# Assessing fit in structural equation models:

# A Monte-Carlo evaluation of RMSEA vs. SRMR confidence intervals and tests of close fit

Alberto Maydeu-Olivares

Dexin Shi

University of South Carolina

Yves Rosseel

Ghent University

Maydeu-Olivares, A., Shi, D., & Rosseel, Y. (2018). Assessing Fit in Structural Equation Models: A Monte-Carlo Evaluation of RMSEA Versus SRMR Confidence Intervals and Tests of Close Fit. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 389–402. http://doi.org/10.1080/10705511.2017.1389611

## Author note

Correspondence concerning this article should be addressed to Alberto Maydeu-Olivares, Dept. of Psychology, University of South Carolina. Barnwell College. 1512 Pendleton St. Columbia, SC 29208. E-mail: amaydeu@sc.edu. This research was supported by the National Science Foundation under Grant No. SES-1659936. The authors acknowledge the Research Computing Center at the University of South Carolina and the High Performance Computing Center at Weifang University for providing the computing resources used to obtain the results reported in this paper.

# Abstract

We compare the accuracy of confidence intervals (CIs) and tests of close fit based on the RMSEA with those based on the SRMR. Investigations used normal and non-normal data with models ranging from $p = 10$ to 60 observed variables. CIs and tests of close fit based on the SRMR are generally accurate across all conditions (even at $p = 60$ with non-normal data). In contrast, CIs and tests of close fit based on the RMSEA are only accurate in small models. In larger models ($p \geq 30$), they incorrectly suggest that models do not fit closely, particularly if sample size is less than 500.

*Keywords*: goodness-of-fit indices, model selection,

# Assessing fit in structural equation models:

# A Monte-Carlo evaluation of RMSEA vs. SRMR confidence intervals and tests of close fit

Structural equation modeling refers to a general set of techniques used to estimate systems of equations, possibly involving latent variables. When the model under consideration does not involve constraints on the mean structure, the model parameters can be estimated by covariance structure modeling, that is, by minimizing a discrepancy function between the sample covariance matrix and the covariance matrix implied by the model. More specifically, let $\mathbf{\Sigma}$ denote the true and unknown population covariance matrix, and let $\mathbf{\Sigma}_0 \equiv \mathbf{\Sigma}(\boldsymbol{\theta})$ denote the covariance matrix specified by the null hypothesis, where $\boldsymbol{\theta}$ denotes the $q$ vector of model parameters to be estimated from the data. In covariance structure modeling, $\boldsymbol{\theta}$ is estimated by minimizing a discrepancy function $F(\mathbf{S},\mathbf{\Sigma}_0)$ where $\mathbf{S}$ denotes the sample covariance matrix. Different discrepancy functions can be employed to this aim. In this paper we focus on the normal theory maximum likelihood discrepancy function

$$F_{ML}(\mathbf{S},\mathbf{\Sigma}_0) = \ln\left|\mathbf{\Sigma}_0\right| - \ln\left|\mathbf{S}\right| + \operatorname{tr}\left(\mathbf{S}\mathbf{\Sigma}_0^{-1}\right) - p\,, \tag{1}$$

where $p$ denotes the number of variables being modeled. This is a discrepancy function suitable for models in which all the outcomes are continuous.

After a model has been fitted, it is necessary to assess the size of its misfit, that is, the magnitude of the discrepancy between the (unknown) data generating process and the model being fitted (i.e., the severity of the misspecification). This is critical in applications, as inferences drawn from poorly fitting models can be misleading (Saris, Satorra, & van der Veld, 2009). In covariance structure modeling, assessing the size of a model's misfit amounts to estimating the discrepancy between $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_0$ using an unstandardized, standardized, or relative effect size of model misfit (Maydeu-Olivares, 2017a). Effect sizes of model misfit are population parameters which can be estimated from the data.

Unstandardized effect sizes of overall misfit express the magnitude of a model's misfit in the original units of the sample statistics used to fit the model. In covariance structure modeling, an unstandardized effect size of a model's overall misfit is the Root Mean Square Residual (RMR)

$$RMR = \sqrt{\frac{1}{t}\sum_{i \le j}\left(\sigma_{ij} - \sigma_{ij}^{0}\right)^{2}} \; , \tag{2}$$

where $\sigma_{ij}$ denotes the population covariance between variables $i$ and $j$ (or variance if $i = j$) and $\sigma_{ij}^{0}$ denotes the population covariance (or variance) under the fitted model, and $t = p(p+1)/2$ denotes the number of non-redundant population variances and covariances. The RMR can be approximately interpreted as the average population residual covariance; however, the magnitude of a covariance is in general uninterpretable. As a result, it is unclear whether the magnitude of a particular value of the RMR, (e.g. RMR = 3), is large or small.

A popular unstandardized effect size of model misfit is the Root Mean Squared Error of approximation (RMSEA: Browne & Cudeck, 1993; Steiger, 1989, 1990)

$$RMSEA = \sqrt{\frac{F(\Sigma, \Sigma_0)}{df}} \tag{3}$$

where $F(\Sigma, \Sigma_0) = F_0$ denotes the discrepancy between $\Sigma$ and $\Sigma_0$. Thus, the RMSEA adjusts the discrepancy between $\Sigma$ and $\Sigma_0$ by the degrees of freedom of the model, $df = t - q$. The definition of the RMSEA given in (3) reveals that the RMSEA depends on the discrepancy used to estimate the model. There is not a unique population RMSEA. Rather, there are as many population RMSEA parameters as discrepancies can be used to estimate the model. For instance, when the ML fitting function (1) is used to estimate the model, the population RMSEA is

$$RMSEA_1 = \sqrt{\frac{\ln\left|\Sigma_0\right| - \ln\left|\Sigma\right| + \operatorname{tr}\left(\Sigma\Sigma_0^{-1}\right) - p}{df}} \; . \tag{4}$$

All possible RMSEA population parameters are unstandardized effect sizes of overall model misfit. This means that their magnitude is uninterpretable and they can only be judged in reference to cut-off values. Is an RMSEA = .3 a large or a small size of model misfit? We believe it is a small size of model misfit because Browne and Cudeck (1993, p. 144) pointed out that in their opinion, an RMSEA ≤ .08 "would indicate a reasonable error of approximation" and that they "would not want to employ a model with RMSEA greater than .1". But because the RMSEA is an unstandardized effect size, any population value (e.g. 0.05) has a different meaning depending on the structure and size of the model (Chen, Curran, Bollen, Kirby, & Paxton, 2008). See Savalaei (2012) for a thorough discussion of this issue.

In contrast, the magnitude of standardized and relative effect sizes of overall misfit can be meaningfully interpreted without the need of cut-off values. Standardized effect sizes overcome the problem of uninterpretability by expressing the magnitude of a model's misfit in standardized units. Relative effect sizes of overall misfit overcome the problem by expressing the magnitude of a model's misfit relative to the magnitude of the misfit of a baseline model.

A standardized effect size of overall misfit suitable for covariance structure models is the Standardized Root Mean squared Residual (SRMR):

$$SRMR = \sqrt{\frac{1}{t}\sum_{i \le j}\frac{\left(\sigma_{ij} - \sigma_{ij}^{0}\right)^{2}}{\sqrt{\sigma_{ii}\sigma_{jj}}}}\,. \tag{5}$$

The SRMR can approximately be interpreted as the average population standardized residual covariance. Another standardized effect size of overall misfit suitable for covariance structure models is the Correlation Root Mean squared Residual (CRMR):

$$CRMR = \sqrt{\frac{1}{t-p}\sum_{i<j}\left(\rho_{ij} - \rho_{ij}^{0}\right)^{2}}\,, \tag{6}$$

where $\rho_{ij}$ denotes the unknown population correlation between variables *i* and *j* and $\rho_{ij}^0$ denotes the population correlation under the fitted model. Standardized effect sizes of overall misfit are clearly preferable to unstandardized effect sizes as the magnitude of the former can be readily interpreted. For instance, a CRMR $= .01$ can be approximately interpreted as the average population residual correlation of the model being fitted. This average value will be judged by any researcher as very small. In contrast, an CRMR = .20 will be judged as unacceptably large.

Turning now to relative effect sizes of overall misfit, a suitable relative effect size of overall misfit is

$$\Gamma_1 = \frac{p}{\operatorname{tr}\left(\Sigma\Sigma_0^{-1}\right)^2}, \tag{7}$$

which can be described as a weighted population coefficient of determination of the fitted model (Steiger, 1989, p. 84). $\Gamma_1$ can also be described as the population counterpart of the Goodness of Fit Index (GFI) (Maiti & Mukherjee, 1990; Steiger, 1989). An alternative relative effect size of model misfit is (Bentler, 1990; Zhang & Savalei, 2016)

$$\Delta = 1 - \frac{F_0}{F_B} \tag{8}$$

where $F_0 = F(\Sigma, \Sigma_0)$ and and $F_B = F(\Sigma, \Sigma_B)$ denotes the discrepancy between $\Sigma$ and some baseline model $\Sigma_B$ (usually the independence model). $\Delta$ is the population Comparative Fit Index (CFI: Bentler, 1990).

In practice, researchers are to choose an effect size of overall model misfit and estimate a confidence interval for it. Using a parameter drift assumption, asymptotic statistical theory has been put forth (Browne & Cudeck, 1993; Steiger, 1989, 1990) that enables users to obtain a confidence interval for unstandardized effect sizes such as the population discrepancy function value ($F_0$), the non-centrality parameter ($\lambda = F_0 / N$, where

*N* denotes sample size), the Cross-Validation Index (CVI: Browne & Cudeck, 1993; Cudeck & Browne, 1983), or the RMSEA. These methods also enable obtaining confidence intervals for a relative effect size of overall misfit, the population Goodness of Fit Index (GFI) of Equation (7) (Maiti & Mukherjee, 1990; Steiger, 1989). It does not seem possible to apply these methods to obtain confidence intervals for the population Comparative Fit Index (CFI) of Equation (8) if the independence model is used as baseline. This is because the use of this model almost certainly violates the parameter drift assumptions used to derive confidence intervals using asymptotic methods (for a discussion, see Bentler, 1990). To date, the only effect size of overall misfit that is widely used in applications is the RMSEA.

Recently, Maydeu-Olivares (2017a) has provided statistical theory for obtaining confidence intervals for standardized effect sizes of overall misfit, such as the SRMR and the CRMR, as well as to perform tests of close fit (i.e., whether the population parameter is smaller than some arbitrary value). However, although the use of the SRMR/CRMR over the RMSEA to assess the effect size of model misfit is preferable on substantive (interpretation) grounds, if coverage rates for the RMSEA are substantially more precise than for the SRMR/CRMR, use of the RMSEA may be preferable on statistical grounds. It is therefore necessary to compare the accuracy of statistical procedures to obtain confidence intervals and tests of close fit for the SRMR/CRMR and the RMSEA. This is the aim of this paper.

