Finite Sampling Properties of the Point Estimates and Confidence Intervals of the RMSEA

PATRICK J. CURRAN
KENNETH A. BOLLEN
University of North Carolina at Chapel Hill

FEINIAN CHEN
Texas A&M University

PAMELA PAXTON
The Ohio State University

JAMES B. KIRBY
Agency for Healthcare, Research & Quality

A key advantage of the root mean square error of approximation (RMSEA) is that under certain assumptions, the sample estimate has a known sampling distribution that allows for the computation of confidence intervals. However, little is known about the finite sampling behaviors of this measure under violations of these ideal asymptotic conditions. This information is critical for developing optimal criteria for using the RMSEA to evaluate model fit in practice. Using data generated from a computer simulation study, the authors empirically tested a set of theoretically generated research hypotheses about the sampling characteristics of the RMSEA under conditions commonly encountered in applied social science research. The results suggest that both the sample estimates and confidence intervals are accurate for sample sizes of $n = 200$ and higher, but caution is warranted in the use of these measures at smaller sample sizes, at least for the types of models considered here.

Keywords: *RMSEA; SEM; goodness of fit; computer simulations; noncentrality*

Assessing model fit is one of the most controversial issues in structural equation modeling (SEM), and a long and rich line of research has

AUTHORS' NOTE: *This work was funded in part by grant DA13148 awarded by the National Institute on Drug Abuse to the first two authors. The authors would like to thank the editor and anonymous reviewers for the detailed and extremely helpful feedback. Correspondence should be addressed to Patrick J. Curran, Department of Psychology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3270; e-mail: curran@unc.edu.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 32, No. 2, November 2003 208-252
DOI: 10.1177/0049124103256130

addressed this challenging problem (e.g., Bollen and Long 1993; Jöreskog 1993; Kaplan 1990; MacCallum 1990). One measure of fit that has become increasingly used in SEM analysis is the root mean squared error of approximation, or RMSEA. Although the asymptotic statistical properties of the sample estimate of the RMSEA are well established under ideal conditions, much less is known about the sampling properties in conditions encountered in practice. Of particular interest are the influences of sample size, model complexity, and model misspecification on the estimation of the RMSEA. Knowledge of these influences is crucial for the development of well-informed guidelines for using estimates of the RMSEA to evaluate model fit in applied research.

We open with a brief review of the general goals of SEM and the challenges of assessing model fit. We then summarize the background and development of the RMSEA and review existing simulation studies of this measure. Finally, we combine information from both statistical theory and prior research to generate a set of specific research hypotheses to be empirically tested using computer simulation methodology. Taken together, we hope this work will help us gain a better understanding about the estimation and the informed use of the RMSEA in applied social science research.

BACKGROUND AND DEVELOPMENT

Given a set of p-observed variables, the covariance structure hypothesis in SEM states that $\Sigma_{\mathbf{0}} = \Sigma(\theta)$, where $\Sigma_{\mathbf{0}}$ represents the $p \times p$ covariance matrix of the observed variables in the population, $\Sigma(\theta)$ represents the covariance matrix implied by the hypothesized model, and θ represents a vector of free parameters in the hypothesized model.[1] The covariance structure hypothesis thus posits that the covariance matrix implied by the hypothesized model is equal to the population covariance matrix of the observed variables. The sample estimator of θ is denoted $\hat{\theta}$ and is calculated to minimize the discrepancy between the covariance matrix implied by the model (denoted $\Sigma(\hat{\theta})$) and the covariance matrix observed in the sample (denoted $\mathbf{S}$).

More formally, the parameter estimates in $\hat{\theta}$ are obtained by minimizing a suitable discrepancy function denoted $\hat{F}(\mathbf{S}, \Sigma(\hat{\theta}))$. Although

there are several important functions from which to choose, the maximum likelihood (ML) estimator remains the dominant one. The ML fitting function is as follows:

$$\hat{F}(\mathbf{S}, \Sigma(\hat{\theta})) = \log|\Sigma(\hat{\theta})| + tr(\mathbf{S}\Sigma^{-1}(\hat{\theta})) - \log|\mathbf{S}| - p, \quad (1)$$

where $\mathbf{S}$, $\Sigma(\hat{\theta})$, and p are defined as before. Assuming no excess multivariate kurtosis, adequate sample size, and proper model specification, ML parameter estimates of $\hat{\theta}$ are asymptotically unbiased, consistent, and asymptotically efficient (Bollen 1989a; Browne 1984). Furthermore, the minimum of the ML function is a scalar value denoted $\hat{F}$ that is equal to zero only when the observed covariance matrix $\mathbf{S}$ is equal to the reproduced covariance matrix $\Sigma(\hat{\theta})$. The larger the discrepancy between $\mathbf{S}$ and $\Sigma(\hat{\theta})$, the greater the value of $\hat{F}$. Because of this property, the minimum of the discrepancy function has long been used as the basis for many measures of overall model fit.

