	8.4%	24.7%	35.5%
13	30	240	225	1	30	12.97	4.7%	15.8%	25.5%
14	30	240	225	2	15	12.56	4.8%	13.3%	24.2%
15	30	240	225	3	10	13.55	4.4%	15.2%	24.8%
16	30	240	225	5	6	14.32	4.6%	16.3%	27.0%
17	30	240	225	6	5	13.73	5.1%	16.7%	25.6%
18	30	240	225	10	3	14.56	5.9%	16.5%	27.2%
19	60	1	1,829	1	60	230.15	89.9%	97.0%	98.8%
20	60	1	1,829	2	30	230.15	89.9%	96.9%	98.8%
21	60	1	1,829	3	20	230.14	89.8%	96.9%	98.9%
22	60	1	1,829	6	10	230.15	89.8%	96.9%	98.8%
23	60	1	1,829	10	6	230.14	89.8%	97.0%	98.9%
24	60	1	1,829	20	3	230.11	89.7%	97.1%	98.8%
25	60	120	1,710	1	60	228.48	91.6%	97.4%	98.8%
26	60	120	1,710	2	30	227.93	91.1%	97.8%	98.7%
27	60	120	1,710	3	20	228.36	92.0%	97.7%	98.8%
28	60	120	1,710	6	10	228.15	91.7%	98.1%	98.7%
29	60	120	1,710	10	6	227.97	91.3%	97.5%	98.5%
30	60	120	1,710	20	3	227.90	92.0%	97.4%	98.8%
31	60	600	1,230	1	60	169.48	81.2%	94.2%	96.9%
32	60	600	1,230	2	30	173.22	84.4%	94.3%	97.7%
33	60	600	1,230	3	20	172.48	83.4%	95.3%	98.2%
34	60	600	1,230	6	10	173.90	85.5%	95.3%	98.1%
35	60	600	1,230	10	6	173.43	82.6%	94.4%	97.6%
36	60	600	1,230	20	3	177.03	85.1%	95.0%	98.1%

Note. N = 200. p = number of observed variables; q = number of estimated parameters; df = degrees of freedom; f = number of factors; p/f = number of observed variables per factor; Δ_{Mean} = mean of the empirical chi-square statistics distribution minus degrees of freedom. 1%, 5%, and 10% indicate the 1%, 5%, and 10% empirical rejection rates, respectively.

adjustments for the model size effect. We considered four proposed corrections: the Bartlett-corrected statistic (1950), the Swain-corrected statistic (1975), the Yuan-corrected statistic (2005), and the Yuan et al. (2015) empirical corrected statistic. We employed the same method and procedures as in Study I. To obtain more generalized results, we included more comprehensive conditions by manipulating four variables as follows:

- *Sample size (N)*: We manipulated the sample size at six different levels, namely N = 200, 400, 600, 800, 1,000, and 2,000.

- *Number of observed variables (p)*: We manipulated p at eight different levels, namely p = 10, 20, 30, 40, 50, 60, 90, and 120.
- *Number of free parameters (q)*: For each fixed number of observed variables (p), we manipulated the number of free parameters (q) so as to take six different values. The actual values of q depend on the number of variables, p, with the largest range of q going from 1 to 2,550 (p = 120).
- *Number of factors (f)*: For most of the conditions, we considered single-factor CFA models only. We also included three large f conditions: f = 20 (p = 60),

TABLE 2
Effect of *p* on the Likelihood Ratio Test Statistic

| Model ID | p | q | df | Δ_{Mean} | Empirical Rejection Rate % | | |
					1%	5%	10%
1	10	1	54	0.48	1.0%	6.8%	11.6%
2	20	1	209	6.96	2.3%	10.0%	16.7%
3	30	1	464	29.07	8.6%	23.4%	34.7%
4	40	1	819	66.63	24.3%	46.7%	62.0%
5	50	1	1,274	135.41	59.3%	81.7%	89.5%
6	60	1	1,829	225.84	88.8%	96.6%	98.6%
7	90	1	4,094	842.86	100.0%	100.0%	100.0%
8	120	1	7,259	2229.71	100.0%	100.0%	100.0%
9	10	52	3	0.02	1.3%	6.0%	10.6%
10	20	52	158	5.61	1.6%	9.2%	16.4%
11	30	52	413	24.97	6.7%	21.2%	33.0%
12	40	52	768	62.02	23.4%	44.0%	59.9%
13	50	52	1,223	131.13	59.9%	79.9%	89.7%
14	60	52	1,778	230.19	91.0%	97.9%	98.8%
15	90	52	4,043	845.81	100.0%	100.0%	100.0%
16	120	52	7,208	2230.40	100.0%	100.0%	100.0%

Note. $N = 200$. p = number of observed variables; q = number of estimated parameters; df = degrees of freedom; Δ_{Mean} = mean of the empirical chi-square statistics distribution minus degrees of freedom. 1%, 5%, and 10% indicate the 1%, 5%, and 10% empirical rejection rates, respectively.

$f = 30$ ($p = 90$), and $f = 40$ ($p = 120$). According to the results from Study I, the number of latent factors (f) and thus the ratio of the observed variables to latent factors (p/f) are not influential factors for the model size effect. However, the number of latent factors (f) is included in the formula for Bartlett's and Yuan's corrected statistics. Therefore, we manipulated the number of factors (f) to better understand the performance of those two corrections.

In total, we included 396 different scenarios. Table 5 summarizes the results of the analysis. Due to space constraints, we only report the absolute difference between the empirical and theoretical means (Δ_{mean}) and the 5% empirical rejection rate in Table 5. To better demonstrate the comparisons, we considered empirical rejection rates between 2% and 8% acceptable and these cases are indicated in bold in Table 5.

For the uncorrected chi-square statistics, in line with what we observed in Study I, a larger number of observed variables (p) was associated with a poor approximation to the asymptotic chi-square distribution. In addition, increasing the number of estimated parameters (q) resulted in a closer approximation to the asymptotic chi-square distribution. Note that p appears to be the most important determinant of the inflated Type I error rates. For example, when $N = 200$ and $p = 10$, the empirical rejection rates were close to the 5% nominal rates across all six levels of q (Model IDs 1–6). As p was increased from 10 to 20, the empirical rejection rates were inflated, except when $q \geq 151$ (Model IDs 11 &12). Nevertheless, when $p \geq 40$, all empirical

rejection rates were noticeably larger than the nominal rates, even when q was extremely large (Model IDs 19–66). We also see that the empirical likelihood ratio statistics show better approximation to the theoretical chi-square distribution as sample size increased. To achieve acceptable Type I error rates, the required sample size (N) can be roughly approximated by a quadratic function of p (i.e., $N \geq p^2$). For example, we observed that when $p = 20$, the empirical rejection rates became close to the nominal level as $N \geq 400$. As p increased to 30 and 40 (regardless of q), acceptable Type I error rates could be obtained as N reached approximately 800 and 2,000, respectively. Nevertheless, when $p \geq 60$, inflated Type I errors were generally observed even when the sample size reached 2,000 (at any level of q).

Compared to the uncorrected chi-squares, all four corrected statistics resulted in more reasonable empirical rejection rates across conditions. As we expected, in one-factor models, Bartlett's (1950) and Yuan's (2005) corrected statistics yielded exactly the same results. In general, these two corrected statistics could reduce the inflation of Type I rates; for example, when $p = 40$ and $N = 200$ (Model IDs 19–24), without corrections, the 5% empirical rejection rates ranged from 14.10% to 45.10% (across different q); under Bartlett's (1950) or Yuan's (2005) correction, for the same conditions, the rejection rates increased from 3.4% to 8.1%. However, both corrections failed to achieve acceptable rejection rates when the $p:N$ ratio was relatively large. For example, under one-factor models, inflated Type I error rates could be observed with p larger than 50 ($N = 200$), even after corrections. In

TABLE 3
Effect of q on the Likelihood Ratio Test Statistic

Model ID	p	q	df	Δ_{Mean}	Empirical Rejection Rate %		
					1%	5%	10%
1	10	1	54	0.48	1.0%	6.8%	11.6%
2	10	12	43	0.48	1.4%	6.9%	11.8%
3	10	22	33	0.67	1.0%	7.7%	12.3%
4	10	32	23	0.32	1.4%	7.2%	11.7%
5	10	42	13	−0.06	1.3%	5.2%	9.3%
6	10	52	3	0.02	1.3%	6.0%	10.6%
7	20	1	209	8.24	2.6%	10.8%	18.2%
8	20	31	179	7.73	2.5%	11.2%	18.6%
9	20	71	139	6.02	2.3%	9.9%	18.3%
10	20	111	99	3.73	1.8%	8.4%	14.7%
11	20	151	59	1.77	1.5%	6.6%	14.0%
12	20	191	19	0.09	0.9%	4.9%	10.4%
13	30	1	819	28.34	7.7%	23.1%	34.3%
14	30	135	685	25.82	7.8%	22.5%	35.3%
15	30	260	560	18.65	5.6%	18.3%	27.8%
16	30	385	435	12.49	4.6%	15.2%	24.7%
17	30	510	310	7.14	3.7%	10.8%	18.7%
18	30	635	185	2.00	1.1%	6.6%	13.3%
19	40	1	819	64.65	23.0%	45.0%	59.9%
20	40	135	685	61.89	22.2%	48.5%	62.0%
21	40	260	560	50.78	20.4%	42.0%	57.2%
22	40	385	435	36.03	13.5%	32.7%	45.3%
23	40	510	310	21.67	9.0%	19.4%	32.5%
24	40	635	185	10.25	4.0%	14.1%	21.9%
25	50	1	1,274	133.44	58.8%	80.8%	89.0%
26	50	180	1,095	128.64	62.5%	82.3%	89.9%
27	50	355	920	110.83	57.2%	78.7%	86.9%
28	50	530	745	87.85	46.5%	70.7%	80.3%
29	50	705	570	57.47	26.0%	50.7%	64.1%
30	50	880	395	33.43	13.0%	32.0%	44.3%
31	60	1	1,829	225.84	88.8%	96.6%	98.6%
32	60	320	1,510	210.56	90.2%	96.7%	98.1%
33	60	430	1,400	191.30	86.1%	95.1%	97.3%
34	60	630	1,200	166.50	80.4%	94.2%	97.3%
35	60	830	1,000	123.08	62.9%	83.1%	89.9%
36	60	1,030	800	69.68	25.9%	52.0%	66.7%
37	90	1	4,094	230.03	89.9%	97.1%	98.8%
38	90	340	3,755	227.62	91.3%	97.1%	98.5%
39	90	660	3,435	221.20	90.7%	97.5%	98.7%
40	90	980	3,115	212.01	90.4%	97.2%	98.9%
41	90	1,280	2,815	199.11	90.3%	97.4%	98.7%
42	90	1,580	2,515	177.03	85.1%	95.0%	98.1%
43	120	1	7,259	842.86	100.0%	100.0%	100.0%
44	120	550	6,710	827.85	100.0%	100.0%	100.0%
45	120	1,050	6,210	779.89	100.0%	100.0%	100.0%
46	120	1,550	5,710	743.87	100.0%	100.0%	100.0%
47	120	2,050	5,210	626.09	100.0%	100.0%	100.0%
48	120	2,550	4,710	508.98	100.0%	100.0%	100.0%

Note. N = 200. p = number of observed variables; q = number of estimated parameters; df = degrees of freedom; Δ_{Mean} = mean of the empirical chi-square statistics distribution minus degrees of freedom. 1%, 5%, and 10% indicate the 1%, 5%, and 10% empirical rejection rates, respectively.

addition, Bartlett's (1950) and Yuan's (2005) corrected statistics are sensitive to the number of factors (f); as f increased considerably, both test statistics were overcorrected. For example, under the conditions with N = 200, p = 90, and f = 30 (Model IDs 49–54), the average values of the empirical distributions after applying Bartlett's (1950) or Yuan's (2005) corrections were far smaller than the corresponding theoretical means.