The remainder of this paper is organized as follows: First, we summarize statistical theory for obtaining confidence intervals (CIs) for the RMSEA and the SRMR/CRMR. Both normal and non-normal outcomes are considered, as non-normal outcomes are the norm rather than the exception in applications. Next, we provide an example in which we obtain CIs for the RMSEA and the SRMR/CRMR to clarify our discussion. We provide the computer code and data used in this example as supplementary materials to this article. Then, we report the results of an extensive simulation study comparing the accuracy of confidence intervals, as well as of tests of close fit, for the RMSEA and SRMR/CRMR. All results were obtained using the *lavaan* package in R (R Development Core Team, 2015; Rosseel, 2012). We conclude with a discussion of the findings and suggestions for applied researchers.

## The Root Mean Square Error of Approximation (RMSEA)

Covariance structure models can be estimated using a variety of discrepancy functions, $F(\mathbf{S},\mathbf{\Sigma}_0)$, between the sample covariance matrix and the covariance matrix specified by the null hypothesis. One such function is the normal theory maximum likelihood discrepancy function given in Equation (1). When the fitted model is correct, for some combinations of discrepancy functions and distributional assumptions, the distribution of the estimated discrepancy function multiplied by sample size, $N\hat{F}$, can be approximated asymptotically using a chi-square distribution. This is the so-called chi-square test statistic in the structural equation modeling literature. Under the same conditions, $N\hat{F}$ can be approximated asymptotically using a non-central chi-square distribution when the fitted model is slightly incorrect (i.e. under parameter drift assumptions; for details on the conditions that must be met for $N\hat{F}$ to follow a non-central chi-square distribution see Browne & Cudeck, 1993; Maydeu-Olivares, 2017a). When $N\hat{F}$ can be approximated asymptotically by a non-central chi-square distribution, the population RMSEA can be estimated using the sample RMSEA (Steiger, 1989; Browne & Cudeck, 1993)

$$\widehat{RMSEA} = \sqrt{\max\left(\frac{N\hat{F} - df}{N \times df}, 0\right)}, \tag{9}$$

and confidence intervals for the population RMSEA can be obtained using a non-central chi-square distribution. For instance, when the model is estimated using the ML fitting function (1), $N\hat{F}$ is the likelihood ratio test statistic. For normally distributed data, the likelihood ratio test statistic follows asymptotically a chi square distribution when the model is correctly specified and follows a non-central chi-square distribution under parameter drift assumptions.

**Confidence intervals for the RMSEA**

Provided $N\hat{F}$ can be well approximated in large samples using a non-central chi-square distribution under parameter drift assumptions, a (100 – α)% confidence interval (CI) for the RMSEA is given by

$$\left(\sqrt{\frac{\hat{L}}{N \times df}};\sqrt{\frac{\hat{U}}{N \times df}}\right). \tag{10}$$

$\hat{L}$ and $\hat{U}$ are the solution to

$$F_{\chi^2}\left(N\hat{F};df,\hat{L}\right) = 1-\alpha/2, \quad \text{and} \quad F_{\chi^2}\left(N\hat{F};df,\hat{U}\right) = \alpha/2, \tag{11}$$

respectively, where $F_{\chi^2}\left(\cdot;df,\lambda\right)$ is the non-central chi-square distribution function with df degrees of freedom and non-centrality parameter λ.

Alternatively, we may wish to test whether the fitted model is a good enough approximation to the population covariance matrix. That is, we may be interested in the following test of close fit $H_0: RMSEA \leq RMSEA_0$ vs. $H_1: RMSEA > RMSEA_0$, where $RMSEA_0 > 0$ is an arbitrary value of the population RMSEA. In this case, the asymptotic *p*-value for the test of close fit is

$$p = 1 - F_{\chi^2}\left(N\hat{F};df, N \times df \times RMSEA_0^2\right). \tag{12}$$

**Estimation of the RMSEA with non-normal data**

When $N\hat{F}$ does not follow asymptotically a chi-square distribution under the null hypothesis, the sample RMSEA (9) is not a suitable estimator of the corresponding population parameter. For instance, when the model is estimated using unweighted least squares under normality assumptions, (9) cannot be used to estimate the population RMSEA as $N\hat{F}$ does not follow a chi-square distribution when the model is correct.

When data are not normal, the normal theory maximum likelihood discrepancy function (1) can still be used as it yields consistent estimates of the model parameters. However, standard errors (SEs) computed under normality are incorrect. Also, $N\hat{F}$ no longer follows asymptotically a chi-square distribution. Asymptotically correct SEs (the so called ‘robust’ SEs) can be obtained under the asymptotically distribution free (ADF) assumptions set forth by Browne (1982). ADF assumptions just require that all eighth-order moments of the distribution of the data are finite.

Under ADF assumptions, the most widely used goodness of fit test statistics when the normal theory maximum likelihood discrepancy function (1) is used are the mean corrected likelihood ratio test statistic $T_s = \hat{c}^{-1}N\hat{F}$, and the mean and variance adjusted likelihood ratio test statistic $T_a$, as described in Satorra and Bentler (1994). When the mean corrected test statistic is used, Li and Bentler (2006, see also Brosseau-Liard et al., 2012) showed that the ML population RMSEA (4) is to be estimated using

$$\widehat{RMSEA}_s = \sqrt{\max\left(\frac{\hat{F}}{df} - \frac{\hat{c}}{n}, 0\right)}\ , \tag{13}$$

instead of simply by applying equation (9) with $T_s = \hat{c}^{-1}N\hat{F}$ in lieu of $n\hat{F}$. That is, when the mean corrected statistic $T_s$ is used, the use of

$$\widehat{RMSEA}_{s2} = \sqrt{\max\left(\frac{T_s - df}{N \times df}, 0\right)}\ , \tag{14}$$

is incorrect, as it does not consistently estimate the population RMSEA, it estimates a different parameter. To the best of our knowledge, no theory has been proposed for how to estimate (4) when a Satorra-Bentler mean and variance adjusted test statistic is used. In the absence of statistical theory, structural equation modeling programs compute

$$\widehat{RMSEA}_a = \sqrt{\max\left(\frac{T_a - df}{N \times df}, 0\right)} \;, \tag{15}$$

but this formula is likely to be a poor estimate of the population ML RMSEA as no statistical theory supports its use.

## The Standardized Root Mean Squared Error (SRMR) and Correlation Root Mean Squared Error (CRMR)

Let $\mathbf{e}_s$ be the *t* vector of standardized residual covariances with elements $\frac{s_{ij} - \hat{\sigma}_{ij}}{\sqrt{s_{ii}s_{jj}}}$,

asymptotic covariance matrix $\Xi_s$. Maydeu-Olivares (2017a) showed that regardless of the discrepancy function and distributional assumptions used, an asymptotically unbiased estimate of the population SRMR (5), henceforth called SRMR*u*, is

$$\widehat{SRMR}_u = \hat{k}_s^{-1}\sqrt{\frac{\max\left(\mathbf{e}_s{}'\mathbf{e}_s - \mathrm{tr}(\hat{\Xi}_s), 0\right)}{t}}, \qquad \text{where} \qquad \hat{k}_s = 1 - \frac{\mathrm{tr}(\hat{\Xi}_s^2) + 2\mathbf{e}_s'\hat{\Xi}_s\mathbf{e}_s}{4\left(\mathbf{e}_s'\mathbf{e}_s\right)^2}. \tag{16}$$

Although the population SRMR is unique and its unbiased estimator (16) is unique, different estimates will be obtained depending on the discrepancy function used to estimate the model and on whether normality or ADF assumptions are used.

Structural equation modeling software compute a sample counterpart of the population SRMR (5), henceforth SRMR*b*,

$$\widehat{SRMR}_b = \sqrt{\frac{\mathbf{e}_s{}'\mathbf{e}_s}{t}} = \sqrt{\frac{1}{t}\sum_{i \le j}\frac{\left(s_{ij} - \hat{\sigma}_{ij}\right)^2}{\sqrt{s_{ii}s_{jj}}}} \;,. \tag{17}$$

However, this is a biased estimate of the population SRMR, and the magnitude of the bias increases as sample size decreases (Maydeu-Olivares 2017a).

Similarly, an unbiased estimate of the CRMR, henceforth CRMR*u*, is

$$\widehat{CRMR_u} = \hat{k}_r^{-1}\sqrt{\frac{\max\left(\mathbf{e}_r{}'\mathbf{e}_r - \mathrm{tr}(\hat{\Xi}_r),0\right)}{t-p}}, \quad \text{where} \qquad \hat{k}_r = 1 - \frac{\mathrm{tr}(\hat{\Xi}_r^2) + 2\mathbf{e}_r'\hat{\Xi}_r\mathbf{e}_r}{4\left(\mathbf{e}_r'\mathbf{e}_r\right)^2}. \tag{18}$$

In (18), $\mathbf{e}_r$ denotes the $t-p$ vector of residual correlations with elements $r_{ij} - \hat{\rho}_{ij}$, and $\Xi_r$ denotes its asymptotic covariance matrix. The sample CRMR, henceforth CRMR*b*, is

$$\widehat{CRMR_b} = \sqrt{\frac{\mathbf{e}_r{}'\mathbf{e}_r}{t-p}} = \sqrt{\frac{1}{t-p}\sum_{i<j}\left(r_{ij} - \hat{\rho}_{ij}\right)^2}, \tag{19}$$

and this is also a biased estimator of the population parameter (Maydeu-Olivares, 2017a).

### Confidence intervals for the SRMR and CRMR

Confidence intervals for the SRMR and CRMR population parameters (as well as tests of close fit) are obtained via the unbiased estimates of these parameters and their asymptotic standard errors using a reference normal distribution. In particular, in large samples, a (100 – α)% confidence interval for the SRMR can be obtained using

$$\Pr\left(\widehat{SRMR_u} - z_{\alpha/2}SE(\widehat{SRMR_u}) \le SRMR \le \widehat{SRMR_u} + z_{\alpha/2}SE(\widehat{SRMR_u})\right) = 1-\alpha, \tag{20}$$

where SE() denotes asymptotic standard error.

If a test of close fit based on the SRMR is desired, for example, of the type

$$H_0 : SRMR \le SRMR_0 \text{ vs. } H_1 : SRMR > SRMR_0, \tag{21}$$

where $SRMR_0 > 0$ is an arbitrary value of the SRMR, *p*-values are obtained using $p = 1 - \Phi(z)$, where $\Phi()$ denotes a standard normal distribution function and $z = \frac{\widehat{SRMR_u} - SRMR_0}{\mathrm{SE}(\widehat{SRMR_u})}$. Analogous procedures may be used with the CRMR.

The asymptotic standard error for the SRMR*u* and CRMR*u* are (Maydeu-Olivares, 2017a)

$$\mathrm{SE}(\widehat{SRMR_u}) = \sqrt{k_s^{-2}\frac{\mathrm{tr}(\Xi_s^2) + 2\mathbf{e}_s'\Xi_s\mathbf{e}_s}{2t\,\mathbf{e}_s'\mathbf{e}_s}}, \quad \mathrm{SE}(\widehat{CRMR_u}) = \sqrt{k_r^{-2}\frac{\mathrm{tr}(\Xi_r^2) + 2\mathbf{e}_r'\Xi_r\mathbf{e}_r}{2(t-p)\,\mathbf{e}_r'\mathbf{e}_r}}. \tag{22}$$

## A numerical example: Fitting a factor model to the short form of the Social Problem Solving Inventory-Revised (SPSI-R)

The short form of the SPSI-R (D'Zurilla, Nezu, & Maydeu-Olivares, 2002) is a 25 item questionnaire measuring five attributes, each by 5 items. Each item consist of five ordered categories and for this analysis, they were treated as continuous. We fitted a confirmatory five-factor model to the Spanish normative sample of this questionnaire (Maydeu-Olivares et al., 2000) matching the underlying theoretical model. Sample size is 741. Maximum likelihood estimation was used as implemented in the *lavaan* package in R (R Development Core Team, 2015; Rosseel, 2012). Maximum item kurtosis was 1.6; maximum item skewness was 1.1. Therefore, the model could be estimated under normality assumptions. In so doing, we obtained $X^2 = 902.68$ on 265 df. The sample RMSEA (9) estimate was .057. The 90% CI for the RMSEA population parameter (4) was (.053; .061), and the $p$-value for testing RMSEA $\leq .05$ was <.01. The sample (biased) SRMR (17) was .061.