LIKELIHOOD RATIO STATISTIC T

The most common measure of fit based on $\hat{F}$ is the likelihood ratio test statistic that is defined as $T = \hat{F}(N-1)$, where N represents total sample size. Under assumptions of sufficiently large sample size, no excess multivariate kurtosis, and proper model specification (Browne 1984), T is asymptotically distributed as a central chi-square with degrees of freedom denoted as df. Given the known asymptotic sampling distribution of T under proper model specification, this test statistic allows researchers to test the null hypothesis $\Sigma_{\mathbf{0}} = \Sigma(\theta)$ that the population covariance matrix equals the covariance matrix implied by the population model parameters. Rejection casts doubt on the model that leads to $\Sigma(\theta)$.

In practice, several factors complicate the interpretation and use of this test statistic. Most important, all empirical models are likely misspecified to some degree, and a formal test of exact fit is often not an optimal method of model evaluation. More specifically, the high statistical power of the test in moderate to large samples can lead to the rejection of models with even trivial misspecifications. A further problem is that the T statistic tends to be larger than a chi-square

variate at smaller sample sizes and in the presence of excess multivariate kurtosis (Anderson and Gerbing 1984; Muthén and Kaplan 1985, 1992; Curran, West, and Finch 1996). Because of these and other limitations, a variety of alternative measures of model fit have been proposed to augment the likelihood ratio test T.

Baseline fit indices represent a family of such measures (e.g., Bentler 1990; Bentler and Bonett 1980; Bollen 1989b; McDonald and Marsh 1990; Tucker and Lewis 1973). These measures share the use of a baseline model to which to compare the relative fit of the hypothesized model. Typically, the baseline model is defined such that all variables are uncorrelated and only the variances are estimated. Most fit indices range between 0 and 1, with higher values indicating better model fit and values exceeding .90 or .95 indicating acceptable fit of the model to the sample data. Although there are a number of advantages to the baseline fit indices, there are several limitations as well. We refer the reader to other sources for further discussion of these baseline fit indices (e.g., Gerbing and Anderson 1993; Rigdon 1998; Tanaka 1993).

ROOT MEAN SQUARE ERROR OF APPROXIMATION (RMSEA)

Absolute fit indices refer to another family of alternative measures that do not measure fit relative to some baseline model. One such measure of fit that has become increasingly popular is the RMSEA. The seminal work on the RMSEA was first presented by Steiger and Lind (1980) and was later elaborated by Browne and Cudeck (1993) and Steiger (2000). The computation of the sample estimate of the RMSEA is based on the relation between the test statistic T and the noncentral chi-square distribution. Under the assumptions of large sample size, no excess multivariate kurtosis, and *proper* model specification, T follows a central chi-square distribution with expected value df and variance $2df$. However, under the assumptions of large sample size, no excess multivariate kurtosis, but *improper* specification, the test statistic T follows a noncentral chi-square with df and noncentrality parameter λ with expected value $df + \lambda$ and variance $2df + 4\lambda$ (Steiger, Shapiro, and Browne 1985). The noncentrality parameter λ is a measure of the degree of misspecification of a hypothesized model (Bentler 1990; McDonald and Marsh 1990; Steiger and Lind 1980).

Browne and Cudeck (1993) distinguish between two types of error in SEM: errors of estimation and errors of approximation. Adopting the notation of Browne and Cudeck, $\Sigma_{\mathbf{0}}$ is the population covariance matrix, $\tilde{\Sigma}_{\mathbf{0}}$ represents the best fit of the hypothesized model to the population covariance matrix, and $\hat{\Sigma}$ represents the best fit of the hypothesized model to the observed covariance matrix **S**. The first two matrices are fixed in the population, and the third is random. *Errors of approximation* represent the degree of misfit between the population covariance matrix $\Sigma_{\mathbf{0}}$ and the population model–implied covariance matrix $\tilde{\Sigma}_{\mathbf{0}}$, and this degree of misfit is estimated as a discrepancy due to approximation via a suitable discrepancy function $F_0 = (\Sigma_{\mathbf{0}}, \tilde{\Sigma}_{\mathbf{0}})$. In contrast, *errors of estimation* are the degree of misfit between the sample model implied covariance matrix $\hat{\Sigma}$ and the population model implied covariance matrix $\tilde{\Sigma}_{\mathbf{0}}$. As with errors of approximation, errors of estimation are calculated via a suitable discrepancy function $F = (\tilde{\Sigma}_{\mathbf{0}}, \hat{\Sigma})$.