TABLE 4
Effect of *df* on the Likelihood Ratio Test Statistic

Model ID	p	q	df	Δ_{Mean}	Empirical Rejection Rate %		
					1%	5%	10%
P1	10	1	54	0.48	1.0%	6.8%	11.6%
	20	156	54	1.53	1.5%	6.6%	11.3%
P2	20	1	209	6.96	2.3%	10.0%	16.7%
	30	256	209	11.73	4.4%	14.8%	24.2%
P3	30	1	464	29.07	8.6%	23.4%	34.7%
	40	365	464	41.89	16.3%	37.6%	50.7%
P4	40	1	819	66.63	24.3%	46.7%	62.0%
	50	465	819	96.57	1.0%	6.8%	11.6%
P5	50	1	1,274	135.41	1.5%	6.6%	11.3%
	60	556	1,274	168.30	2.3%	10.0%	16.7%

Note. $N = 200$. p = number of observed variables; q = number of estimated parameters; df = degrees of freedom; Δ_{Mean} = mean of the empirical chi-square statistics distribution minus degrees of freedom. 1%, 5%, and 10% indicate the 1%, 5%, and 10% empirical rejection rates, respectively.

Swain's corrected statistic is generally superior to Bartlett's and Yuan (2005)'s corrections. However, Swain's correction started exhibiting inadequate performance when p was equal to or greater than 90, especially when sample size was small (i.e., $N \leq 400$). For example, for Model IDs 49 to 54, when $p = 90$ and $N = 200$, the empirical Type I error rates were substantially inflated (up to about 27%). When p reached 120, inflated Type I error rates could still be observed, unless $N \geq 1,000$.

Overall, the Yuan et al. (2015) empirical statistic yields the best performance among the four corrections in terms of controlling for Type I error rates and producing smaller mean differences. Particularly, when the sample size was larger than 400, the largest Type I error rate ($\alpha = .05$) produced by Yuan et al.'s (2015) empirical correction was 9.0% (Model ID 377). The only conditions under which the Yuan et al. (2015) empirical statistics failed to produce acceptable Type I error rates was when p was very large (i.e., $p \geq 90$) and N very small (i.e., $N = 200$), where the Yuan et al. (2015) empirical corrected statistics could reject the correctly specified model too often (Model IDs 43–66). However, even under these conditions, the Yuan et al. (2015) empirical corrected statistics seems to outperform the other corrections. For example, we can see that compared to the results from the other three corrections, the average values of the empirical distributions after applying the Yuan et al. (2015) empirical correction were closer to the corresponding theoretical means. It is also noted that under conditions in which sample size was small and q was extremely large, Yuan et al.'s empirical corrected statistic tended to overshrink the chi-square statistic (e.g., Model IDs 30, 36, 48, 60, 102, 114, and 126), which led to the empirical rejection rates being noticeably smaller than its nominal level.

DISCUSSION AND CONCLUSION

In this article, we first identified the unique effects among potential factors for model size effects on the likelihood ratio chi-square statistic distribution. The results showed that (a) bias in the likelihood ratio statistic is not substantially influenced by the number of latent factors (f) or the ratio of the observed variables to latent factors (p/f); (b) model size effects are associated with the number of observed variables (p), as generally a larger p is associated with a poor approximation to the asymptotic chi-square distribution; (c) model size effects also appear to be a function of the number of estimated parameters (q), such that increasing the number of estimated parameters reduces the inflated Type I error rates; and (d) p and q have unique effects on the likelihood ratio chi-square statistic distribution, even when the df is held constant. In Study II, we evaluated the performance of four proposed corrections for the model size effect, including the Bartlett-corrected statistic (1950), the Swain-corrected statistic (1975), the Yuan-corrected statistic (2005), and the Yuan et al. (2015) empirical corrected statistic. The results indicate that the Yuan et al. (2015) empirically corrected statistic generally yields the best performance in controlling the Type I error rates when fitting large SEM models.

The findings of this study expand on conclusions from previous methodological research, and also offer implications for empirical studies. First, consistent with Moshagen (2012), our results suggest that the number of observed variables (p) is the most important determinant of model size effects. When p becomes very large (e.g., $p \geq 60$), the empirical Type I error rates of the chi-square test statistics can be inflated dramatically (regardless of q), even with a relatively large sample size (e.g., $N = 2,000$). In practice, p

TABLE 5
Differences Between the Empirical and Theoretical Means (Δ_{mean}) and the 5% Empirical Rejection Rates of Five Likelihood Ratio Test Statistics

						No Correction		Bartlett (1950)		Swain (1975)		Yuan (2005)		Yuan et al. (2015)	
ID	N	p	q	f	df	Δ_{Mean}	5%	Δ_{Mean}	5%	Δ_{Mean}	5%	Δ_{Mean}	5%	Δ_{Mean}	5%
1	200	10	1	1	54	0.48	6.8%	-0.84	4.7%	-0.49	5.2%	-0.84	4.7%	-0.89	4.4%
2	200	10	12	1	43	0.48	6.9%	-0.57	5.4%	-0.39	5.9%	-0.57	5.4%	-0.61	5.3%
3	200	10	22	1	33	0.67	7.7%	-0.15	5.6%	-0.07	5.9%	-0.15	5.6%	-0.19	5.6%
4	200	10	32	1	23	0.32	7.2%	-0.24	5.9%	-0.22	6.0%	-0.24	5.9%	-0.27	5.9%
5	200	10	42	1	13	-0.06	5.2%	10.62	4.2%	-0.38	4.2%	-0.38	4.2%	-0.39	4.1%
6	200	10	52	1	3	0.02	6.0%	-0.05	5.4%	-0.06	5.4%	-0.05	5.4%	-0.05	5.4%
7	200	20	1	1	209	8.24	10.8%	-0.68	5.3%	0.76	6.1%	-0.68	5.3%	-1.15	4.9%
8	200	20	31	1	179	7.73	11.2%	0.07	5.2%	0.65	5.5%	0.07	5.2%	-0.43	4.9%
9	200	20	71	1	139	6.02	9.9%	0.07	5.1%	-0.02	5.1%	0.07	5.1%	-0.40	5.0%
10	200	20	111	1	99	3.73	8.4%	-0.48	4.3%	-0.87	3.8%	-0.48	4.3%	-0.88	3.8%
11	200	20	151	1	59	1.77	6.6%	-0.72	4.5%	-1.13	3.9%	-0.72	4.5%	-1.00	4.1%
12	200	20	191	1	19	0.09	4.9%	-0.69	3.9%	-0.87	3.7%	-0.69	3.9%	-0.79	3.9%
13	200	30	1	1	464	28.34	23.1%	-0.12	5.5%	3.15	6.1%	-0.12	5.5%	-1.88	4.9%
14	200	30	80	1	385	25.82	22.5%	2.08	5.7%	2.30	5.8%	2.08	5.7%	0.12	5.0%
15	200	30	160	1	305	18.65	18.3%	-0.06	4.8%	-1.46	4.5%	-0.06	4.8%	-1.99	4.3%
16	200	30	240	1	225	12.49	15.2%	-1.23	4.7%	-3.28	4.0%	-1.23	4.7%	-2.94	4.1%
17	200	30	320	1	145	7.14	10.8%	-1.65	4.7%	-3.55	3.7%	-1.65	4.7%	-2.93	4.2%
18	200	30	400	1	65	2.00	6.6%	-1.88	2.8%	-2.95	2.2%	-1.88	2.8%	-2.52	2.4%
19	200	40	1	1	819	64.65	45.0%	-1.21	5.3%	4.67	6.9%	-1.21	5.3%	-5.61	4.4%
20	200	40	135	1	685	61.89	48.5%	6.21	8.1%	5.31	7.8%	6.21	8.1%	0.99	6.5%
21	200	40	260	1	560	50.78	42.0%	5.26	7.4%	0.96	5.8%	5.26	7.4%	-0.17	5.5%
22	200	40	385	1	435	36.03	32.7%	0.92	5.2%	-4.81	3.5%	0.92	5.2%	-4.15	3.7%
23	200	40	510	1	310	21.67	19.4%	-3.05	4.8%	-8.64	2.9%	-3.05	4.8%	-7.25	3.3%
24	200	40	635	1	185	10.25	14.1%	-4.30	3.4%	-8.45	1.8%	-4.30	3.4%	-7.14	2.1%
25	200	50	1	1	1,274	133.44	80.8%	4.95	6.3%	14.34	8.2%	4.95	6.3%	-4.01	4.4%
26	200	50	180	1	1,095	128.64	82.3%	16.93	10.3%	14.54	9.2%	16.93	10.3%	5.84	6.2%
27	200	50	355	1	920	110.83	78.7%	16.72	10.0%	7.76	6.6%	16.72	10.0%	4.66	5.7%
28	200	50	530	1	745	87.85	70.7%	11.82	8.4%	-0.38	4.4%	11.82	8.4%	-0.12	4.4%
29	200	50	705	1	570	57.47	50.7%	0.19	4.5%	-12.42	2.5%	0.19	4.5%	-10.46	2.5%
30	200	50	880	1	395	33.43	32.0%	-5.68	3.0%	-16.47	0.7%	-5.68	3.0%	-14.09	1.0%
31	200	60	1	1	1,829	225.84	96.6%	3.83	5.2%	17.54	8.0%	3.83	5.2%	-12.11	2.6%
32	200	60	320	1	1,510	210.56	96.7%	24.67	12.1%	14.73	8.0%	24.67	12.1%	3.04	4.7%
33	200	60	430	1	1,400	191.30	95.1%	19.38	9.3%	4.55	4.5%	19.38	9.3%	-3.26	3.0%
34	200	60	630	1	1,200	166.50	94.2%	18.87	9.7%	-2.00	4.4%	18.87	9.7%	-4.69	3.8%
35	200	60	830	1	1,000	123.08	83.1%	1.75	5.6%	-21.55	1.3%	1.75	5.6%	-21.00	1.4%
36	200	60	1,030	1	800	69.68	52.0%	-24.28	0.7%	-46.78	0.1%	-24.28	0.7%	-44.51	0.1%
37	200	60	1	20	1,829	230.03	97.1%	-123.49	0.0%	21.31	9.0%	-57.96	0.2%	-8.40	2.8%
38	200	60	123	20	1,707	227.62	97.1%	-104.53	0.0%	21.15	8.8%	-42.96	0.9%	0.04	3.6%
39	200	60	223	20	1,607	221.20	97.5%	-92.69	0.0%	19.23	8.8%	-34.51	1.1%	3.37	4.5%
40	200	60	340	20	1,490	212.01	97.2%	-80.21	0.1%	17.14	8.0%	-26.05	1.7%	6.22	5.2%
41	200	60	471	20	1,359	199.11	97.4%	-68.40	0.1%	14.28	7.4%	-18.82	1.8%	7.64	5.9%
42	200	60	600	20	1,230	177.03	95.0%	-64.54	0.0%	4.74	5.2%	-19.76	1.3%	1.39	4.3%
43	200	90	1	1	4,094	842.86	100.0%	61.40	15.6%	94.39	26.7%	61.40	15.6%	2.50	4.3%
44	200	90	340	1	3,755	827.85	100.0%	102.43	31.8%	88.42	25.9%	102.43	31.8%	24.33	7.0%