For comparison, we also estimated the model under ADF assumptions. The estimated mean adjusted $X^2$ (Satorra & Bentler, 1994) was 775.37, the sample RMSEA (14) was .051 and the 90% CI for the RMSEA (4) was (.047; .055). The $p$-value for testing RMSEA $\leq .05$ was .34. The sample (biased) SRMR (17) is unaffected by the distributional assumptions used to estimate the model as only parameter estimates enter in its computation. Therefore, the sample SRMR estimate under ADF assumptions is also .061.

We note that although the data appears normally distributed, whether normality or ADF assumptions are used to assess the goodness of fit has a substantial effect on the results. For a model of this size and at this sample size, Maydeu-Olivares' (2017b) results suggest that even when data is normally distributed the mean corrected $X^2$ computed under ADF assumptions yields more accurate results than the $X^2$ computed under normality. Therefore, the results obtained using the mean corrected $X^2$ should be used. However, in this case, we should estimate the RMSEA using (13), and not (14), obtaining RMSEA = .055, CI = (.050; .059). There is a substantial difference when the correct formula is used to estimate the RMSEA using the mean corrected $X^2$ in this application.

Using *lavaan*, we can also obtain the unbiased estimates (16) and (18) as well as CIs for the population SRMR and CRMR parameters (5) and (6). The unbiased estimate of the SRMR (and CRMR) depends on distributional assumptions. Under normality assumptions, the unbiased estimate of the SRMR is .058, 90% CI = (.052, .063); under ADF assumptions, $SRMR_u$ = .057, 90% CI = (.050, .063). Also, under normality assumptions, $CRMR_u$ = .060, 90% CI = (.054, .066); under ADF assumptions, $CRMR_u$ = .059, 90% CI = (.052, .066).

To what extent can the estimated SRMR, CRMR and RMSEA confidence intervals be trusted? How accurate are they? Put differently, if a test of close fit is of interest (i.e., whether the SRMR, CRMR or RMSEA parameters are smaller than some arbitrary value), how accurate are the p-values of these tests? These are the main questions we try to address in this paper.

## Previous research on the accuracy of CIs and tests of close fit for the RMSEA and SRMR/CRMR

A number of simulation studies has examined the performance of point estimates, confidence intervals and tests of close fit using the RMSEA under maximum likelihood estimation (Curran, Bollen, Chen, Paxton, & Kirby, 2003; Fan, Thompson, & Wang, 1999; Hu & Bentler, 1998; Kenny, Kaniskan, & McCoach, 2015; Kenny & McCoach, 2003; Nasser & Wisenbaker, 2003; Nevitt & Hancock, 2000). However, with few exceptions they focus on normally distributed data, and the maximum number of observed variables considered was 15. For instance, Curran et al. (2003) examined the accuracy of CIs and *p*-values for tests of close fit for the RMSEA with normally distributed data. CIs for the RMSEA were found to be adequate for sample sizes larger than 200. However, only models with 8 and 15 variables were considered. Brosseau-Liard et al. (2012) described how to estimate the population RMSEA with non-normal data and examined the coverage of the RMSEA in models involving $p$ = 12 variables. None of these results are applicable to our example involving $p$ = 25 observed variables. Kenny and McCoach (2003) examined the behavior of the sample RMSEA in larger models (up to $p$ = 30) but they did not examine CIs or tests of close fit. In empirical research, often it is of interest to model a larger number of variables. Finally, as non-normal data are prevalent in many applications, researchers increasingly rely on robust standard errors and corrected goodness of fit test (Satorra & Bentler, 1994) –and there is evidence that they perform better than methods based on normality assumptions, even when data are normal (Maydeu-Olivares, 2017b).

A number of simulation studies have also examined the performance of the sample SRMR (Beauducel & Wittmann, 2005; Fan & Sivo, 2005, 2007; Garrido, Abad, & Ponsoda, 2016; Hu & Bentler, 1998, 1999) and the statistic has received praise as one of the most effective goodness of fit statistics (e.g., Hu & Bentler, 1998, 1999). Previous research on the accuracy of CIs and $p$-values for tests of close fit for the SRMR/CRMR is limited to Maydeu-Olivares (2017a) who investigated a few conditions involving normal as well as non-normal data. Only models with $p$ = 8 observed variables were considered. CIs for the SRMR/CRMR were found to be adequate except when population model misspecification was small (population SRMR = .025) and sample size was small ($N$ < 100).

## Simulation study

We performed an extensive simulation study to investigate the accuracy with which confidence intervals for the population RMSEA, SRMR, and CRMR can be estimated when a structural equation model for continuous outcomes is estimated by maximum likelihood estimation. The population parameters under consideration are given by Equations (3), (5), and (6). We also investigated the accuracy for *p*-values of tests of close fit for these parameters as applied users often report those instead of confidence intervals.

We simulated multivariate normal data with mean zero and a covariance structure conforming to an independent cluster model with two correlated factors with the same number of indicators per factor. In order to introduce non-normality, the continuous data were discretized into seven categories coded 0 to 6. Since the number of categories is large, it is appropriate to treat the discretized data as continuous when conducting data analysis (Muthén & Kaplan, 1985; Rhemtulla, Brosseau-Liard, & Savalei, 2012). Model misspecification was introduced by fitting a one-factor model (i.e., by ignoring the multidimensional structure of the data). We set the population values for all factor loadings to 0.70 and all error variances to 0.51. For both factors, their factor variances were set to one. The population values of the inter-factor correlations varied according to different levels of model misspecification as described below.

## Simulation Conditions

The simulation conditions were obtained by manipulating four variables:

1. *Sample size*. Sample sizes included 100, 200, 500, and 1,000.

2. *Model size*. Model size refers to the total number of observed variables ($p$), including small ($p = 10$), medium ($p = 30$), and large ($p = 60$) models.

3. *Degree of non-normality.* Three sets of skewness and (excess) kurtosis values were examined: skewness = 0.00, (excess) kurtosis = 0.00 (i.e. normal data); skewness = 0.00, kurtosis = 3.30; skewness = -2.00, kurtosis = 3.30. The item level non-normal data with desired skewness and kurtosis were generated by discretizing continuous data through selected thresholds (DiStefano & Morgan, 2014; Maydeu-Olivares, 2017b; Muthén & Kaplan, 1985). The threshold values used for each combination of skewness and kurtosis can

be found in Table 1. The computation of the population skewness and kurtosis is described in Maydeu-Olivares, Coffman and Hartmann (2007).

4. *Degree of model misspecification.* Three levels of model misspecification were obtained by manipulating the interfactor correlation in the population model: small ($\rho = 0.9$), medium ($\rho = 0.8$), and large ($\rho = 0.7$).

For every possible condition, the population values of the RMSEA, SRMR and CRMR were computed and values are reported in Table 1. Across conditions, population RMSEA values ranged from 0.023 to 0.103; population SRMR values ranged from 0.018 to 0.068; and population CRMR values ranged from 0.020 to 0.069. It is worth noting that the population RMSEA decreased as the inter-factor correlation increased, but also as model size increased (see Savalei, 2012 for more discussion). In addition, the population values for the non-normal conditions were slightly smaller than those obtained from normal data.

----------------------------------

Insert Table 1 around here

----------------------------------

In total, the simulation study consisted of a crossed design with 108 (4 × 3 × 3 × 3) conditions. For each simulated condition, 1,000 replications were generated using the *simsem* package in R (Pornprasertmanit, Miller, & Schoemann, 2012; R Development Core Team, 2015).

We investigated the effect of sample size because statistical theory for obtaining confidence intervals and *p*-values for tests of close fit is asymptotic. Therefore, results should be accurate if sample size is large enough, but it is necessary to investigate how large is “large enough” as applied researchers often use rather small samples. We also considered models ranging up to 60 observed variables to reflect conditions typically found in test development applications. In addition, observed data is often non-normal and we generated non-normal data with the skewness and kurtosis typically found in item level data. Finally, we considered a variety of values for each of the population parameters within the range typically reported in applications.

### *Data Analysis*

A one factor model was fitted in all cases with ML estimation using the *lavaan* package in R (R Development Core Team, 2015; Rosseel, 2012). All replications converged for all conditions. More specifically, for each replication, the population RMSEA, SRMR and CRMR were estimated under both normality and ADF assumptions. The 90% and 95% confidence intervals (CIs) were estimated for each parameter. The RMSEA was estimated under normality assumptions using (9) and under ADF assumptions using (13), that is via the Satorra-Bentler scaled test statistic. The SRMR and CRMR were estimated using (16) and (18). Finally, for each condition, we computed the *p*-value of a test of close fit in which the RMSEA and SRMR/CRMR equal its population value.

The outcomes of SRMR and CRMR were very similar to each other; therefore, we only report results for SRMR. However, all results observed for the SRMR are also applicable to the CRMR.

## Results

Table 2 provides the average of RMSEA and $\text{SRMR}_u$ across replications along with the population values. We have highlighted in this table those conditions where a relative bias less than 10% was obtained. We see in this table that when data is normally distributed, both the RMSEA and $\text{SRMR}_u$ provide, on average, estimates close to their population values, regardless of whether the estimates are obtained under normal theory or ADF assumptions. When non-normality was present, the estimates under normality assumptions could produce upwardly biased mean estimates, especially when sample size was small (e.g. kurtosis = 3.0, skewness = -2.0, $N = 100$). However, we note that the normal theory $\text{SRMR}_u$ is more robust than the normal theory RMSEA (i.e., bias is smaller) in the presence of non-normal data.

In general, both the ADF RMSEA and $\text{SRMR}_u$ provide unbiased estimates across all levels of model misspecification. However, the sample $\text{SRMR}_u$ converges faster to its population value than the sample RMSEA, especially when model size was large. That is, often a larger sample is needed for the sample RMSEA to reach a similar level of bias than for the $\text{SRMR}_u$. The use of $\text{SRMR}_b$ currently provided in many SEM software packages should be avoided, as it is severely biased upwards even in samples of size 1,000 (suggesting that the model fits more poorly than it does). We provide in Table A.1 in the supplementary materials to this article the average value $\text{SRMR}_b$ for each of the conditions of our study. We see in this table for instance that when kurtosis = 3, skewness = 2, $p = 60$, and population SRMR = .022, $\text{SRMR}_b = .043$ on average when $N = 500$, but it reaches an average of .087 when $N = 100$. In contrast, the average of the ADF $\text{SRMR}_u$ in these conditions is .022 and .028, respectively.