45	200	90	660	1	3,435	779.89	100.0%	112.71	38.7%	67.53	19.5%	112.71	38.7%	20.56	**8.0%**
46	200	90	980	1	3,115	743.87	100.0%	133.04	50.9%	64.94	19.6%	133.04	50.9%	30.05	10.3%
47	200	90	1280	1	2,815	626.09	100.0%	81.39	26.8%	−0.09	**4.2%**	81.39	26.8%	−26.01	**2.0%**
48	200	90	1580	1	2,515	508.98	100.0%	30.31	10.9%	−58.49	0.7%	9.31	10.9%	−77.75	0.3%
49	200	90	1	30	4,094	836.33	100.0%	−423.09	0.0%	88.85	24.3%	−183.60	0.0%	−2.92	**3.9%**
50	200	90	230	30	3,865	824.62	100.0%	−373.32	0.0%	81.51	23.5%	−145.51	0.0%	10.16	**4.9%**
51	200	90	386	30	3,709	828.66	100.0%	−330.46	0.0%	91.25	27.0%	−110.04	0.1%	29.92	8.5%
52	200	90	555	30	3,540	811.03	100.0%	−300.41	0.0%	86.18	26.2%	−89.05	0.1%	34.06	9.3%
53	200	90	734	30	3,361	790.14	100.0%	−270.24	0.0%	81.68	25.4%	−68.60	0.3%	37.66	11.4%
54	200	90	906	30	3,189	758.92	100.0%	−249.55	0.0%	70.45	21.2%	−57.78	0.6%	33.04	10.0%
55	200	120	1	1	7,259	2227.82	100.0%	249.41	65.6%	312.84	81.8%	249.41	65.6%	96.67	19.5%
56	200	120	550	1	6,710	2202.95	100.0%	344.22	88.6%	297.16	81.5%	344.22	88.6%	126.95	29.2%
57	200	120	1,050	1	6,210	2138.69	100.0%	397.63	96.7%	277.09	78.5%	397.63	96.7%	131.19	30.8%
58	200	120	1,550	1	5,710	2048.03	100.0%	430.15	99.3%	253.55	75.3%	430.15	99.3%	124.08	31.2%
59	200	120	2,050	1	5,210	1880.89	100.0%	402.14	99.0%	185.81	55.5%	402.14	99.0%	68.94	17.1%
60	200	120	2,550	1	4,710	1592.54	100.0%	278.19	86.3%	39.93	11.6%	278.19	86.3%	−65.47	0.8%
61	200	120	1	40	7,259	2229.71	100.0%	−988.82	0.0%	314.35	82.2%	−368.96	0.0%	98.14	20.1%
62	200	120	234	40	7,026	2214.64	100.0%	−919.75	0.0%	298.63	80.6%	−316.09	0.0%	106.33	22.4%
63	200	120	549	40	6,711	2178.85	100.0%	−836.55	0.0%	278.18	76.9%	−255.81	0.0%	108.36	22.4%
64	200	120	800	40	6,460	2163.28	100.0%	−761.70	0.0%	278.83	78.9%	−198.37	0.0%	122.25	26.7%
65	200	120	1,050	40	6,210	2131.54	100.0%	−697.88	0.0%	271.54	77.2%	−152.95	0.2%	125.76	29.1%
66	200	120	1,287	40	5,973	2097.38	100.0%	−640.06	0.0%	265.31	77.2%	−112.85	0.3%	127.96	30.5%
67	400	10	1	1	54	0.79	**5.5%**	0.13	**4.8%**	0.30	**5.2%**	0.13	**4.8%**	0.10	**4.7%**
68	400	10	12	1	43	0.37	**5.6%**	−0.16	**4.4%**	−0.06	**4.9%**	−0.16	**4.4%**	−0.18	**4.3%**
69	400	10	22	1	33	0.31	**5.7%**	−0.09	**5.0%**	−0.05	**5.0%**	−0.09	**5.0%**	−0.11	**5.0%**
70	400	10	32	1	23	0.25	**4.9%**	−0.03	**4.6%**	−0.02	**4.6%**	−0.03	**4.6%**	−0.05	**4.6%**
71	400	10	42	1	13	**0.02**	**5.1%**	−0.14	**4.8%**	−0.14	**4.8%**	−0.14	**4.8%**	−0.15	**4.8%**
72	400	10	52	1	3	0.01	**5.8%**	−0.03	**5.5%**	−0.03	**5.4%**	−0.03	**5.5%**	−0.03	**5.4%**
73	400	20	1	40	209	4.47	**7.4%**	0.10	**5.5%**	0.80	**5.9%**	0.10	**5.5%**	−0.13	**5.5%**
74	400	20	31	40	179	4.31	8.2%	0.56	**6.0%**	0.84	**6.1%**	0.56	**6.0%**	0.32	**6.0%**
75	400	20	71	40	139	3.17	**7.2%**	0.26	**5.5%**	0.22	**5.5%**	0.26	**5.5%**	**0.03**	**5.5%**
76	400	20	111	40	99	1.88	**6.8%**	−0.19	**4.3%**	−0.37	**4.1%**	−0.19	**4.3%**	−0.38	**4.1%**
77	400	20	151	40	59	0.95	**5.9%**	−0.28	**4.1%**	−0.48	**3.8%**	−0.28	**4.1%**	−0.41	**3.8%**
78	400	20	191	40	19	−0.16	**4.5%**	−0.54	**3.9%**	−0.63	**3.7%**	−0.54	**3.9%**	−0.59	**3.8%**
79	400	30	1	1	464	13.41	11.1%	−0.35	**5.3%**	1.23	**5.7%**	−0.35	**5.3%**	−1.21	**5.3%**
80	400	30	80	1	385	12.32	10.9%	0.87	**5.9%**	0.97	**5.9%**	0.87	**5.9%**	−0.08	**5.7%**
81	400	30	160	1	305	9.41	9.8%	0.35	**4.8%**	−0.33	**4.8%**	0.35	**4.8%**	−0.59	**4.8%**
82	400	30	240	1	225	6.59	9.0%	−0.08	**4.6%**	−1.08	**4.3%**	−0.08	**4.6%**	−0.91	**4.3%**
83	400	30	320	1	145	3.37	**7.5%**	−0.90	**4.2%**	−1.83	**3.7%**	−0.90	**4.2%**	−1.53	**4.0%**
84	400	30	400	1	65	0.99	**6.0%**	−0.91	**4.3%**	−1.44	**3.7%**	−0.91	**4.3%**	−1.23	**4.1%**
85	400	40	1	1	819	31.96	20.9%	0.33	**5.4%**	3.15	**7.1%**	0.33	**5.4%**	−1.79	**4.9%**
86	400	40	135	1	685	30.19	21.8%	3.60	**6.4%**	3.17	**6.4%**	3.60	**6.4%**	1.10	**5.6%**
87	400	40	260	1	560	25.00	20.2%	3.25	**6.4%**	1.20	**5.6%**	3.25	**6.4%**	0.66	**5.2%**
88	400	40	385	1	435	18.79	17.7%	1.92	**6.0%**	−0.83	**5.5%**	1.92	**6.0%**	−0.52	**5.6%**
89	400	40	510	1	310	11.88	13.3%	−0.09	**5.6%**	−2.79	**4.4%**	−0.09	**5.6%**	−2.12	**4.6%**
90	400	40	635	1	185	5.11	9.1%	−1.96	**4.1%**	−3.97	**2.9%**	−1.96	**4.1%**	−3.34	**2.9%**
91	400	50	1	1	1,274	63.39	34.7%	2.49	**6.7%**	6.94	**7.9%**	2.49	**6.7%**	−1.75	**5.7%**
92	400	50	180	1	1,095	59.94	35.0%	7.35	**7.8%**	6.23	**7.1%**	7.35	**7.8%**	2.13	**6.0%**
93	400	50	355	1	920	51.49	31.7%	7.26	8.2%	3.05	**7.5%**	7.26	8.2%	1.59	**7.1%**
94	400	50	530	1	745	41.19	28.7%	5.40	**6.8%**	−0.35	**5.2%**	5.40	**6.8%**	−0.23	**5.2%**

(Continued)

TABLE 5
(Continued)