Tables 3 and 4 summarize the coverage rates for 90% and 95% confidence intervals around the population RMSEA and SRMR. For 90% CIs, coverage rates between 85% and 95% were considered acceptable and were highlighted in Table 3. We considered coverage rates between 90% and 99% for 95% CIs as acceptable and were highlighted in Table 4. Also, we have underlined in these tables the conditions in which coverage rates for the RMSEA are more accurate than for the SRMR. Similar patterns were observed across 90% and 95% CIs.

We see in these tables that, not surprisingly, CIs obtained under normality assumptions performed poorly with non-normal data. Under ADF assumptions, the coverage rates of CIs for the RMSEA were only acceptable when the model size was small. When fitting medium ($p = 30$) to large ($p = 60$) models, the coverage rates of RMSEA were noticeably lower than their nominal rates, even in large samples ($N = 1,000$). On the other hand, the CIs of robust SRMR could produce acceptable coverage rates across most simulation conditions, including cases where the number of observed variables was very large ($p = 60$). Note that the only situations where CIs for the RMSEA seemed to perform better than for the SRMR involved very small levels of model misspecification ($\rho = 0.9$) in small models ($p = 10$).

Table 5 summarizes the results of tests of close fit. More specifically, this table depicts the empirical rejection rates at a 5% significance level of tests of the RMSEA and SRMR being equal to their population values. These ranged from 0.023 to 0.103 for the RMSEA, and from 0.018 to 0.068 for the SRMR. We considered a statistic to be reasonably accurate if its rejection rate ranged between 0.01 and 0.10 at the 5% significance level; those cases are highlighted in Table 5. Note that for the SRMR, the 5% empirical rejection rates were equivalent to the non-coverage rates using 95% CIs (i.e. one minus the corresponding 95% converge rates in Table 4). This relationship does not hold for the RMSEA. Table 5 shows that for both the SRMR and RMSEA, the empirical rejection rates obtained under normality were close to its nominal rates only when small models ($p = 10$) were fitted to normal data. On the other hand the SRMR under ADF assumptions consistently yielded empirical rejection rates close to the nominal level (5%). The robust RMSEA, however, generally rejected the model too often as the number of observed variables reached 30 and beyond. For instance, for normally distributed data and $p = 30$, empirical rejection rates at the 5% significance level for testing whether RMSEA $\leq .054$ where .16 at $N = 500$ but .51 at $N = 100$. In contrast, empirical rejection rates at the 5% level for testing whether the equivalent SRMR $\leq .044$ where .08 at both $N = 500$ and $N = 100$.

## Discussion and conclusions

A great deal of research has been devoted to assess the degree of misfit of structural equation models, and applied researchers use an array of goodness of fit indices to this purpose. Goodness of fit indices are sample statistics used in conjunction with some fixed cutoff values that have been proposed in the literature (e.g., Hu & Bentler, 1999). If a goodness of fit index meets the recommended cutoff values, the model is retained (Barrett, 2007; Marsh, Hau, & Wen, 2004); otherwise, it is rejected. There are several problems with these practices (Maydeu-Olivares, 2017); for instance: a) the parameter being estimated (the effect size of model misfit) is often not described, nor is care taken so that the sample statistic consistently estimates the population parameter, b) the sampling variability of the sample statistic (the goodness of fit index) is ignored.

In this paper we distinguish between effect sizes of model misfit and goodness of fit indices. Effect sizes of model misfit are population parameters that capture the discrepancy between the fitted model and the data generating process for which statistical theory is available, thus enabling the construction of confidence intervals, and, if of interest, statistical tests (Maydeu-Olivares, 2017). We reserve the term goodness of fit index to refer to sample statistics used to adjudge model fit disregarding their sampling variability and without referencing any population parameter. Good overviews of the array of goodness of fit indices that have been proposed in the literature are Bollen and Long (1993) and Marsh, Hau and Grayson (2005). In applications, the only effect size of model misfit that is widely used is the RMSEA.

A drawback of the RMSEA is that it is an unstandardized effect size. As a result, population values of the RMSEA cannot be substantively interpreted. Their magnitude can only be judged in reference to cut-off values. But precisely because the RMSEA is an unstandardized effect size, any population value (say 0.05) has a different meaning depending on the structure and size of the model: For instance, most researchers would consider a one factor model to be a close approximation to a two factor model whose factors correlate .9. Fewer researchers would consider a one factor model to be a close approximation when the two factors correlate .8. However, the first combination of true and fitted model yields a population RMSEA of .042 when $p = 10$ and data is normal, but the second combination yields a population RMSEA of .037 when $p = 60$, kurtosis = 3 and skewness = -2 (see Table 1). Thus, a smaller population RMSEA does not necessarily mean a more closely fitting model. To overcome this problem Maydeu-Olivares (2017a) has advocated the use of standardized effect sizes of model misfit such as the Standardized Root Mean Squared Error (SRMR) or the Correlation Root Mean Squared Error (CRMR). Unlike the RMSEA, the SRMR or the CRMR can be substantively interpreted and therefore can convey the degree of misfit without the need of cut-off values. Statistical theory is available to obtain confidence intervals for the SRMR and CRMR parameters, as well as to perform tests of close fit (i.e., whether the population parameter is smaller than some arbitrary value). In this article we report the results of an extensive simulation study comparing the accuracy with which CIs for these population parameters can be obtained, using both normal and non-normal data.

Our simulation results show that across almost every condition tested, confidence intervals and test of close fit based on the $\text{SRMR}_u$ are more accurate than those based on the RMSEA. The latter fail in models involving a medium or large number of variables, and generally fail when data are non-normal. In fact, confidence intervals and test for the RMSEA only outperform those for the SRMR when the number of variables is small ($p = 10$), the degree of misfit is small (a one factor model is fitted to a model with two factors that correlate 0.9), and generally in small samples ($N = 100, 200$).

Our conclusions are somewhat limited by the conditions of this study. For instance, the accuracy of RMSEA CIs noticeably worsens from 10 to 30 observed variables, but we do not know at what model size they remain adequate, as we did not include conditions with between 10 and 30 variables in our study. In addition, only one type of misspecification was considered (misspecification of the interfactor correlations). To explore whether the results presented here would generalize to other setups we performed some additional simulation studies where we misspecified confirmatory factor models by omitting cross-loadings. Results are presented as supplementary materials and are consistent with the results presented in this paper. This leads us to conjecture that holding effect size of misfit, sample size, degree of non-normality, and model size fixed, the accuracy of CIs for the SRMR and RMSEA does not depend on the source of misspecification (e.g. misspecified dimensionality, omitted cross loadings, etc. ).

What do the simulation results inform us about the results obtained in our numerical example? Given our model size and our sample, it is likely that the estimated RMSEA is slightly larger than it should be and that its estimated CI is narrower than it should be. As a result, the RMSEA results incorrectly make the fit of our model appear poorer than what it is. In contrast, the unbiased estimate of the SRMR and its associated CI accurately convey the degree of misfit of our fitted model.

Why do confidence intervals for the RMSEA fail in the setup considered in this paper? Because the RMSEA is based on the likelihood ratio (LR) test statistic and therefore it will fail whenever the asymptotic approximation to the empirical sampling distribution of the LR fails. The asymptotic approximation to the distribution of the likelihood ratio test is known to fail in correctly specified models with a large number of variables (Herzog, Boomsma, & Reinecke, 2007; Kenny & McCoach, 2003; Shi, Lee, & Terry, 2017) even when data is normally distributed. In addition, when data is non-normal, only the mean adjusted LR can be used to estimate the population RMSEA as no statistical theory exists for estimating the RMSEA using the mean and variance adjusted LR, which is known to yield better results than the mean adjusted LR (Maydeu-Olivares, 2017b). As a result, confidence intervals and tests for the RMSEA fail in models with large number of variables (probably over 15 or 20) particularly with non-normal data.

Fortunately, the results presented here show that assessing the degree of misfit of a SEM model can be safely assessed using the SRMR, even in large models and non-normal data. In closing, the use of the sample SRMR currently provided by most software programs is to be avoided as it overestimates the population SRMR and suggests that the model fits more poorly than it does.

# References

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, *42*(5), 815–824. http://doi.org/10.1016/j.paid.2006.09.018

Beauducel, A., & Wittmann, W. W. (2005). Simulation Study on Fit Indexes in CFA Based on Data With Slightly Distorted Simple Structure. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(1), 41–75. http://doi.org/10.1207/s15328007sem1201_3

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. http://doi.org/10.1037/0033-2909.107.2.238

Bollen, K. A., & Long, J. S. (1993). *Testing Structural Equation Models*. Newbury Park, CA: Sage.

Brosseau-Liard, P. E., Savalei, V., & Li, L. (2012). An Investigation of the Sample Performance of Two Nonnormality Corrections for RMSEA. *Multivariate Behavioral Research*, *47*(6), 904–930. http://doi.org/10.1080/00273171.2012.715252

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141). Cambridge, UK: Cambridge University Press.

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., 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*(4), 462–494. http://doi.org/10.1177/0049124108314720

Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research*, *18*, 147–167. http://doi.org/10.1207/s15327906mbr1802

Curran, P. J., Bollen, K. A., Chen, F., Paxton, P., & Kirby, J. B. (2003). Finite Sampling

Properties of the Point Estimates and Confidence Intervals of the RMSEA. *Sociological Methods & Research*, *32*(2), 208–252. http://doi.org/10.1177/0049124103256130

D'Zurilla, T. J., Nezu, A. M., & Maydeu-Olivares, A. (2002). Manual of the Social Problem-Solving Inventory-Revised. North Tonawanda, NY: Multi-Health Systems, Inc.

DiStefano, C., & Morgan, G. B. (2014). A Comparison of Diagonal Weighted Least Squares Robust Estimation Techniques for Ordinal Data. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(3), 425–438. http://doi.org/10.1080/10705511.2014.915373

Fan, X., & Sivo, S. A. (2005). Sensitivity of Fit Indexes to Misspecified Structural or Measurement Model Components: Rationale of Two-Index Strategy Revisited. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(3), 343–367. http://doi.org/10.1207/s15328007sem1203_1

Fan, X., & Sivo, S. A. (2007). Sensitivity of Fit Indices to Model Misspecification and Model Types. *Multivariate Behavioral Research*, *42*(3), 509–529. http://doi.org/10.1080/00273170701382864

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 56–83. http://doi.org/10.1080/10705519909540119

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2016). Are Fit Indices Really Fit to Estimate the Number of Factors With Categorical Variables ? Some Cautionary Findings via Monte Carlo Simulation. *Psychological Methods*, *21*(1), 93–111. http://doi.org/10.1037/met0000064

Herzog, W., Boomsma, A., & Reinecke, S. (2007). The Model-Size Effect on Traditional and Modified Tests of Covariance Structures. *Structural Equation Modeling: A*

*Multidisciplinary Journal*, *14*(3), 361–390. http://doi.org/10.1080/10705510701301602

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453. http://doi.org/10.1037//1082-989X.3.4.424

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. http://doi.org/10.1080/10705519909540118

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The Performance of RMSEA in Models With Small Degrees of Freedom. *Sociological Methods & Research*, *44*(3), 486–507. http://doi.org/10.1177/0049124114543236

Kenny, D. A., & McCoach, D. B. (2003). Effect of the Number of Variables on Measures of Fit in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*(3), 333–351. http://doi.org/10.1207/S15328007SEM1003_1

Li, L., & Bentler, P. M. (2006). *Robust statistical tests for evaluating the hypothesis of close fit of misspecified mean and covariance structural models* (UCLA Statistics Preprint No. 506). Los Angeles, CA.