| ID | N | p | q | f | df | No Correction | | Corrections | | | | | | | |
| | | | | | | | | Bartlett (1950) | | Swain (1975) | | Yuan (2005) | | Yuan et al. (2015) | |
						Δ_{Mean}	5%	Δ_{Mean}	5%	Δ_{Mean}	5%	Δ_{Mean}	5%	Δ_{Mean}	5%
95	400	50	705	1	570	28.47	21.7%	1.22	**5.3%**	-4.78	**3.7%**	1.22	**5.3%**	-3.84	**4.0%**
96	400	50	880	1	395	17.43	15.0%	-1.35	**5.2%**	-6.53	**2.7%**	-1.35	**5.2%**	-5.38	**3.3%**
97	400	60	1	1	1,829	103.42	53.2%	0.23	**4.1%**	6.67	**5.1%**	0.23	**4.1%**	-7.25	**3.4%**
98	400	60	320	1	1,510	96.09	53.5%	9.54	**5.5%**	4.92	**4.6%**	9.54	**5.5%**	-0.53	**4.1%**
99	400	60	430	1	1,400	86.16	48.9%	6.08	**4.8%**	-0.82	**3.3%**	6.08	**4.8%**	-4.46	**3.0%**
100	400	60	630	1	1,200	73.80	43.4%	5.16	**6.0%**	-4.54	**3.7%**	5.16	**6.0%**	-5.79	**3.7%**
101	400	60	830	1	1,000	55.67	35.1%	-1.21	**4.3%**	-12.14	**2.5%**	-1.21	**4.3%**	-11.88	**2.6%**
102	400	60	1,030	1	800	33.07	21.0%	-11.82	**2.4%**	-22.57	1.6%	-11.82	**2.4%**	-11.49	1.6%
103	400	60	1	20	1,829	103.17	52.6%	-62.28	0.4%	5.49	**4.4%**	-31.61	1.1%	-8.42	**2.9%**
104	400	60	123	20	1,707	102.23	54.7%	-52.69	0.5%	5.93	**4.5%**	-23.97	1.4%	-3.92	**3.4%**
105	400	60	223	20	1,607	99.83	55.2%	-46.32	0.8%	5.79	**4.8%**	-19.23	1.9%	-1.59	**3.7%**
106	400	60	340	20	1,490	95.53	53.9%	-40.24	0.7%	5.01	**4.6%**	-15.07	1.8%	-0.08	**3.5%**
107	400	60	471	20	1,359	89.85	53.0%	-34.21	0.7%	4.13	**4.9%**	-11.22	**3.0%**	1.05	**4.3%**
108	400	60	600	20	1,230	80.33	48.8%	-31.88	1.2%	0.30	**3.8%**	-11.08	**2.5%**	-1.25	**3.7%**
109	400	90	1	1	4,094	364.75	98.9%	12.74	**6.9%**	27.60	9.0%	12.74	**6.9%**	-13.79	**3.9%**
110	400	90	340	1	3,755	356.42	99.2%	31.84	10.2%	25.57	9.3%	31.84	10.2%	-3.11	**5.3%**
111	400	90	660	1	3,435	334.23	98.7%	36.65	10.8%	16.50	**7.7%**	36.65	10.8%	-4.45	**4.3%**
112	400	90	980	1	3,115	317.25	98.9%	46.29	14.3%	16.08	**7.6%**	46.29	14.3%	0.60	**5.1%**
113	400	90	1,280	1	2,815	270.76	97.3%	27.15	10.4%	-9.30	**4.4%**	27.15	10.4%	-20.89	**2.9%**
114	400	90	1,580	1	2,515	222.21	91.9%	6.11	**6.5%**	-33.98	1.6%	6.11	**6.5%**	-42.67	1.3%
115	400	90	1	30	4,094	365.30	98.7%	-202.82	0.0%	28.12	9.1%	-94.78	0.2%	-13.28	**4.0%**
116	400	90	230	30	3,865	359.64	98.8%	-178.59	0.0%	25.76	9.9%	-76.24	0.8%	-6.29	**5.0%**
117	400	90	386	30	3,709	355.50	98.9%	-162.33	0.0%	26.07	9.9%	-63.86	1.1%	-1.33	**5.3%**
118	400	90	555	30	3,540	347.73	98.9%	-147.57	0.0%	24.71	10.1%	-53.38	1.3%	1.48	**5.6%**
119	400	90	734	30	3,361	339.01	98.7%	-132.38	0.0%	24.07	9.4%	-42.74	1.4%	4.50	**6.4%**
120	400	90	906	30	3,189	329.01	98.6%	-119.19	0.1%	23.03	10.0%	-33.96	1.8%	6.40	**6.6%**
121	400	120	1	1	7,259	892.27	100.0%	44.46	9.1%	71.64	14.1%	44.46	9.1%	-21.00	**3.9%**
122	400	120	550	1	6,710	879.43	100.0%	90.06	17.2%	70.07	14.0%	90.06	17.2%	-2.22	**5.2%**
123	400	120	1,050	1	6,210	851.34	99.9%	116.89	26.7%	66.04	13.8%	116.89	26.7%	4.49	**6.0%**
124	400	120	1,550	1	5,710	809.68	100.0%	131.57	32.3%	57.54	12.9%	131.57	32.3%	3.28	**5.5%**
125	400	120	2,050	1	5,210	741.77	100.0%	122.72	30.6%	32.17	8.7%	122.72	30.6%	-16.76	**3.5%**
126	400	120	2,550	1	4,710	634.99	100.0%	79.06	19.8%	-21.72	**3.8%**	79.06	19.8%	-66.30	1.5%
127	400	120	1	40	7,259	895.41	100.0%	-484.10	0.0%	74.46	14.6%	-218.42	0.0%	-18.21	**4.2%**
128	400	120	234	40	7,026	888.10	100.0%	-450.75	0.0%	69.68	14.3%	-192.90	0.0%	-12.46	**4.5%**
129	400	120	549	40	6,711	871.73	100.0%	-411.06	0.0%	63.16	13.3%	-164.01	0.1%	-9.08	**5.1%**
130	400	120	800	40	6,460	864.93	100.0%	-374.25	0.0%	66.58	13.7%	-135.59	0.1%	0.24	**5.4%**
131	400	120	1,050	40	6,210	849.94	100.0%	-344.41	0.0%	64.80	14.1%	-114.39	0.4%	3.26	**5.4%**
132	400	120	1,287	40	5,973	836.08	100.0%	-315.83	0.0%	65.15	14.5%	-93.98	0.5%	7.36	**5.4%**
133	600	10	1	1	54	0.77	**5.9%**	0.33	**5.5%**	0.45	**5.6%**	0.33	**5.5%**	0.32	**5.5%**
134	600	10	12	1	43	0.31	**5.5%**	-0.03	**4.7%**	0.03	**4.7%**	-0.03	**4.7%**	-0.05	**4.7%**
135	600	10	22	1	33	0.40	**4.6%**	0.13	**4.2%**	0.15	**4.4%**	0.13	**4.2%**	0.11	**4.2%**
136	600	10	32	1	23	0.13	**4.3%**	-0.06	**3.9%**	-0.05	**3.9%**	-0.06	**3.9%**	-0.07	**3.9%**
137	600	10	42	1	13	-0.06	**4.6%**	-0.16	**4.5%**	-0.16	**4.5%**	-0.16	**4.5%**	-0.17	**4.5%**
138	600	10	52	1	3	0.01	**4.9%**	-0.02	**4.9%**	-0.02	**4.9%**	-0.02	**4.9%**	-0.02	**4.9%**

139	600	20	1	1	209	3.82	8.0%	0.92	6.2%	1.39	6.6%	0.92	6.2%	0.77	6.1%
140	600	20	31	1	179	3.49	7.8%	1.00	6.3%	1.19	6.5%	1.00	6.3%	0.84	6.3%
141	600	20	71	1	139	2.48	7.1%	0.55	5.7%	0.52	5.7%	0.55	5.7%	0.40	5.6%
142	600	20	111	1	99	1.16	5.5%	-0.20	4.9%	-0.33	4.8%	-0.20	4.9%	-0.33	4.8%
143	600	20	151	1	59	0.65	5.2%	-0.16	3.9%	-0.29	3.8%	-0.16	3.9%	-0.25	3.8%
144	600	20	191	1	19	-0.06	5.4%	-0.32	4.7%	-0.38	4.5%	-0.32	4.7%	-0.35	4.6%
145	600	30	1	1	464	10.07	8.4%	0.96	4.6%	2.01	4.9%	0.96	4.6%	0.40	4.5%
146	600	30	80	1	385	9.28	8.7%	1.71	4.9%	1.78	4.9%	1.71	4.9%	1.09	4.5%
147	600	30	160	1	305	6.48	8.1%	0.50	4.0%	0.05	3.7%	0.50	4.0%	-0.12	3.7%
148	600	30	240	1	225	4.12	7.9%	-0.28	5.5%	-0.93	4.9%	-0.28	5.5%	-0.83	4.9%
149	600	30	320	1	145	2.25	7.1%	-0.58	4.6%	-1.19	4.3%	-0.58	4.6%	-0.99	4.4%
150	600	30	400	1	65	0.97	5.7%	-0.29	4.3%	-0.64	3.8%	-0.29	4.3%	-0.50	4.0%
151	600	40	1	1	819	21.48	13.6%	0.66	5.1%	2.52	6.2%	0.66	5.1%	-0.73	5.1%
152	600	40	135	1	685	21.24	13.6%	3.75	6.3%	3.75	6.2%	3.46	6.3%	2.10	5.7%
153	600	40	260	1	560	18.27	14.5%	3.95	6.8%	2.60	6.7%	3.95	6.8%	2.25	6.7%
154	600	40	385	1	435	14.08	12.1%	2.96	6.1%	1.14	5.4%	2.96	6.1%	1.35	5.5%
155	600	40	510	1	310	8.47	10.2%	0.59	5.7%	-1.20	4.9%	0.59	5.7%	-0.75	5.0%
156	600	40	635	1	185	3.71	6.8%	-0.96	4.2%	-2.29	3.4%	-0.96	4.2%	-1.87	3.7%
157	600	50	1	1	1,274	43.53	22.1%	3.57	5.4%	6.49	6.0%	3.57	5.4%	0.79	4.8%
158	600	50	180	1	1,095	40.47	21.5%	6.03	5.1%	15.30	4.9%	6.03	5.1%	2.61	4.4%
159	600	50	355	1	920	34.16	19.2%	5.22	6.4%	2.47	5.0%	5.22	6.4%	1.51	4.7%
160	600	50	530	1	745	27.61	17.5%	4.18	5.8%	0.42	5.0%	4.18	5.8%	0.50	5.0%
161	600	50	705	1	570	19.31	13.2%	1.44	4.7%	-2.49	3.4%	1.44	4.7%	-1.88	3.7%
162	600	50	880	1	395	12.17	10.7%	-0.18	4.7%	-3.58	3.5%	-0.18	4.7%	-2.83	3.9%
163	600	60	1	1	1,829	72.23	31.5%	3.99	5.7%	8.20	6.4%	3.99	5.7%	-0.91	4.6%
164	600	60	320	1	1,510	65.61	33.1%	9.06	6.4%	6.03	5.9%	9.06	6.4%	2.48	5.4%
165	600	60	430	1	1,400	59.66	29.9%	7.27	6.4%	2.75	5.8%	7.27	6.4%	0.37	5.3%
166	600	60	630	1	1,200	51.16	28.4%	6.25	5.5%	-0.10	4.8%	6.25	5.5%	-0.92	4.6%
167	600	60	830	1	1,000	38.64	22.0%	1.36	4.3%	-5.80	2.6%	1.36	4.3%	-5.63	2.6%
168	600	60	1,030	1	800	23.78	13.8%	-5.78	3.1%	-12.87	2.4%	-5.78	3.1%	-12.15	2.4%
169	600	60	1	20	1,829	71.53	30.4%	-36.88	1.2%	7.52	5.4%	-16.78	2.3%	-1.59	4.2%
170	600	60	123	20	1,707	71.04	32.4%	-30.38	1.2%	7.99	6.5%	-11.58	3.2%	1.55	4.8%
171	600	60	223	20	1,607	68.51	32.9%	-27.06	1.3%	7.02	6.2%	-9.35	3.0%	2.19	4.8%
172	600	60	340	20	1,490	65.17	31.6%	-23.54	1.5%	6.03	6.7%	-7.09	3.8%	2.70	5.9%
173	600	60	471	20	1,359	61.47	30.7%	-19.55	1.8%	5.49	6.5%	-4.54	4.7%	3.48	5.9%
174	600	60	600	20	1,230	55.27	27.3%	-18.04	2.4%	2.99	5.5%	-4.45	4.3%	1.97	5.2%
175	600	90	1	1	4,094	232.75	79.7%	5.22	6.8%	14.82	8.0%	5.22	6.8%	-11.93	4.4%
176	600	90	340	1	3,755	227.25	81.0%	17.83	8.3%	13.79	7.4%	17.83	8.3%	-4.71	4.8%
177	600	90	660	1	3,435	213.50	80.1%	21.63	9.6%	8.64	6.5%	21.63	9.6%	-4.87	5.2%
178	600	90	980	1	3,115	202.39	78.7%	27.94	11.6%	8.49	6.9%	27.94	11.6%	-1.48	5.7%
179	600	90	1,280	1	2,815	173.40	70.4%	16.25	8.9%	-7.26	5.5%	16.25	8.9%	-14.74	4.3%
180	600	90	1,580	1	2,515	142.11	61.8%	2.38	6.0%	-23.54	2.8%	2.38	6.0%	-29.16	2.4%
181	600	90	1	30	4,094	234.05	80.3%	-133.24	0.0%	16.06	7.1%	-63.40	1.4%	-10.70	4.1%
182	600	90	230	30	3,865	230.71	81.2%	-116.87	0.0%	15.10	7.0%	-50.77	1.7%	-5.61	4.5%
183	600	90	386	30	3,709	227.63	82.1%	-106.45	0.0%	15.09	7.0%	-42.92	2.4%	-2.58	4.6%
184	600	90	555	30	3,540	222.04	82.6%	-97.22	0.1%	13.82	6.8%	-36.51	2.2%	-1.15	4.6%
185	600	90	734	30	3,361	217.47	82.5%	-86.21	0.0%	14.58	6.9%	-28.46	2.3%	1.97	5.4%
186	600	90	906	30	3,189	210.58	81.4%	-77.92	0.0%	13.62	6.7%	-23.06	2.6%	2.92	5.0%
187	600	120	1	1	7,259	553.25	99.4%	20.37	6.6%	37.75	9.0%	20.37	6.6%	-21.46	4.1%
188	600	120	550	1	6,710	555.52	99.5%	52.15	11.3%	39.41	9.5%	52.15	11.3%	-6.69	4.8%