Maiti, S. S., & Mukherjee, B. N. (1990). A note on distributional properties of the Jöreskog-Sörbom fit indices. *Psychometrika*, *55*(4), 721–726. http://doi.org/10.1007/BF02294619

Marsh, H. W., Hau, K., & Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(3), 320–341. http://doi.org/10.1207/s15328007sem1103_2

Maydeu-Olivares, A. (2017a). Assessing the Size of Model Misfit in Structural Equation Models. *Psychometrika*, *82*(3), 533–558. http://doi.org/10.1007/s11336-016-9552-7

Maydeu-Olivares, A. (2017b). Maximum likelihood estimation of structural equation models for continuous data: Standard errors and goodness of fit. *Structural Equation Modeling*, *24*(3), 383–394. http://doi.org/10.1080/10705511.2016.1269606

Maydeu-Olivares, A., Coffman, D. L., & Hartmann, W. M. (2007). Asymptotically distribution-free (ADF) interval estimation of coefficient alpha. *Psychological Methods*, *12*(2), 157–176. http://doi.org/10.1037/1082-989X.12.2.157

Maydeu-Olivares, A., Rodriguez-Fornells, A., Gomez-Benito, J., D'Zurilla, T. J., Gómez-Benito, J., & D'Zurilla, T. J. (2000). Psychometric properties of the Spanish adaptation of the Social Problem-Solving Inventory-Revised (SPSI-R). *Personality and Individual Differences*, *29*(4), 699–708. http://doi.org/10.1016/S0191-8869(99)00226-3

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*(2), 171–189. http://doi.org/10.1111/j.2044-8317.1985.tb00832.x

Nasser, F., & Wisenbaker, J. (2003). A Monte Carlo Study Investigating the Impact of Item Parceling on Measures of Fit in Confirmatory Factor Analysis. *Educational and Psychological Measurement*, *63*(5), 729–757. http://doi.org/10.1177/0013164403258228

Nevitt, J., & Hancock, G. R. (2000). Improving the Root Mean Square Error of Approximation for Nonnormal Conditions in Structural Equation Modeling. *The Journal of Experimental Education*, *68*(3), 251–268. http://doi.org/10.1080/00220970009600095

Pornprasertmanit, S., Miller, P., & Schoemann, A. M. (2012). R packagesimsem: SIMulated structural equation modeling. Available from the Comprehensive R Archive Network: http://cran. r-project. org.

R Development Core Team. (2015). R: A Language and Environment for Statistical Computing. Vienna, Austria.

Rhemtulla, M., Brosseau-Liard, P. É. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical

SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–73. http://doi.org/10.1037/a0029315

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. http://doi.org/10.18637/jss.v048.i02

Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing Structural Equation Models or Detection of Misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 561–582. http://doi.org/10.1080/10705510903203433

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. (2012). The Relationship Between Root Mean Square Error of Approximation and Model Misspecification in Confirmatory Factor Analysis Models. *Educational and Psychological Measurement*, *72*(6), 910–932. http://doi.org/10.1177/0013164412452564

Shi, D., Lee, T., & Terry, R. A. (2017). Revisiting the Model Size Effect in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *50*(1), 1–20. http://doi.org/10.1080/10705511.2017.1369088

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

Steiger, J. H. (1990). Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behavioral Research*, *25*(2), 173–180. http://doi.org/10.1207/s15327906mbr2502_4

Zhang, X., & Savalei, V. (2016). Bootstrapping Confidence Intervals for Fit Indexes in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(3), 392–408. http://doi.org/10.1080/10705511.2015.1118692

Table 1. *Population thresholds, skewness, kurtosis, RMSEA, SRMR, and CRMR*

| **Kur.** | **Skew.** | **Thresholds** | **p** | **ρ** | **RMSEA** | **SRMR** | **CRMR** |
|---|---|---|---|---|---|---|---|
| 0.0 | 0.0 | -1.64, -1.08, -0.52, 0.52, 1.08, 1.64 | 10 | 0.9 | 0.042 | 0.021 | 0.023 |
| | | | 10 | 0.8 | 0.075 | 0.041 | 0.046 |
| | | | 10 | 0.7 | 0.103 | 0.062 | 0.068 |
| | | | 30 | 0.9 | 0.032 | 0.022 | 0.023 |
| | | | 30 | 0.8 | 0.054 | 0.044 | 0.046 |
| | | | 30 | 0.7 | 0.070 | 0.067 | 0.069 |
| | | | 60 | 0.9 | 0.027 | 0.023 | 0.023 |
| | | | 60 | 0.8 | 0.042 | 0.045 | 0.046 |
| | | | 60 | 0.7 | 0.054 | 0.068 | 0.069 |
| 3.3 | 0.0 | -2.33, -1.64, -1.04, 1.04, 1.64, 2.33 | 10 | 0.9 | 0.032 | 0.018 | 0.020 |
| | | | 10 | 0.8 | 0.058 | 0.036 | 0.040 |
| | | | 10 | 0.7 | 0.080 | 0.053 | 0.059 |
| | | | 30 | 0.9 | 0.026 | 0.019 | 0.020 |
| | | | 30 | 0.8 | 0.043 | 0.039 | 0.040 |
| | | | 30 | 0.7 | 0.057 | 0.057 | 0.059 |
| | | | 60 | 0.9 | 0.022 | 0.020 | 0.020 |
| | | | 60 | 0.8 | 0.035 | 0.039 | 0.040 |
| | | | 60 | 0.7 | 0.045 | 0.058 | 0.059 |
| 3.3 | -2.0 | -2.33, -1.88, -1.55, -1.17, -0.84, -0.55 | 10 | 0.9 | 0.035 | 0.020 | 0.022 |
| | | | 10 | 0.8 | 0.063 | 0.039 | 0.043 |
| | | | 10 | 0.7 | 0.086 | 0.057 | 0.063 |
| | | | 30 | 0.9 | 0.028 | 0.021 | 0.022 |
| | | | 30 | 0.8 | 0.046 | 0.042 | 0.043 |
| | | | 30 | 0.7 | 0.061 | 0.062 | 0.064 |
| | | | 60 | 0.9 | 0.023 | 0.022 | 0.022 |
| | | | 60 | 0.8 | 0.037 | 0.043 | 0.043 |
| | | | 60 | 0.7 | 0.047 | 0.063 | 0.064 |

Table 2. *Population RMSEA and SRMR and average of RMSEA and SRMR$_u$ across replications*