(Continued)

TABLE 5
(Continued)

ID	N	p	q	f	df	No Correction		Corrections							
								Bartlett (1950)		Swain (1975)		Yuan (2005)		Yuan et al. (2015)	
						Δ_{Mean}	5%	Δ_{Mean}	5%	Δ_{Mean}	5%	Δ_{Mean}	5%	Δ_{Mean}	5%
189	600	120	1,050	1	6,210	536.39	99.6%	68.99	15.8%	36.63	9.8%	68.99	15.8%	−2.54	**5.1%**
190	600	120	1,550	1	5,710	509.46	99.6%	78.56	18.2%	31.53	9.7%	78.56	18.2%	−2.95	**5.6%**
191	600	120	2,050	1	5,210	463.25	99.4%	72.06	15.9%	14.54	**7.2%**	72.06	15.9%	−16.54	**4.2%**
192	600	120	2,550	1	4,710	400.58	98.5%	46.51	12.8%	−17.68	**4.5%**	46.51	12.8%	−46.07	**2.0%**
193	600	120	1	40	7,259	560.61	99.5%	−320.57	0.0%	36.22	8.7%	−150.86	0.2%	−22.98	**3.8%**
194	600	120	234	40	7,026	555.67	99.6%	−298.69	0.0%	33.41	8.5%	−134.15	0.2%	−19.00	**3.9%**
195	600	120	549	40	6,711	544.71	99.5%	−272.92	0.0%	29.34	8.9%	−115.45	0.4%	−16.71	**3.8%**
196	600	120	800	40	6,460	540.16	99.7%	−248.67	0.0%	31.95	8.9%	−96.75	0.3%	−10.28	**4.7%**
197	600	120	1,050	40	6,210	530.35	99.6%	−229.20	0.0%	31.03	8.7%	−82.92	0.6%	−8.10	**4.1%**
198	600	120	1,287	40	5,973	521.96	99.7%	−209.94	0.0%	32.13	8.2%	−68.98	0.7%	−4.60	**3.8%**
199	800	10	1	1	54	1.07	**6.7%**	0.74	**6.3%**	0.83	**6.4%**	0.74	**6.3%**	0.73	**6.3%**
200	800	10	12	1	43	0.74	**6.2%**	0.48	**5.8%**	0.52	**5.8%**	0.48	**5.8%**	0.47	**5.8%**
201	800	10	22	1	33	0.47	**6.1%**	0.27	**5.5%**	0.29	**5.5%**	0.27	**5.5%**	0.26	**5.5%**
202	800	10	32	1	23	0.11	**5.0%**	−0.03	**4.7%**	−0.02	**4.8%**	−0.03	**4.7%**	−0.04	**4.7%**
203	800	10	42	1	13	−0.09	**5.1%**	−0.17	**4.9%**	−0.17	**4.9%**	−0.17	**4.9%**	−0.18	**4.9%**
204	800	10	52	1	3	**0.03**	**4.6%**	0.01	**4.5%**	0.01	**4.5%**	0.01	**4.5%**	0.01	**4.5%**
205	800	20	1	1	209	2.86	**6.7%**	0.67	**5.3%**	1.04	**5.8%**	0.67	**5.3%**	0.58	**5.3%**
206	800	20	31	1	179	2.49	**6.4%**	0.63	**5.8%**	0.77	**5.9%**	0.63	**5.8%**	0.51	**5.8%**
207	800	20	71	1	139	1.97	**6.5%**	0.52	**4.9%**	0.50	**4.9%**	0.52	**4.9%**	0.41	**4.8%**
208	800	20	111	1	99	1.10	**5.7%**	**0.07**	**4.3%**	−0.02	**4.3%**	**0.07**	**4.3%**	−0.02	**4.3%**
209	800	20	151	1	59	0.51	**4.8%**	−0.10	**4.1%**	−0.20	**3.9%**	−0.10	**4.1%**	−0.17	**4.0%**
210	800	20	191	1	19	−0.06	**5.1%**	−0.25	**4.8%**	−0.29	**4.6%**	−0.25	**4.8%**	−0.27	**4.7%**
211	800	30	1	1	464	6.34	**5.9%**	−0.43	**3.8%**	0.35	**3.8%**	−0.43	**3.8%**	−0.85	**3.7%**
212	800	30	80	1	385	5.73	**5.9%**	0.11	**3.5%**	0.16	**3.5%**	0.11	**3.5%**	−0.36	**3.4%**
213	800	30	160	1	305	4.11	**6.6%**	−0.34	**4.0%**	−0.63	**3.7%**	−0.34	**4.0%**	−0.80	**3.7%**
214	800	30	240	1	225	2.52	**5.9%**	−0.75	**4.3%**	−1.24	**4.2%**	−0.70	**4.3%**	−1.16	**4.2%**
215	800	30	320	1	145	1.37	**5.7%**	−0.74	**5.0%**	−1.19	**4.7%**	−0.74	**5.0%**	−1.04	**4.8%**
216	800	30	400	1	65	0.62	**5.3%**	−0.33	**4.8%**	−0.59	**4.6%**	−0.33	**4.8%**	−0.48	**4.6%**
217	800	40	1	1	819	15.67	10.5%	0.17	**5.6%**	1.56	**5.9%**	0.17	**5.6%**	−0.86	**5.4%**
218	800	40	135	1	685	14.73	11.4%	1.74	**6.2%**	1.53	**6.2%**	1.74	**6.2%**	0.52	**5.6%**
219	800	40	260	1	560	12.74	11.4%	2.11	**6.5%**	1.11	**6.2%**	2.11	**6.5%**	0.84	**6.0%**
220	800	40	385	1	435	9.64	8.9%	1.39	**5.2%**	**0.04**	**4.6%**	1.39	**5.2%**	0.20	**4.6%**
221	800	40	510	1	310	5.75	8.8%	−0.11	**4.9%**	−1.44	**4.2%**	−0.11	**4.9%**	−1.10	**4.2%**
222	800	40	635	1	185	2.63	**7.4%**	−0.86	**5.1%**	−1.85	**4.5%**	−0.86	**5.1%**	−1.53	**4.9%**
223	800	50	1	1	1,274	31.76	15.2%	2.07	**5.5%**	4.24	**5.8%**	2.07	**5.5%**	0.00	**5.4%**
224	800	50	180	1	1,095	29.69	15.0%	4.11	**6.3%**	3.57	**6.1%**	4.11	**6.3%**	1.57	**5.4%**
225	800	50	355	1	920	25.03	14.4%	3.55	**5.5%**	1.50	**5.0%**	3.55	**5.5%**	0.79	**4.9%**
226	800	50	530	1	745	21.03	13.7%	3.61	**6.4%**	0.82	**5.7%**	3.61	**6.4%**	0.88	**5.7%**
227	800	50	705	1	570	14.62	10.6%	1.33	**4.5%**	−1.60	**3.6%**	1.33	**4.5%**	−1.15	**3.7%**
228	800	50	880	1	395	9.30	8.2%	0.11	**4.6%**	−2.42	**3.4%**	0.11	**4.6%**	−1.86	**3.6%**
229	800	60	1	1	1,829	50.26	20.9%	−0.31	**4.6%**	2.82	**5.0%**	−0.31	**4.6%**	−3.94	**4.4%**
230	800	60	320	1	1,510	46.48	21.7%	4.60	**5.3%**	2.36	**5.1%**	4.60	**5.3%**	−0.28	**4.6%**
231	800	60	430	1	1,400	42.04	21.3%	3.23	**4.3%**	−0.11	**4.0%**	3.23	**4.3%**	−1.88	**3.7%**
232	800	60	630	1	1,200	36.35	19.3%	3.08	**5.8%**	−1.62	**5.1%**	3.08	**5.8%**	−2.23	**4.9%**

233	800	60	830	1	1,000	27.98	14.8%	0.32	**4.4%**	−4.99	**4.0%**	0.32	**4.4%**	−4.87	**4.0%**
234	800	60	1,030	1	800	16.92	10.4%	−5.07	**2.8%**	−10.33	1.9%	−5.07	**2.8%**	−9.80	1.9%
235	800	60	1	20	1,829	51.73	21.0%	−28.70	1.7%	4.25	**5.7%**	−13.79	**3.5%**	−2.51	**4.8%**
236	800	60	123	20	1,707	51.38	22.9%	−23.82	**2.5%**	4.64	**5.2%**	−9.88	**3.2%**	−0.14	**4.1%**
237	800	60	223	20	1,607	49.98	21.7%	−20.87	**2.5%**	4.39	**5.2%**	−7.74	**4.0%**	0.81	**4.6%**
238	800	60	340	20	1,490	48.25	21.7%	−17.53	**2.8%**	4.39	**5.9%**	−5.34	**4.3%**	1.92	**5.3%**
239	800	60	471	20	1,359	44.78	21.4%	−15.25	**2.7%**	3.30	**5.2%**	−4.12	**3.8%**	1.81	**4.7%**
240	800	60	600	20	1,230	39.61	19.3%	−14.68	**2.4%**	0.89	**5.3%**	−4.62	**4.3%**	0.14	**5.2%**
241	800	90	1	1	4,094	173.12	60.3%	4.90	**4.8%**	12.00	**5.3%**	4.90	**4.8%**	−7.78	**3.9%**
242	800	90	340	1	3,755	169.40	62.6%	14.69	**6.1%**	11.70	**5.7%**	14.69	**6.1%**	−1.97	**4.6%**
243	800	90	660	1	3,435	157.27	60.7%	15.65	**6.4%**	6.06	**4.8%**	15.65	**6.4%**	−3.91	**4.2%**
244	800	90	980	1	3,115	148.45	58.4%	19.80	**7.9%**	5.45	**6.1%**	19.80	**7.9%**	−1.90	**5.4%**
245	800	90	1,280	1	2,815	128.53	50.2%	12.48	**6.7%**	−4.88	**5.3%**	12.48	**6.7%**	−10.40	**4.5%**
246	800	90	1,580	1	2,515	106.93	41.9%	3.56	**5.5%**	−15.62	**3.5%**	3.56	**5.5%**	−19.77	3.1%
247	800	90	1	30	4,094	174.47	60.8%	−97.10	0.0%	13.29	**5.9%**	−45.45	1.5%	−6.50	**4.3%**
248	800	90	230	30	3,865	171.91	62.6%	−84.92	0.2%	12.59	**6.1%**	−36.08	1.4%	−2.70	**4.9%**
249	800	90	386	30	3,709	168.48	63.5%	−78.21	0.2%	11.54	**5.4%**	−31.30	1.7%	−1.51	**4.2%**
250	800	90	555	30	3,540	165.51	63.1%	−70.23	0.2%	11.76	**5.6%**	−25.40	**2.1%**	0.71	**4.3%**
251	800	90	734	30	3,361	161.90	63.1%	−62.23	0.4%	12.15	**5.8%**	−19.61	**2.2%**	2.85	**4.3%**
252	800	90	906	30	3,189	157.43	63.5%	−55.47	0.6%	12.09	**5.6%**	−14.98	**2.3%**	4.19	**4.2%**
253	800	120	1	1	7,259	407.98	94.8%	9.75	**6.2%**	22.52	**7.2%**	9.75	**6.2%**	−20.99	3.5%
254	800	120	550	1	6,710	402.61	95.3%	33.19	9.8%	23.83	8.3%	33.19	9.8%	−10.00	**4.2%**
255	800	120	1,050	1	6,210	389.03	95.2%	46.28	12.2%	22.55	8.3%	46.28	12.2%	−6.18	**4.8%**
256	800	120	1,550	1	5,710	370.62	95.0%	54.79	13.7%	20.32	8.7%	54.79	13.7%	−4.96	**4.9%**
257	800	120	2,050	1	5,210	339.92	94.1%	51.66	13.9%	9.49	7.0%	51.66	13.9%	−13.30	**4.7%**
258	800	120	2,550	1	4,710	294.16	90.2%	34.25	10.3%	−12.87	**4.8%**	34.25	10.3%	−33.71	2.6%
259	800	120	1	40	7,259	404.79	95.0%	−242.65	0.0%	19.50	**6.9%**	−117.96	0.4%	−24.00	3.5%
260	800	120	234	40	7,026	402.58	94.7%	−224.99	0.0%	18.95	**7.8%**	−104.13	0.5%	−19.55	**3.5%**
261	800	120	549	40	6,711	393.50	94.9%	−206.69	0.0%	−434.81	**7.2%**	−91.10	0.5%	−18.61	**3.7%**
262	800	120	800	40	6,460	389.61	94.9%	−189.05	0.0%	16.80	8.1%	−77.61	0.7%	−14.18	**4.4%**
263	800	120	1,050	40	6,210	383.16	95.5%	−173.83	0.0%	17.01	8.1%	−66.56	0.7%	−11.69	**4.6%**
264	800	120	1,287	40	5,973	376.96	95.1%	−159.49	0.0%	17.94	8.6%	−56.17	1.4%	−8.98	**4.9%**
265	1,000	10	1	1	54	0.85	**7.0%**	0.59	**6.6%**	0.66	**6.8%**	0.59	**6.6%**	0.58	**6.6%**
266	1,000	10	12	1	43	0.68	**5.5%**	0.47	**5.2%**	0.50	**5.2%**	0.47	**5.2%**	0.46	**5.2%**
267	1,000	10	22	1	33	0.57	**5.4%**	0.41	**5.0%**	0.42	**5.0%**	0.41	**5.0%**	0.40	**5.0%**
268	1,000	10	32	1	23	0.20	**5.7%**	**0.08**	**5.5%**	0.09	**5.5%**	**0.08**	**5.5%**	**0.08**	**5.4%**
269	1,000	10	42	1	13	0.10	**4.9%**	**0.04**	**4.7%**	0.04	**4.7%**	**0.04**	**4.7%**	**0.03**	**4.7%**
270	1,000	10	52	1	3	0.01	**4.9%**	−0.01	**4.8%**	−0.01	**4.8%**	−0.01	**4.8%**	−0.01	**4.8%**
271	1,000	20	1	1	209	2.87	**7.0%**	1.14	**5.5%**	1.42	**5.6%**	1.14	**5.5%**	1.05	**5.5%**
272	1,000	20	31	1	179	2.12	**5.5%**	0.64	**4.4%**	0.75	**4.6%**	0.64	**4.4%**	0.55	**4.4%**
273	1,000	20	71	1	139	1.54	**5.5%**	0.39	**4.6%**	0.37	**4.6%**	0.39	**4.6%**	0.30	**4.6%**
274	1,000	20	111	1	99	0.81	**4.9%**	−0.01	**4.6%**	−0.08	**4.6%**	−0.01	**4.6%**	−0.08	**4.6%**
275	1,000	20	151	1	59	0.37	**4.9%**	−0.11	**4.4%**	−0.19	**4.4%**	−0.11	**4.4%**	−0.17	**4.4%**
276	1,000	20	191	1	19	−0.02	**4.6%**	−0.18	**4.5%**	−0.21	**4.4%**	−0.18	**4.5%**	−0.20	**4.4%**
277	1,000	30	1	1	464	5.42	**6.4%**	0.01	**4.8%**	0.63	**4.8%**	0.01	**4.8%**	−0.32	**4.7%**
278	1,000	30	80	1	385	4.45	**6.2%**	−0.04	**4.3%**	0.00	**4.3%**	−0.04	**4.3%**	−0.41	**4.2%**
279	1,000	30	160	1	305	3.02	**6.6%**	−0.53	**5.2%**	−0.80	**5.1%**	−0.53	**5.2%**	−0.89	**5.1%**
280	1,000	30	240	1	225	1.20	**6.0%**	−1.40	**5.5%**	−1.79	**5.4%**	−1.40	**5.5%**	−1.73	**5.4%**
281	1,000	30	320	1	145	0.16	**5.3%**	−1.51	**4.6%**	−1.87	**4.5%**	−1.51	**4.6%**	−1.75	**4.5%**
282	1,000	30	400	1	65	**0.05**	**5.3%**	−0.70	**4.6%**	−0.91	**4.4%**	−0.70	**4.6%**	−0.83	**4.6%**

(Continued)

TABLE 5
(Continued)