| Kur. | Skew. | ρ | N | *p* = 10 | | | | | | *p* = 30 | | | | | | *p* = 60 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | RMSEA | | | SRMR$_u$ | | | RMSEA | | | SRMR$_u$ | | | RMSEA | | | SRMR$_u$ | | |
| | | | | POP | NT | ADF | POP | NT | ADF | POP | NT | ADF | POP | NT | ADF | POP | NT | ADF | POP | NT | ADF |
| 0.0 | 0.0 | 0.9 | 100 | **0.042** | 0.042 | 0.042 | **0.021** | 0.018 | 0.019 | **0.032** | 0.050 | 0.051 | **0.022** | 0.022 | 0.023 | **0.027** | 0.065 | 0.065 | **0.023** | 0.023 | 0.024 |
| | | | 200 | **0.042** | 0.040 | 0.040 | **0.021** | 0.019 | 0.019 | **0.032** | 0.037 | 0.037 | **0.022** | 0.022 | 0.022 | **0.027** | 0.038 | 0.037 | **0.023** | 0.023 | 0.023 |
| | | | 500 | **0.042** | 0.041 | 0.041 | **0.021** | 0.020 | 0.020 | **0.032** | 0.033 | 0.033 | **0.022** | 0.022 | 0.022 | **0.027** | 0.029 | 0.029 | **0.023** | 0.023 | 0.023 |
| | | | 1000 | **0.042** | 0.042 | 0.041 | **0.021** | 0.021 | 0.021 | **0.032** | 0.033 | 0.033 | **0.022** | 0.022 | 0.022 | **0.027** | 0.027 | 0.027 | **0.023** | 0.023 | 0.023 |
| | | 0.8 | 100 | **0.075** | 0.073 | 0.073 | **0.041** | 0.039 | 0.040 | **0.054** | 0.066 | 0.066 | **0.044** | 0.044 | 0.045 | **0.042** | 0.072 | 0.072 | **0.045** | 0.045 | 0.045 |
| | | | 200 | **0.075** | 0.074 | 0.074 | **0.041** | 0.041 | 0.041 | **0.054** | 0.057 | 0.057 | **0.044** | 0.045 | 0.045 | **0.042** | 0.050 | 0.050 | **0.045** | 0.045 | 0.045 |
| | | | 500 | **0.075** | 0.075 | 0.075 | **0.041** | 0.041 | 0.041 | **0.054** | 0.055 | 0.054 | **0.044** | 0.045 | 0.045 | **0.042** | 0.044 | 0.043 | **0.045** | 0.045 | 0.045 |
| | | | 1000 | **0.075** | 0.075 | 0.075 | **0.041** | 0.041 | 0.041 | **0.054** | 0.054 | 0.054 | **0.044** | 0.045 | 0.045 | **0.042** | 0.043 | 0.043 | **0.045** | 0.045 | 0.045 |
| | | 0.7 | 100 | **0.103** | 0.101 | 0.101 | **0.062** | 0.061 | 0.061 | **0.070** | 0.080 | 0.080 | **0.067** | 0.066 | 0.067 | **0.054** | 0.080 | 0.079 | **0.068** | 0.067 | 0.067 |
| | | | 200 | **0.103** | 0.102 | 0.102 | **0.062** | 0.062 | 0.062 | **0.070** | 0.073 | 0.073 | **0.067** | 0.067 | 0.067 | **0.054** | 0.060 | 0.060 | **0.068** | 0.067 | 0.067 |
| | | | 500 | **0.103** | 0.103 | 0.102 | **0.062** | 0.062 | 0.062 | **0.070** | 0.071 | 0.071 | **0.067** | 0.067 | 0.067 | **0.054** | 0.055 | 0.055 | **0.068** | 0.068 | 0.068 |
| | | | 1000 | **0.103** | 0.103 | 0.103 | **0.062** | 0.062 | 0.062 | **0.070** | 0.071 | 0.071 | **0.067** | 0.067 | 0.067 | **0.054** | 0.054 | 0.054 | **0.068** | 0.068 | 0.068 |
| 3.0 | 0.0 | 0.9 | 100 | **0.032** | 0.058 | 0.039 | **0.018** | 0.028 | 0.019 | **0.026** | 0.068 | 0.050 | **0.019** | 0.035 | 0.023 | **0.022** | 0.079 | 0.065 | **0.020** | 0.036 | 0.024 |
| | | | 200 | **0.032** | 0.046 | 0.033 | **0.018** | 0.024 | 0.017 | **0.026** | 0.047 | 0.033 | **0.019** | 0.028 | 0.020 | **0.022** | 0.049 | 0.035 | **0.020** | 0.029 | 0.021 |
| | | | 500 | **0.032** | 0.038 | 0.031 | **0.018** | 0.021 | 0.017 | **0.026** | 0.035 | 0.027 | **0.019** | 0.024 | 0.020 | **0.022** | 0.033 | 0.024 | **0.020** | 0.024 | 0.020 |
| | | | 1000 | **0.032** | 0.036 | 0.032 | **0.018** | 0.020 | 0.018 | **0.026** | 0.031 | 0.027 | **0.019** | 0.022 | 0.020 | **0.022** | 0.027 | 0.023 | **0.020** | 0.022 | 0.020 |
| | | 0.8 | 100 | **0.058** | 0.075 | 0.060 | **0.036** | 0.041 | 0.035 | **0.043** | 0.075 | 0.061 | **0.039** | 0.048 | 0.040 | **0.035** | 0.083 | 0.070 | **0.039** | 0.049 | 0.041 |
| | | | 200 | **0.058** | 0.067 | 0.058 | **0.036** | 0.039 | 0.035 | **0.043** | 0.058 | 0.048 | **0.039** | 0.043 | 0.039 | **0.035** | 0.055 | 0.044 | **0.039** | 0.044 | 0.040 |
| | | | 500 | **0.058** | 0.063 | 0.059 | **0.036** | 0.037 | 0.036 | **0.043** | 0.050 | 0.045 | **0.039** | 0.041 | 0.039 | **0.035** | 0.042 | 0.037 | **0.039** | 0.041 | 0.039 |
| | | | 1000 | **0.058** | 0.062 | 0.060 | **0.036** | 0.037 | 0.036 | **0.043** | 0.047 | 0.045 | **0.039** | 0.040 | 0.039 | **0.035** | 0.039 | 0.036 | **0.039** | 0.040 | 0.039 |
| | | 0.7 | 100 | **0.080** | 0.092 | 0.081 | **0.053** | 0.056 | 0.052 | **0.057** | 0.083 | 0.071 | **0.057** | 0.064 | 0.058 | **0.045** | 0.086 | 0.074 | **0.058** | 0.064 | 0.059 |
| | | | 200 | **0.080** | 0.087 | 0.081 | **0.053** | 0.055 | 0.053 | **0.057** | 0.069 | 0.061 | **0.057** | 0.061 | 0.058 | **0.045** | 0.061 | 0.052 | **0.058** | 0.061 | 0.058 |
| | | | 500 | **0.080** | 0.085 | 0.082 | **0.053** | 0.055 | 0.054 | **0.057** | 0.062 | 0.059 | **0.057** | 0.059 | 0.058 | **0.045** | 0.051 | 0.046 | **0.058** | 0.060 | 0.058 |
| | | | 1000 | **0.080** | 0.084 | 0.083 | **0.053** | 0.054 | 0.054 | **0.057** | 0.060 | 0.059 | **0.057** | 0.058 | 0.058 | **0.045** | 0.048 | 0.046 | **0.058** | 0.059 | 0.059 |
| 3.0 | -2.0 | 0.9 | 100 | **0.035** | 0.098 | 0.048 | **0.020** | 0.049 | 0.023 | **0.028** | 0.106 | 0.064 | **0.021** | 0.057 | 0.026 | **0.023** | 0.121 | 0.086 | **0.022** | 0.059 | 0.028 |
| | | | 200 | **0.035** | 0.071 | 0.036 | **0.020** | 0.037 | 0.019 | **0.028** | 0.073 | 0.038 | **0.021** | 0.043 | 0.022 | **0.023** | 0.075 | 0.044 | **0.022** | 0.045 | 0.023 |
| | | | 500 | **0.035** | 0.052 | 0.033 | **0.020** | 0.028 | 0.019 | **0.028** | 0.049 | 0.029 | **0.021** | 0.032 | 0.021 | **0.023** | 0.048 | 0.027 | **0.022** | 0.033 | 0.022 |
| | | | 1000 | **0.035** | 0.045 | 0.034 | **0.020** | 0.024 | 0.019 | **0.028** | 0.040 | 0.028 | **0.021** | 0.027 | 0.021 | **0.023** | 0.037 | 0.024 | **0.022** | 0.028 | 0.022 |
| | | 0.8 | 100 | **0.063** | 0.108 | 0.066 | **0.039** | 0.059 | 0.038 | **0.046** | 0.110 | 0.072 | **0.042** | 0.068 | 0.044 | **0.037** | 0.121 | 0.088 | **0.043** | 0.069 | 0.045 |
| | | | 200 | **0.063** | 0.088 | 0.062 | **0.039** | 0.050 | 0.038 | **0.046** | 0.080 | 0.053 | **0.042** | 0.056 | 0.042 | **0.037** | 0.079 | 0.051 | **0.043** | 0.058 | 0.043 |
| | | | 500 | **0.063** | 0.073 | 0.062 | **0.039** | 0.044 | 0.039 | **0.046** | 0.061 | 0.047 | **0.042** | 0.048 | 0.042 | **0.037** | 0.055 | 0.039 | **0.043** | 0.049 | 0.042 |
| | | | 1000 | **0.063** | 0.068 | 0.063 | **0.039** | 0.042 | 0.039 | **0.046** | 0.054 | 0.047 | **0.042** | 0.045 | 0.042 | **0.037** | 0.046 | 0.037 | **0.043** | 0.046 | 0.043 |
| | | 0.7 | 100 | **0.086** | 0.119 | 0.085 | **0.057** | 0.071 | 0.055 | **0.061** | 0.114 | 0.079 | **0.062** | 0.081 | 0.062 | **0.047** | 0.122 | 0.090 | **0.063** | 0.082 | 0.063 |
| | | | 200 | **0.086** | 0.103 | 0.083 | **0.057** | 0.065 | 0.056 | **0.061** | 0.088 | 0.065 | **0.062** | 0.072 | 0.061 | **0.047** | 0.082 | 0.057 | **0.063** | 0.073 | 0.062 |
| | | | 500 | **0.086** | 0.093 | 0.085 | **0.057** | 0.060 | 0.057 | **0.061** | 0.072 | 0.061 | **0.062** | 0.066 | 0.062 | **0.047** | 0.061 | 0.048 | **0.063** | 0.067 | 0.062 |
| | | | 1000 | **0.086** | 0.089 | 0.085 | **0.057** | 0.059 | 0.057 | **0.061** | 0.066 | 0.061 | **0.062** | 0.064 | 0.062 | **0.047** | 0.054 | 0.047 | **0.063** | 0.065 | 0.062 |

*Notes: Shaded results indicate relative bias* <|*.10*|.

Table 3. *Coverage rates for 90% confidence intervals around the population RMSEA and SRMR*

| Kur. | Skew. | ρ | N | *p* = 10 | | | | *p* = 30 | | | | *p* = 60 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | NT | | ADF | | NT | | ADF | | NT | | ADF | |
| | | | | RMSEA | SRMR | RMSEA | SRMR | RMSEA | SRMR | RMSEA | SRMR | RMSEA | SRMR | RMSEA | SRMR |
| 0.0 | 0.0 | 0.9 | 100 | 0.92 | 0.69 | 0.92 | 0.75 | 0.40 | 0.75 | 0.40 | 0.84 | 0.00 | 0.77 | 0.00 | 0.92 |
| | | | 200 | 0.93 | 0.78 | 0.94 | 0.79 | 0.75 | 0.80 | 0.76 | 0.84 | 0.04 | 0.76 | 0.05 | 0.85 |
| | | | 500 | 0.88 | 0.85 | 0.88 | 0.85 | 0.83 | 0.83 | 0.85 | 0.84 | 0.64 | 0.79 | 0.68 | 0.81 |
| | | | 1000 | 0.86 | 0.86 | 0.86 | 0.86 | 0.83 | 0.86 | 0.83 | 0.86 | 0.75 | 0.83 | 0.77 | 0.84 |
| | | 0.8 | 100 | 0.86 | 0.83 | 0.87 | 0.84 | 0.50 | 0.80 | 0.51 | 0.84 | 0.00 | 0.80 | 0.00 | 0.86 |
| | | | 200 | 0.85 | 0.87 | 0.86 | 0.87 | 0.71 | 0.83 | 0.72 | 0.84 | 0.14 | 0.82 | 0.17 | 0.84 |
| | | | 500 | 0.82 | 0.88 | 0.83 | 0.88 | 0.75 | 0.86 | 0.76 | 0.86 | 0.63 | 0.87 | 0.67 | 0.87 |
| | | | 1000 | 0.83 | 0.87 | 0.83 | 0.87 | 0.74 | 0.89 | 0.75 | 0.89 | 0.64 | 0.88 | 0.65 | 0.88 |
| | | 0.7 | 100 | 0.83 | 0.87 | 0.84 | 0.87 | 0.54 | 0.82 | 0.56 | 0.84 | 0.00 | 0.82 | 0.00 | 0.84 |
| | | | 200 | 0.83 | 0.90 | 0.84 | 0.90 | 0.69 | 0.86 | 0.71 | 0.85 | 0.25 | 0.85 | 0.31 | 0.86 |
| | | | 500 | 0.83 | 0.91 | 0.84 | 0.91 | 0.70 | 0.89 | 0.72 | 0.89 | 0.59 | 0.89 | 0.61 | 0.89 |
| | | | 1000 | 0.82 | 0.89 | 0.82 | 0.89 | 0.71 | 0.90 | 0.72 | 0.89 | 0.57 | 0.89 | 0.59 | 0.89 |
| 3.3 | 0.0 | 0.9 | 100 | 0.70 | 0.74 | 0.92 | 0.96 | 0.01 | 0.15 | 0.33 | 0.93 | 0.00 | 0.01 | 0.00 | 0.99 |
| | | | 200 | 0.72 | 0.68 | 0.93 | 0.79 | 0.06 | 0.23 | 0.74 | 0.90 | 0.00 | 0.02 | 0.06 | 0.98 |
| | | | 500 | 0.78 | 0.77 | 0.92 | 0.82 | 0.13 | 0.42 | 0.82 | 0.89 | 0.00 | 0.15 | 0.64 | 0.93 |
| | | | 1000 | 0.75 | 0.80 | 0.87 | 0.86 | 0.22 | 0.58 | 0.81 | 0.87 | 0.00 | 0.39 | 0.74 | 0.90 |
| | | 0.8 | 100 | 0.75 | 0.80 | 0.91 | 0.86 | 0.02 | 0.67 | 0.43 | 0.96 | 0.00 | 0.52 | 0.00 | 0.99 |
| | | | 200 | 0.77 | 0.85 | 0.90 | 0.90 | 0.14 | 0.73 | 0.73 | 0.92 | 0.00 | 0.64 | 0.14 | 0.95 |
| | | | 500 | 0.77 | 0.84 | 0.84 | 0.89 | 0.27 | 0.79 | 0.78 | 0.91 | 0.00 | 0.76 | 0.63 | 0.92 |
| | | | 1000 | 0.74 | 0.84 | 0.84 | 0.90 | 0.34 | 0.83 | 0.75 | 0.92 | 0.03 | 0.81 | 0.64 | 0.91 |
| | | 0.7 | 100 | 0.78 | 0.86 | 0.89 | 0.90 | 0.06 | 0.82 | 0.47 | 0.95 | 0.00 | 0.83 | 0.00 | 0.96 |
| | | | 200 | 0.77 | 0.87 | 0.87 | 0.90 | 0.23 | 0.82 | 0.72 | 0.92 | 0.00 | 0.84 | 0.24 | 0.92 |
| | | | 500 | 0.77 | 0.87 | 0.85 | 0.91 | 0.37 | 0.85 | 0.72 | 0.92 | 0.05 | 0.83 | 0.60 | 0.93 |
| | | | 1000 | 0.75 | 0.86 | 0.81 | 0.90 | 0.42 | 0.86 | 0.71 | 0.94 | 0.11 | 0.85 | 0.55 | 0.92 |
| 3.3 | -2.0 | 0.9 | 100 | 0.21 | 0.27 | 0.90 | 0.96 | 0.00 | 0.00 | 0.17 | 0.99 | 0.00 | 0.00 | 0.00 | 1.00 |
| | | | 200 | 0.29 | 0.33 | 0.92 | 0.92 | 0.00 | 0.00 | 0.70 | 0.93 | 0.00 | 0.00 | 0.01 | 0.99 |
| | | | 500 | 0.41 | 0.48 | 0.94 | 0.82 | 0.00 | 0.00 | 0.86 | 0.91 | 0.00 | 0.00 | 0.64 | 0.97 |
| | | | 1000 | 0.52 | 0.58 | 0.87 | 0.84 | 0.00 | 0.05 | 0.86 | 0.90 | 0.00 | 0.00 | 0.79 | 0.92 |
| | | 0.8 | 100 | 0.33 | 0.60 | 0.91 | 0.89 | 0.00 | 0.05 | 0.30 | 0.98 | 0.00 | 0.00 | 0.00 | 1.00 |
| | | | 200 | 0.45 | 0.64 | 0.92 | 0.89 | 0.00 | 0.18 | 0.75 | 0.96 | 0.00 | 0.04 | 0.05 | 0.97 |
| | | | 500 | 0.60 | 0.74 | 0.87 | 0.89 | 0.01 | 0.45 | 0.80 | 0.91 | 0.00 | 0.31 | 0.71 | 0.93 |
| | | | 1000 | 0.65 | 0.76 | 0.86 | 0.90 | 0.05 | 0.62 | 0.79 | 0.93 | 0.00 | 0.53 | 0.70 | 0.92 |
| | | 0.7 | 100 | 0.50 | 0.75 | 0.91 | 0.89 | 0.00 | 0.43 | 0.43 | 0.97 | 0.00 | 0.27 | 0.00 | 0.99 |
| | | | 200 | 0.57 | 0.75 | 0.86 | 0.88 | 0.00 | 0.60 | 0.76 | 0.94 | 0.00 | 0.50 | 0.17 | 0.94 |
| | | | 500 | 0.66 | 0.78 | 0.84 | 0.91 | 0.06 | 0.70 | 0.77 | 0.93 | 0.00 | 0.70 | 0.71 | 0.94 |
| | | | 1000 | 0.66 | 0.78 | 0.82 | 0.91 | 0.19 | 0.75 | 0.72 | 0.95 | 0.00 | 0.72 | 0.64 | 0.93 |