| | | | | | | No Correction | | Corrections | | | | | | | | |
| | | | | | | | | Bartlett (1950) | | Swain (1975) | | Yuan (2005) | | Yuan et al. (2015) | |
ID	N	p	f	q	df	Δ_{Mean}	5%	Δ_{Mean}	5%	Δ_{Mean}	5%	Δ_{Mean}	5%	Δ_{Mean}	5%
283	1,000	40	1	1	819	14.29	8.5%	1.92	4.8%	3.02	4.9%	1.91	4.8%	1.09	4.7%
284	1,000	40	1	135	685	13.68	9.4%	3.31	5.0%	3.14	5.0%	3.31	5.0%	2.33	4.7%
285	1,000	40	1	260	560	12.23	9.3%	3.73	5.8%	2.93	5.6%	3.73	5.8%	2.72	5.6%
286	1,000	40	1	385	435	9.56	10.2%	2.96	6.8%	1.89	6.7%	2.96	6.8%	2.01	6.8%
287	1,000	40	1	510	310	5.90	8.8%	1.21	6.0%	0.15	5.7%	1.21	6.0%	0.41	5.8%
288	1,000	40	1	635	185	2.70	5.5%	-0.09	4.4%	-0.88	3.9%	-0.09	4.4%	-0.63	4.1%
289	1,000	50	1	1	1,274	26.83	12.9%	3.18	6.3%	4.90	6.7%	3.18	6.3%	1.53	6.0%
290	1,000	50	1	180	1,095	24.87	12.9%	4.50	6.7%	4.07	6.6%	4.50	6.7%	2.48	6.2%
291	1,000	50	1	355	920	21.42	12.2%	4.30	6.0%	2.67	5.6%	4.30	6.0%	2.10	5.1%
292	1,000	50	1	530	745	17.74	11.6%	3.87	5.8%	1.64	5.7%	3.87	5.8%	1.70	5.7%
293	1,000	50	1	705	570	11.75	8.2%	1.17	4.7%	-1.16	3.6%	1.17	4.7%	-0.80	3.6%
294	1,000	50	1	880	395	7.70	8.3%	0.38	4.8%	-1.64	3.8%	0.38	4.8%	-1.19	4.1%
295	1,000	60	1	1	1,829	43.12	19.6%	2.83	5.9%	5.32	6.4%	2.83	5.9%	-0.06	5.4%
296	1,000	60	1	320	1,510	39.53	18.5%	6.18	7.3%	4.40	6.6%	6.18	7.3%	2.30	6.2%
297	1,000	60	1	430	1,400	35.83	17.6%	4.93	7.1%	2.26	6.3%	4.93	7.1%	0.86	5.9%
298	1,000	60	1	630	1,200	31.16	17.5%	4.66	7.4%	0.92	6.5%	4.66	7.4%	0.43	6.3%
299	1,000	60	1	830	1,000	23.97	14.3%	1.93	6.4%	-2.30	4.8%	1.93	6.4%	-2.20	4.8%
300	1,000	60	1	1,030	800	14.28	10.6%	-3.25	5.2%	-7.44	3.8%	-3.25	5.2%	-7.02	4.3%
301	1,000	60	20	1	1,829	43.61	19.2%	-20.44	2.8%	5.79	6.7%	-8.57	4.2%	0.41	5.7%
302	1,000	60	20	123	1,707	43.44	18.8%	-16.43	3.4%	6.23	6.4%	-5.33	4.7%	2.42	5.4%
303	1,000	60	20	223	1,607	42.10	19.7%	-14.30	2.9%	5.81	5.9%	-3.84	3.7%	2.96	5.5%
304	1,000	60	20	340	1,490	40.09	18.5%	-12.24	3.4%	5.20	5.9%	-2.54	4.3%	3.24	5.7%
305	1,000	60	20	471	1,359	37.54	16.8%	-10.23	3.5%	4.54	5.7%	-1.37	4.3%	3.35	5.4%
306	1,000	60	20	600	1,230	33.30	16.1%	-9.91	3.3%	2.49	5.5%	-1.90	4.8%	1.89	5.4%
307	1,000	90	1	1	4,094	132.93	42.9%	-0.35	5.6%	5.28	5.9%	-0.35	5.2%	-10.39	4.2%
308	1,000	90	1	340	3,755	129.89	44.9%	7.40	7.0%	5.03	6.2%	7.40	7.0%	-5.79	4.9%
309	1,000	90	1	660	3,435	121.16	42.6%	9.03	6.9%	1.43	5.2%	9.03	6.9%	-6.46	4.6%
310	1,000	90	1	980	3,115	113.88	42.5%	12.06	7.6%	0.71	6.1%	12.06	7.6%	-5.10	5.0%
311	1,000	90	1	1,280	2,815	98.76	36.4%	6.88	5.7%	-6.86	3.9%	6.88	5.7%	-11.23	3.4%
312	1,000	90	1	1,580	2,515	81.23	31.1%	-0.63	3.9%	-15.82	2.8%	-0.63	3.9%	-19.11	2.7%
313	1,000	90	30	1	4,094	132.43	42.4%	-82.63	0.3%	4.79	5.9%	-41.74	1.7%	-10.88	4.2%
314	1,000	90	30	230	3,865	130.24	43.3%	-73.05	0.3%	4.13	5.6%	-34.39	1.6%	-7.97	4.1%
315	1,000	90	30	386	3,709	128.14	43.1%	-67.11	0.3%	3.93	4.8%	-29.98	2.4%	-6.40	4.0%
316	1,000	90	30	555	3,540	124.58	43.0%	-61.89	0.3%	2.97	5.4%	-26.43	2.8%	-5.78	4.4%
317	1,000	90	30	734	3,361	121.44	44.8%	-55.76	0.4%	3.05	4.7%	-22.06	2.8%	-4.31	4.2%
318	1,000	90	30	906	3,189	117.29	43.5%	-50.95	0.9%	2.44	5.3%	-18.95	2.4%	-3.80	4.6%
319	1,000	120	1	1	7,259	324.17	83.1%	9.15	4.9%	19.25	6.1%	9.15	4.9%	-15.17	3.7%
320	1,000	120	1	550	6,710	320.42	85.3%	28.37	7.9%	20.98	7.0%	28.37	7.9%	-5.77	4.4%
321	1,000	120	1	1,050	6,210	310.08	85.6%	39.22	9.3%	20.47	6.7%	39.22	9.3%	-2.23	4.6%
322	1,000	120	1	1,550	5,710	292.46	84.7%	43.11	11.7%	15.89	6.4%	43.11	11.7%	-4.07	4.4%
323	1,000	120	1	2,050	5,210	267.09	81.7%	39.56	11.9%	6.28	6.2%	39.56	11.9%	-11.70	4.2%
324	1,000	120	1	2,550	4,710	230.14	74.1%	24.92	9.7%	-12.28	4.3%	24.92	9.7%	-28.74	2.5%
325	1,000	120	40	1	7,259	324.84	83.6%	-187.58	0.0%	19.90	6.6%	-88.89	0.4%	-14.53	3.6%
326	1,000	120	40	234	7,026	322.36	83.9%	-174.15	0.0%	18.85	6.2%	-78.53	0.5%	-11.62	3.7%

327	1,000	120	549	40	6,711	361.20	84.1%	−158.61	0.1%	16.92	6.5%	−67.16	0.5%	−9.82	4.2%
328	1,000	120	800	40	6,460	313.92	85.2%	−143.78	0.0%	19.04	6.4%	−55.63	0.6%	−5.46	4.0%
329	1,000	120	1,050	40	6,210	307.57	84.3%	−132.81	0.1%	18.08	7.1%	−47.99	1.2%	−4.61	4.9%
330	1,000	120	1,287	40	5,973	302.18	84.8%	−121.82	0.1%	18.41	6.9%	−40.16	1.4%	−2.86	5.0%
331	2,000	10	1	1	54	0.76	5.8%	0.63	5.6%	0.66	5.7%	0.63	5.6%	0.62	5.6%
332	2,000	10	12	1	43	0.55	5.8%	0.44	5.7%	0.46	5.7%	0.44	5.7%	0.44	5.7%
333	2,000	10	22	1	33	0.33	5.2%	0.25	5.1%	0.25	5.1%	0.25	5.1%	0.24	5.1%
334	2,000	10	32	1	23	0.04	3.9%	−0.01	3.8%	0.00	3.8%	−0.01	3.8%	−0.02	3.8%
335	2,000	10	42	1	13	0.09	4.4%	0.06	4.4%	0.06	4.4%	0.06	4.4%	0.06	4.4%
336	2,000	10	52	1	3	0.13	5.8%	0.13	5.7%	0.13	5.7%	0.13	5.7%	0.13	5.7%
337	2,000	20	1	1	209	2.58	5.9%	1.71	5.6%	1.85	5.7%	1.71	5.6%	1.67	5.6%
338	2,000	20	31	1	179	2.16	5.6%	1.42	5.2%	1.48	5.2%	1.42	5.2%	1.37	5.2%
339	2,000	20	71	1	139	1.95	6.2%	1.38	5.6%	1.37	5.6%	1.38	5.6%	1.33	5.6%
340	2,000	20	111	1	99	1.23	5.8%	0.82	5.6%	0.78	5.6%	0.82	5.6%	0.78	5.6%
341	2,000	20	151	1	59	0.38	5.1%	0.14	5.1%	0.10	5.1%	0.14	5.1%	0.11	5.1%
342	2,000	20	191	1	19	−0.07	4.9%	−0.15	4.8%	−0.17	4.8%	−0.15	4.8%	−0.16	4.8%
343	2,000	30	1	1	464	3.84	6.4%	1.15	5.5%	1.46	5.7%	1.15	5.5%	0.98	5.5%
344	2,000	30	80	1	385	3.61	6.9%	1.38	5.1%	1.40	5.2%	1.38	5.1%	1.19	5.1%
345	2,000	30	160	1	305	2.94	5.9%	1.17	4.8%	1.03	4.7%	1.17	4.8%	0.98	4.7%
346	2,000	30	240	1	225	1.24	5.0%	−0.06	4.0%	−0.26	3.9%	−0.06	4.0%	−0.22	3.9%
347	2,000	30	320	1	145	0.08	6.0%	−0.76	4.9%	−0.94	4.9%	−0.76	4.9%	−0.88	4.9%
348	2,000	30	400	1	65	−0.12	4.6%	−0.49	4.0%	−0.59	4.0%	−0.49	4.0%	−0.55	4.0%
349	2,000	40	1	1	819	7.32	7.2%	1.19	5.0%	1.74	5.1%	1.19	5.0%	0.78	4.8%
350	2,000	40	135	1	685	7.43	6.6%	2.29	4.6%	2.20	4.6%	2.29	4.6%	1.81	4.4%
351	2,000	40	260	1	560	6.68	7.9%	2.47	5.8%	2.07	5.7%	2.47	5.8%	1.97	5.7%
352	2,000	40	385	1	435	5.18	7.5%	1.91	6.7%	1.38	6.2%	1.91	6.7%	1.44	6.2%
353	2,000	40	510	1	310	3.04	6.3%	0.72	4.9%	0.19	4.9%	0.72	4.9%	0.33	4.9%
354	2,000	40	635	1	185	1.47	5.1%	0.09	4.5%	−0.30	4.3%	0.09	4.5%	−0.18	4.3%
355	2,000	50	1	1	1,274	13.43	8.1%	1.73	5.2%	2.58	5.2%	1.73	5.2%	0.91	5.2%
356	2,000	50	180	1	1,095	11.90	8.3%	1.84	5.6%	1.62	5.6%	1.84	5.6%	0.84	5.6%
357	2,000	50	355	1	920	10.65	8.0%	2.20	5.3%	1.39	5.1%	2.20	5.3%	1.11	4.9%
358	2,000	50	530	1	745	8.04	6.5%	1.20	4.4%	0.10	4.2%	1.20	4.4%	0.13	4.2%
359	2,000	50	705	1	570	5.27	7.3%	0.04	5.8%	−1.11	5.2%	0.04	5.8%	−0.93	5.4%
360	2,000	50	880	1	395	3.53	7.1%	−0.09	5.0%	−1.09	4.3%	−0.09	5.0%	−0.87	4.4%
361	2,000	60	1	1	1,829	23.28	11.9%	3.36	6.6%	4.59	6.8%	3.36	6.6%	1.93	6.6%
362	2,000	60	320	1	1,510	21.11	10.7%	4.64	7.1%	3.76	7.0%	4.64	7.1%	2.73	6.2%
363	2,000	60	430	1	1,400	19.42	10.2%	4.15	6.2%	2.83	5.7%	4.15	6.2%	2.14	5.3%
364	2,000	60	630	1	1,200	16.81	9.8%	3.73	5.7%	1.88	5.1%	3.73	5.7%	1.64	5.1%
365	2,000	60	830	1	1,000	13.83	8.4%	2.92	5.6%	0.83	5.1%	2.92	5.6%	0.88	5.2%
366	2,000	60	1,030	1	800	9.06	7.4%	0.36	4.7%	−1.72	4.2%	0.36	4.7%	−1.51	4.3%
367	2,000	60	1	20	1,829	22.17	11.3%	−9.47	5.0%	3.49	6.5%	−3.60	5.4%	0.83	6.4%
368	2,000	60	123	20	1,707	21.43	10.7%	−8.11	4.2%	3.07	6.8%	−2.63	5.6%	1.19	6.5%
369	2,000	60	223	20	1,607	20.49	10.5%	−7.33	3.8%	2.59	6.3%	−2.17	5.2%	1.19	5.8%
370	2,000	60	340	20	1,490	19.27	11.2%	−6.52	4.3%	2.07	6.2%	−1.74	5.3%	1.11	6.2%
371	2,000	60	471	20	1,359	17.09	10.3%	−6.43	4.2%	0.84	5.1%	−2.07	4.7%	0.26	5.1%
372	2,000	60	600	20	1,230	14.74	9.8%	−6.54	4.0%	−0.44	5.2%	−2.60	4.5%	−0.73	5.2%
373	2,000	90	1	1	4,094	62.39	32.6%	−3.10	9.4%	−0.34	9.8%	−3.10	9.4%	−8.04	8.8%
374	2,000	90	340	1	3,755	61.11	34.1%	0.98	9.8%	−0.19	9.2%	0.98	9.8%	−5.50	7.8%
375	2,000	90	660	1	3,435	58.22	34.5%	3.18	9.6%	−0.55	8.7%	3.18	9.6%	−4.43	8.0%
376	2,000	90	980	1	3,115	55.21	32.8%	5.26	10.9%	−0.31	9.5%	5.26	10.9%	−3.17	8.8%