*Notes:* Shaded results indicate acceptable coverage (between ,85 and .95).

Table 4. *Coverage rates for 95% confidence intervals around the population RMSEA and SRMR*

| | | | | *p* = 10 | | | | *p* = 30 | | | | *p* = 60 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Kur. | Skew. | ρ | N | NT | | ADF | | NT | | ADF | | NT | | ADF | |
| | | | | RMSEA | SRMR | RMSEA | SRMR | RMSEA | SRMR | RMSEA | SRMR | RMSEA | SRMR | RMSEA | SRMR |
| 0.0 | 0.0 | 0.9 | 100 | 0.95 | 0.89 | 0.95 | 0.95 | 0.52 | 0.83 | 0.50 | 0.91 | 0.00 | 0.86 | 0.00 | 0.97 |
| | | | 200 | 0.96 | 0.84 | 0.97 | 0.83 | 0.83 | 0.86 | 0.85 | 0.88 | 0.09 | 0.83 | 0.10 | 0.93 |
| | | | 500 | 0.93 | 0.90 | 0.93 | 0.90 | 0.90 | 0.90 | 0.91 | 0.91 | 0.72 | 0.87 | 0.76 | 0.89 |
| | | | 1000 | 0.92 | 0.92 | 0.92 | 0.93 | 0.89 | 0.92 | 0.90 | 0.92 | 0.83 | 0.89 | 0.85 | 0.89 |
| | | 0.8 | 100 | 0.95 | 0.89 | 0.95 | 0.91 | 0.61 | 0.88 | 0.60 | 0.92 | 0.00 | 0.86 | 0.00 | 0.91 |
| | | | 200 | 0.93 | 0.93 | 0.93 | 0.93 | 0.79 | 0.90 | 0.80 | 0.90 | 0.21 | 0.90 | 0.24 | 0.90 |
| | | | 500 | 0.90 | 0.93 | 0.90 | 0.93 | 0.84 | 0.92 | 0.85 | 0.93 | 0.72 | 0.93 | 0.75 | 0.93 |
| | | | 1000 | 0.90 | 0.94 | 0.90 | 0.94 | 0.84 | 0.94 | 0.84 | 0.94 | 0.73 | 0.94 | 0.73 | 0.93 |
| | | 0.7 | 100 | 0.89 | 0.92 | 0.90 | 0.92 | 0.63 | 0.90 | 0.64 | 0.90 | 0.00 | 0.88 | 0.00 | 0.90 |
| | | | 200 | 0.91 | 0.94 | 0.91 | 0.94 | 0.76 | 0.91 | 0.78 | 0.91 | 0.34 | 0.92 | 0.39 | 0.92 |
| | | | 500 | 0.90 | 0.95 | 0.90 | 0.94 | 0.79 | 0.94 | 0.80 | 0.94 | 0.68 | 0.94 | 0.71 | 0.94 |
| | | | 1000 | 0.89 | 0.94 | 0.89 | 0.94 | 0.78 | 0.95 | 0.79 | 0.95 | 0.67 | 0.94 | 0.67 | 0.94 |
| 3.3 | 0.0 | 0.9 | 100 | 0.79 | 0.84 | 0.96 | 0.99 | 0.01 | 0.24 | 0.45 | 0.99 | 0.00 | 0.02 | 0.00 | 1.00 |
| | | | 200 | 0.81 | 0.77 | 0.96 | 0.97 | 0.08 | 0.32 | 0.83 | 0.94 | 0.00 | 0.05 | 0.10 | 0.99 |
| | | | 500 | 0.84 | 0.84 | 0.95 | 0.87 | 0.20 | 0.52 | 0.90 | 0.94 | 0.00 | 0.22 | 0.74 | 0.97 |
| | | | 1000 | 0.84 | 0.87 | 0.92 | 0.92 | 0.30 | 0.69 | 0.88 | 0.93 | 0.00 | 0.48 | 0.82 | 0.96 |
| | | 0.8 | 100 | 0.82 | 0.89 | 0.95 | 0.92 | 0.05 | 0.78 | 0.54 | 0.99 | 0.00 | 0.66 | 0.00 | 1.00 |
| | | | 200 | 0.85 | 0.91 | 0.96 | 0.94 | 0.19 | 0.81 | 0.81 | 0.96 | 0.00 | 0.76 | 0.22 | 0.98 |
| | | | 500 | 0.85 | 0.91 | 0.91 | 0.94 | 0.37 | 0.86 | 0.84 | 0.94 | 0.01 | 0.85 | 0.73 | 0.96 |
| | | | 1000 | 0.83 | 0.92 | 0.91 | 0.95 | 0.43 | 0.89 | 0.83 | 0.96 | 0.06 | 0.87 | 0.73 | 0.97 |
| | | 0.7 | 100 | 0.86 | 0.92 | 0.95 | 0.94 | 0.10 | 0.90 | 0.59 | 0.97 | 0.00 | 0.91 | 0.00 | 0.98 |
| | | | 200 | 0.84 | 0.92 | 0.92 | 0.94 | 0.31 | 0.89 | 0.79 | 0.95 | 0.00 | 0.90 | 0.33 | 0.96 |
| | | | 500 | 0.86 | 0.94 | 0.92 | 0.97 | 0.46 | 0.91 | 0.81 | 0.97 | 0.07 | 0.90 | 0.69 | 0.97 |
| | | | 1000 | 0.82 | 0.91 | 0.89 | 0.95 | 0.51 | 0.92 | 0.78 | 0.97 | 0.15 | 0.91 | 0.65 | 0.96 |
| 3.3 | -2.0 | 0.9 | 100 | 0.30 | 0.38 | 0.94 | 0.99 | 0.00 | 0.00 | 0.26 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| | | | 200 | 0.38 | 0.43 | 0.96 | 0.98 | 0.00 | 0.00 | 0.80 | 0.95 | 0.00 | 0.00 | 0.02 | 1.00 |
| | | | 500 | 0.51 | 0.58 | 0.97 | 0.87 | 0.00 | 0.00 | 0.92 | 0.96 | 0.00 | 0.00 | 0.76 | 0.99 |
| | | | 1000 | 0.59 | 0.66 | 0.93 | 0.90 | 0.00 | 0.08 | 0.93 | 0.95 | 0.00 | 0.00 | 0.87 | 0.96 |
| | | 0.8 | 100 | 0.44 | 0.72 | 0.96 | 0.98 | 0.00 | 0.08 | 0.41 | 0.99 | 0.00 | 0.00 | 0.00 | 1.00 |
| | | | 200 | 0.54 | 0.73 | 0.96 | 0.91 | 0.00 | 0.26 | 0.84 | 0.98 | 0.00 | 0.06 | 0.09 | 0.99 |
| | | | 500 | 0.70 | 0.83 | 0.93 | 0.94 | 0.01 | 0.55 | 0.87 | 0.96 | 0.00 | 0.42 | 0.79 | 0.96 |
| | | | 1000 | 0.74 | 0.83 | 0.92 | 0.95 | 0.07 | 0.71 | 0.86 | 0.97 | 0.00 | 0.61 | 0.78 | 0.96 |
| | | 0.7 | 100 | 0.57 | 0.85 | 0.95 | 0.93 | 0.00 | 0.58 | 0.55 | 0.99 | 0.00 | 0.39 | 0.00 | 1.00 |
| | | | 200 | 0.66 | 0.85 | 0.92 | 0.93 | 0.00 | 0.71 | 0.84 | 0.97 | 0.00 | 0.64 | 0.25 | 0.97 |
| | | | 500 | 0.74 | 0.87 | 0.91 | 0.96 | 0.08 | 0.80 | 0.83 | 0.97 | 0.00 | 0.80 | 0.78 | 0.96 |
| | | | 1000 | 0.74 | 0.85 | 0.89 | 0.94 | 0.25 | 0.84 | 0.81 | 0.97 | 0.00 | 0.83 | 0.73 | 0.97 |

*Notes:* Shaded results indicate acceptable coverage (between .90 and .99).