(Continued)

TABLE 5
(Continued)

ID	N	p	q	f	df	No Correction		Bartlett (1950)		Swain (1975)		Yuan (2005)		Yuan et al. (2015)	
						Δ_{Mean}	5%	Δ_{Mean}	5%	Δ_{Mean}	5%	Δ_{Mean}	5%	Δ_{Mean}	5%
377	2,000	90	1,280	1	2,815	48.05	31.7%	2.93	11.9%	−3.81	9.7%	2.93	11.9%	−5.96	9.0%
378	2,000	90	1,580	1	2,515	39.66	27.9%	−0.59	9.9%	−8.06	8.0%	−0.59	9.9%	−9.68	7.2%
379	2,000	90	1	30	4,094	61.15	16.9%	−44.51	1.6%	−1.56	5.0%	−24.42	2.9%	−9.26	4.3%
380	2,000	90	230	30	3,865	59.64	16.9%	−40.16	1.9%	−2.27	4.9%	−21.19	3.0%	−8.22	4.2%
381	2,000	90	386	30	3,709	58.49	17.1%	−37.31	1.8%	−2.46	5.1%	−19.10	3.2%	−7.53	4.3%
382	2,000	90	555	30	3,540	57.51	17.3%	−33.98	2.2%	−2.16	5.1%	−16.58	3.6%	−6.45	4.5%
383	2,000	90	734	30	3,361	55.65	17.2%	−31.23	2.1%	197.61	4.9%	−14.71	3.6%	−6.00	4.6%
384	2,000	90	906	30	3,189	53.35	16.7%	−29.10	2.4%	−2.94	4.8%	−13.42	3.3%	−6.00	4.1%
385	2,000	120	1	1	7,259	158.08	37.5%	4.10	5.6%	9.03	6.2%	4.10	5.6%	−7.79	4.5%
386	2,000	120	550	1	6,710	155.28	38.0%	12.75	6.4%	9.14	5.6%	12.75	6.4%	−3.91	5.0%
387	2,000	120	1,050	1	6,210	149.58	36.9%	17.55	7.5%	8.41	6.2%	17.55	7.5%	−2.66	4.9%
388	2,000	120	1,550	1	5,710	141.07	36.2%	19.60	8.2%	6.34	5.9%	19.60	8.2%	−3.38	4.8%
389	2,000	120	2,050	1	5,210	129.38	34.9%	18.53	6.9%	2.31	5.4%	18.53	6.9%	−6.45	4.8%
390	2,000	120	2,550	1	4,710	110.62	31.7%	10.54	5.6%	−7.60	3.7%	10.54	5.6%	−15.63	3.3%
391	2,000	120	1	40	7,259	155.35	35.4%	−95.01	0.6%	6.36	6.0%	−46.79	1.8%	−10.46	4.0%
392	2,000	120	234	40	7,026	153.23	35.5%	−89.19	0.8%	5.04	5.4%	−42.51	2.1%	−9.84	3.9%
393	2,000	120	549	40	6,711	150.83	36.2%	−80.87	0.9%	4.79	5.5%	−36.25	2.4%	−8.26	4.4%
394	2,000	120	800	40	6,460	149.53	37.7%	−73.65	1.1%	5.75	5.8%	−30.67	2.8%	−6.20	4.7%
395	2,000	120	1,050	40	6,210	147.55	37.7%	−67.12	1.0%	6.43	6.0%	−25.78	2.9%	−4.63	5.0%
396	2,000	120	1,287	40	5,973	143.82	37.5%	−62.73	1.6%	5.59	6.0%	−22.95	3.6%	−4.48	4.6%

Note. p = number of observed variables; q = number of estimated parameters; df = degrees of freedom; Δ_{Mean} = mean of the empirical chi-square statistics distribution minus degrees of freedom; 5% indicates the 5% empirical rejection rates.

and N can be used as indexes for generally evaluating the risk of inflated Type I error rates when using the ML-based chi-square test statistic. To securely reach acceptable Type I error rates, the sample size (N) should be much larger than the number of observed variables (p). Roughly speaking, our results suggest that $N \geq p^2$.

In addition, by manipulating q to a greater degree, we observed that increasing q could mitigate the inflation of Type I error rates, which conflicts with Moshagen's (2012) conclusion. Interestingly, the pattern we observed is the opposite of the conclusion reached by Jackson (2003), who found that a larger q is associated with a poor approximation to the asymptotic chi-square distribution. The divergence of conclusions could be due to the differences in the range of q manipulated. In Jackson's (2003) study design, the number of free parameters manipulated was somewhat restricted (from 2 to 40) and the overall effect of q on model fit was quite small.[4] According to our findings, although p is the dominant factor for the model size effects, the effect of q does not appear to be negligible after controlling for the influence of p. These findings suggest that both p and q should be accounted for in the formulation of correctional methods.

By including both p and q in its formulation, our results show that the Yuan et al. (2015) empirically corrected statistic generates the best performance (among the four existing corrections). Our results showed that Type I error rates under the Yuan et al. (2015) empirical correction are acceptable across a wide range of simulation conditions, except when p is very large ($p \geq 90$) and N is small (i.e., $N = 200$). In Yuan et al. (2015), the authors also noticed the same issue under a large $p{:}N$ ratio and concluded that the empirical correction can perform reasonably well whenever $N \geq \max(50, 2p)$. In this study, we observed that Type I error rates can still be inflated when p is extremely large (e.g., $p = 90$), even under the condition of $N \geq 2p$ (e.g., $N = 200$). For such conditions (e.g., $p = 90$), our results showed that a larger sample size (e.g., $N = 400$) is required to achieve reasonable Type I error rates. It seems that, for the Yuan et al. (2015) empirical correction, $N \geq 4p$ might be a more reasonable guideline.

Finally, we turn to some limitations of our study and future research directions. First, in our study, we only focused on correctly specified models and therefore only on the issue of Type I error rates. However, correctly rejecting misspecified models (i.e., power) is also important in evaluating model fit. Provided our findings that the Yuan et al. (2015) empirical correction could overcorrect the chi-square statistics under certain conditions (e.g., when q is extremely large), it would be interesting to investigate if applying the empirical correction to the chi-square test statistics would reduce its power (to reject misspecified models). Second, the likelihood ratio chi-square test has been criticized for having many undesirable characteristics. A major criticism is that the chi-square test is sensitive to the sample size (N). That is, researchers using the chi-square exact fit test are very likely to reject a model with a good fit because of a large sample size. Due to this limitation, in practice, the chi-square test is "not always the final word in assessing fit" (West, Taylor, & Wu, 2012, p. 211). Researchers have developed a large number of practical fit indexes in their attempts to rectify some of the problems with the chi-square exact fit test. Many fit indexes are directly based on the chi-square statistic (e.g., CFI; Bentler, 1990) and have been routinely reported in SEM applications as standard tools for evaluating model fit (McDonald & Ho, 2002; Mueller & Hancock, 2010). Future studies should explore the model size effect on practical model fit indexes, which can be practically more meaningful in empirical applications of SEM (Xing & Yuan, 2017).

FUNDING

REFERENCES

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*, 155–173. doi:10.1007/BF02294170

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*, 397–438. doi:10.1080/10705510903008204

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, *42*, 815–824. doi:10.1016/j.paid.2006.09.018

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology*, *3*, 77–85. doi:10.1016/j.paid.2006.09.018

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

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246. doi:10.1037/0033-2909.107.2.238

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

Browne, M. W. (1973). Generalized least squares estimators in the analysis of covariance structures. *ETS Research Report Series*, *1973*(1), 1–36.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*(1), 62–83. doi:10.1111/bmsp.1984.37.issue-1

Costa, P. T., & MacCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI–R) and NEO Five-Factor Inventory (NEO FFI): Professional manual*. Odessa, FL: Psychological Assessment Resources.

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. doi:10.1037/0033-2909.109.3.512

---

[4] For example, in Jackson (2003), for $p = 20$ and $N = 100$, the chi-square bias (i.e., relative difference, which is taking the difference between the mean of the empirical chi-square and its expected value [df] then dividing by the expected value) changed from 0.072 to -0.015 as q decreased from 40 to 2.

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(1), 16–29. doi:10.1037/1082-989X.1.1.16

Ding, L., Velicer, W. F., & Harlow, L. L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling*, 2, 119–143. doi:10.1080/10705519509540000

Fouladi, R. T. (2000). Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate non-normality. *Structural Equation Modeling*, 7, 356–410. doi:10.1207/S15328007SEM0703_2

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. doi:10.1080/10705510701301602

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling An overview and a meta-analysis. *Sociological Methods & Research*, 26, 329–367. doi:10.1177/0049124198026003003

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424–453. doi:10.1037/1082-989X.3.4.424

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. doi:10.1207/S15328007SEM1001_6

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183–202. doi:10.1007/BF02289343

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, 10, 333–351. doi:10.1207/S15328007SEM1003_1

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method* (Vol. 18). London: Butterworths.

Ling, G. (2012). Why the major field test in business does not report subscores: Reliability and construct validity evidence. *ETS Research Report Series*, 2012(1), i–26. doi:10.1002/j.2333-8504.2012.tb02293.x

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. doi:10.1207/s15327906mbr3302_1

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling*, 16, 439–476. doi:10.1080/10705510903008220

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, 7(1), 64. doi:10.1037/1082-989X.7.1.64

Moshagen, M. (2012). The model size effect in SEM: Inflated goodness-of-fit statistics are due to the size of the covariance matrix. *Structural Equation Modeling*, 19, 86–98. doi:10.1080/10705511.2012.634724

Mueller, R. O., & Hancock, G. R. (2010). Structural equation modeling. In Hancock, G. R., & Mueller, R. O. (Eds.). *The reviewer's guide to quantitative methods in the social sciences*. New York, NY: Routledge.

Muthén, L. K., & Muthén, B. O. (1998–2011). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

Saris, W. E., Satorra, A., & Van Der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling*, 16, 561–582. doi:10.1080/10705510903203433

Shi, D., Lee, T., & Terry, R. A. (2015). Revisiting the model size effect in structural equation modeling (SEM). *Multivariate Behavioral Research*, 50, 142. doi:10.1080/00273171.2014.989012

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

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling*. New York, NY: Guilford.

Xing, L., & Yuan, K. H. (2017, July). *Model evaluation with small N and/or large p*. Poster presented at the 2017 Annual Meeting of the Psychometric Society, Zurich, Switzerland.

Yuan, K.-H. (2005). Fit indices versus test statistics. *Multivariate Behavioral Research*, 40, 115–148. doi:10.1207/s15327906mbr4001_5

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., Tian, Y., & Yanagihara, H. (2015). Empirical correction to the likelihood ratio statistic for structural equation modeling with many variables. *Psychometrika*, 80, 379–405. doi:10.1007/s11336-013-9386-5