Table 5: *Test of close fit results. Empirical rejection rates at a 5% significance level of a test that the RMSEA and SRMR equal their population values*

| Kur. | Skew. | ρ | N | *p* = 10 | | | | | | *p* = 30 | | | | | | *p* = 60 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | RMSEA | | | SRMR | | | RMSEA | | | SRMR | | | RMSEA | | | SRMR | | |
| | | | | POP | NT | ADF | POP | NT | ADF | POP | NT | ADF | POP | NT | ADF | POP | NT | ADF | POP | NT | ADF |
| 0.0 | 0.0 | 0.9 | 100 | **0.042** | 0.08 | 0.08 | **0.021** | 0.11 | 0.05 | **0.032** | 0.62 | 0.61 | **0.022** | 0.17 | 0.09 | **0.027** | 1.00 | 0.14 | **0.023** | 1.00 | 0.03 |
| | | | 200 | **0.042** | 0.07 | 0.06 | **0.021** | 0.16 | 0.17 | **0.032** | 0.24 | 0.22 | **0.022** | 0.14 | 0.12 | **0.027** | 0.96 | 0.17 | **0.023** | 0.95 | 0.07 |
| | | | 500 | **0.042** | 0.07 | 0.07 | **0.021** | 0.10 | 0.10 | **0.032** | 0.14 | 0.12 | **0.022** | 0.11 | 0.09 | **0.027** | 0.36 | 0.13 | **0.023** | 0.32 | 0.11 |
| | | | 1000 | **0.042** | 0.08 | 0.08 | **0.021** | 0.08 | 0.08 | **0.032** | 0.11 | 0.10 | **0.022** | 0.08 | 0.08 | **0.027** | 0.21 | 0.11 | **0.023** | 0.18 | 0.11 |
| | | 0.8 | 100 | **0.075** | 0.10 | 0.09 | **0.041** | 0.11 | 0.10 | **0.054** | 0.52 | 0.51 | **0.044** | 0.12 | 0.08 | **0.042** | 1.00 | 0.14 | **0.045** | 1.00 | 0.09 |
| | | | 200 | **0.075** | 0.08 | 0.08 | **0.041** | 0.07 | 0.07 | **0.054** | 0.27 | 0.25 | **0.044** | 0.10 | 0.10 | **0.042** | 0.87 | 0.10 | **0.045** | 0.84 | 0.10 |
| | | | 500 | **0.075** | 0.09 | 0.08 | **0.041** | 0.07 | 0.07 | **0.054** | 0.18 | 0.16 | **0.044** | 0.08 | 0.08 | **0.042** | 0.32 | 0.07 | **0.045** | 0.27 | 0.07 |
| | | | 1000 | **0.075** | 0.09 | 0.09 | **0.041** | 0.06 | 0.07 | **0.054** | 0.16 | 0.14 | **0.044** | 0.06 | 0.06 | **0.042** | 0.24 | 0.07 | **0.045** | 0.21 | 0.07 |
| | | 0.7 | 100 | **0.103** | 0.10 | 0.10 | **0.062** | 0.08 | 0.08 | **0.070** | 0.48 | 0.45 | **0.067** | 0.10 | 0.10 | **0.054** | 1.00 | 0.12 | **0.068** | 1.00 | 0.10 |
| | | | 200 | **0.103** | 0.10 | 0.09 | **0.062** | 0.06 | 0.06 | **0.070** | 0.26 | 0.23 | **0.067** | 0.09 | 0.09 | **0.054** | 0.75 | 0.09 | **0.068** | 0.70 | 0.08 |
| | | | 500 | **0.103** | 0.09 | 0.08 | **0.062** | 0.05 | 0.06 | **0.070** | 0.20 | 0.18 | **0.067** | 0.06 | 0.06 | **0.054** | 0.31 | 0.06 | **0.068** | 0.27 | 0.06 |
| | | | 1000 | **0.103** | 0.10 | 0.10 | **0.062** | 0.06 | 0.06 | **0.070** | 0.17 | 0.15 | **0.067** | 0.05 | 0.06 | **0.054** | 0.26 | 0.07 | **0.068** | 0.22 | 0.06 |
| 3.0 | 0.0 | 0.9 | 100 | **0.032** | 0.31 | 0.08 | **0.018** | 0.16 | 0.01 | **0.026** | 1.00 | 0.68 | **0.019** | 0.76 | 0.01 | **0.022** | 1.00 | 0.98 | **0.020** | 1.00 | 0.00 |
| | | | 200 | **0.032** | 0.28 | 0.07 | **0.018** | 0.24 | 0.03 | **0.026** | 0.95 | 0.27 | **0.019** | 0.68 | 0.06 | **0.022** | 1.00 | 0.95 | **0.020** | 0.95 | 0.01 |
| | | | 500 | **0.032** | 0.23 | 0.08 | **0.018** | 0.16 | 0.13 | **0.026** | 0.87 | 0.15 | **0.019** | 0.48 | 0.06 | **0.022** | 1.00 | 0.78 | **0.020** | 0.36 | 0.03 |
| | | | 1000 | **0.032** | 0.22 | 0.09 | **0.018** | 0.13 | 0.09 | **0.026** | 0.78 | 0.16 | **0.019** | 0.31 | 0.07 | **0.022** | 1.00 | 0.52 | **0.020** | 0.25 | 0.04 |
| | | 0.8 | 100 | **0.058** | 0.26 | 0.10 | **0.036** | 0.11 | 0.08 | **0.043** | 0.98 | 0.59 | **0.039** | 0.22 | 0.01 | **0.035** | 1.00 | 0.34 | **0.039** | 1.00 | 0.00 |
| | | | 200 | **0.058** | 0.21 | 0.07 | **0.036** | 0.09 | 0.06 | **0.043** | 0.87 | 0.26 | **0.039** | 0.19 | 0.04 | **0.035** | 1.00 | 0.24 | **0.039** | 0.87 | 0.02 |
| | | | 500 | **0.058** | 0.20 | 0.10 | **0.036** | 0.09 | 0.06 | **0.043** | 0.73 | 0.19 | **0.039** | 0.14 | 0.06 | **0.035** | 1.00 | 0.15 | **0.039** | 0.36 | 0.04 |
| | | | 1000 | **0.058** | 0.22 | 0.12 | **0.036** | 0.08 | 0.05 | **0.043** | 0.66 | 0.22 | **0.039** | 0.11 | 0.04 | **0.035** | 0.97 | 0.13 | **0.039** | 0.34 | 0.03 |
| | | 0.7 | 100 | **0.080** | 0.21 | 0.09 | **0.053** | 0.08 | 0.06 | **0.057** | 0.95 | 0.54 | **0.057** | 0.10 | 0.03 | **0.045** | 1.00 | 0.09 | **0.058** | 1.00 | 0.02 |
| | | | 200 | **0.080** | 0.20 | 0.09 | **0.053** | 0.09 | 0.06 | **0.057** | 0.78 | 0.27 | **0.057** | 0.11 | 0.05 | **0.045** | 1.00 | 0.10 | **0.058** | 0.78 | 0.04 |
| | | | 500 | **0.080** | 0.20 | 0.10 | **0.053** | 0.06 | 0.03 | **0.057** | 0.63 | 0.25 | **0.057** | 0.09 | 0.03 | **0.045** | 0.95 | 0.10 | **0.058** | 0.36 | 0.03 |
| | | | 1000 | **0.080** | 0.22 | 0.15 | **0.053** | 0.09 | 0.05 | **0.057** | 0.58 | 0.26 | **0.057** | 0.08 | 0.03 | **0.045** | 0.89 | 0.10 | **0.058** | 0.40 | 0.04 |
| 3.0 | -2.0 | 0.9 | 100 | **0.035** | 0.80 | 0.11 | **0.020** | 0.62 | 0.01 | **0.028** | 1.00 | 0.83 | **0.021** | 1.00 | 0.00 | **0.023** | 1.00 | 1.00 | **0.022** | 1.00 | 0.00 |
| | | | 200 | **0.035** | 0.72 | 0.08 | **0.020** | 0.57 | 0.02 | **0.028** | 1.00 | 0.31 | **0.021** | 1.00 | 0.05 | **0.023** | 1.00 | 1.00 | **0.022** | 0.99 | 0.00 |
| | | | 500 | **0.035** | 0.59 | 0.06 | **0.020** | 0.42 | 0.13 | **0.028** | 1.00 | 0.11 | **0.021** | 1.00 | 0.05 | **0.023** | 1.00 | 1.00 | **0.022** | 0.36 | 0.01 |
| | | | 1000 | **0.035** | 0.48 | 0.07 | **0.020** | 0.34 | 0.10 | **0.028** | 1.00 | 0.09 | **0.021** | 0.92 | 0.05 | **0.023** | 1.00 | 1.00 | **0.022** | 0.17 | 0.04 |
| | | 0.8 | 100 | **0.063** | 0.68 | 0.09 | **0.039** | 0.28 | 0.02 | **0.046** | 1.00 | 0.71 | **0.042** | 0.92 | 0.01 | **0.037** | 1.00 | 1.00 | **0.043** | 1.00 | 0.00 |
| | | | 200 | **0.063** | 0.55 | 0.08 | **0.039** | 0.27 | 0.09 | **0.046** | 1.00 | 0.25 | **0.042** | 0.74 | 0.02 | **0.037** | 1.00 | 0.94 | **0.043** | 0.95 | 0.01 |
| | | | 500 | **0.063** | 0.39 | 0.07 | **0.039** | 0.18 | 0.06 | **0.046** | 1.00 | 0.14 | **0.042** | 0.45 | 0.04 | **0.037** | 1.00 | 0.59 | **0.043** | 0.26 | 0.04 |
| | | | 1000 | **0.063** | 0.32 | 0.08 | **0.039** | 0.17 | 0.05 | **0.046** | 0.95 | 0.12 | **0.042** | 0.29 | 0.03 | **0.037** | 1.00 | 0.40 | **0.043** | 0.18 | 0.04 |
| | | 0.7 | 100 | **0.086** | 0.51 | 0.09 | **0.057** | 0.15 | 0.07 | **0.061** | 1.00 | 0.58 | **0.062** | 0.42 | 0.01 | **0.047** | 1.00 | 0.61 | **0.063** | 1.00 | 0.00 |
| | | | 200 | **0.086** | 0.41 | 0.07 | **0.057** | 0.15 | 0.07 | **0.061** | 1.00 | 0.22 | **0.062** | 0.29 | 0.03 | **0.047** | 1.00 | 0.36 | **0.063** | 0.84 | 0.03 |
| | | | 500 | **0.086** | 0.30 | 0.07 | **0.057** | 0.13 | 0.04 | **0.061** | 0.94 | 0.14 | **0.062** | 0.20 | 0.04 | **0.047** | 1.00 | 0.20 | **0.063** | 0.20 | 0.04 |
| | | | 1000 | **0.086** | 0.26 | 0.08 | **0.057** | 0.15 | 0.06 | **0.061** | 0.81 | 0.14 | **0.062** | 0.16 | 0.03 | **0.047** | 1.00 | 0.17 | **0.063** | 0.15 | 0.03 |

*Notes:* Shaded results indicate acceptable rejection rates (between .01 and .10).