Steiger and Lind (1980) and Browne and Cudeck (1993) argue that error of approximation is of key interest in model evaluation and thus define the RMSEA to be $\varepsilon = \sqrt{F_0/df}$ where ε is the population value of the RMSEA. F_0 represents a suitably weighted sum of squared deviations between the population covariance matrix and the covariance matrix implied by the best fit of the hypothesized model to $\Sigma_{\mathbf{0}}$. As an adjustment for parsimony, F_0 is divided by df to incorporate information about model complexity such that a penalty is imposed for more complicated models, as evidenced through fewer df.

F_0 is an unknown population value that must be estimated from the sample data. Because the minimum of the discrepancy function $\hat{F}$ is a biased estimate of this type of error (see McDonald 1989), a less biased estimate of errors of approximation is $\hat{F}_0 = \hat{F} - \frac{df}{N-1}$. This estimate of $\hat{F}_0$ lies at the heart of the estimate of the RMSEA.

The sample estimate of the RMSEA is

$$\hat{\varepsilon} = \sqrt{\frac{\hat{F}_0}{df}} = \sqrt{\frac{T - df}{df(N-1)}} = \sqrt{\frac{\hat{\lambda}}{df(N-1)}}, \tag{2}$$

Given that it is possible that df can exceed T, resulting in a negative numerator, an added condition is that if $T - df$ is equal to or less

than zero, then $\hat{\varepsilon}$ is set to zero (Browne and Cudeck 1993, equation (9)). The sample estimate $\hat{\varepsilon}$ thus ranges from zero to positive infinity; a value of zero denotes perfect model fit, and larger values reflect poorer model fit. General "rules of thumb" have recommended values of less than .05 to denote good model fit, values less than .08 to denote adequate model fit, and values exceeding .10 to denote poor model fit, although clear warnings are given about the subjectivity of these critical values (Browne and Cudeck 1993; Steiger 1989).

A key strength of $\hat{\varepsilon}$ is that, unlike most baseline fit indices, under certain assumptions the sampling distribution is known. Specifically, the asymptotic distribution of $\hat{\varepsilon}$ is a rescaled noncentral chi-square for a given sample size df and noncentrality parameter λ. (For ease of communication, we will refer to this rescaled asymptotic noncentral chi-square distribution as the sampling distribution of $\hat{\varepsilon}$.) Knowledge of this sampling distribution allows for the computation of confidence intervals around the sample point estimate. As is detailed in Browne and Cudeck (1993, equations (10), (11), (14)), these confidence intervals (CIs) are obtained by reference to the noncentral chi-square distribution, which are transformed into the metric of the RMSEA. More specifically, the lower and upper CI values of $\hat{\varepsilon}$ are given as

$$CI = \left(\sqrt{\frac{\hat{\lambda}_L}{df(N-1)}} \quad ; \quad \sqrt{\frac{\hat{\lambda}_U}{df\,(N-1)}} \right), \tag{3}$$

where $\hat{\lambda}_L$ and $\hat{\lambda}_U$ are the respective lower and upper limits of the noncentral chi-square under consideration (Browne and Cudeck 1993, equation (14)).

Given certain assumptions, the noncentral chi-square distribution provides a clear understanding of the sampling distribution of $\hat{\varepsilon}$. This knowledge allows for the computation of both point estimates of the RMSEA and the calculation of the precision of these point estimates via the corresponding confidence intervals. These intervals can be used in a variety of interesting ways, including the evaluation of exact model fit, "close" and "not close" model fit, and the computation of omnibus statistical power (see MacCallum, Browne, and Sugawara 1996 for details about all of these potential uses).

However, the use of the noncentral chi-square distribution relies on the assumptions of no excess multivariate kurtosis, adequate sample size, and errors of approximation that are "not great" relative to errors of estimation (e.g., Browne and Cudeck 1993:141; Steiger et al. 1985:259). Given that any or all of these conditions might be violated in applied research, it is critical that an understanding be gained about the sampling characteristics of the RMSEA point estimates and CIs under less than ideal experimental conditions. These finite sampling characteristics must be better understood prior to using the RMSEA to evaluate model fit in a thoughtful and informed way. Because asymptotic statistical theory might not characterize finite sampling conditions (e.g., small sample size or specific multivariate distribution) (Mariano 1982), computer simulation experiments provide a powerful method with which to empirically examine these types of research questions.

PREVIOUS SIMULATION STUDIES

Although there are a number of well-designed simulation studies examining the finite sampling properties of the likelihood ratio test statistic and a variety of incremental fit indices, much less study has focused on the RMSEA. Furthermore, the limited amount of research that has examined the RMSEA tends to consider only bias in point estimation or issues of statistical power; other important aspects of this index, such as sampling variability and confidence interval estimation, have yet to be closely considered.

Hu and Bentler (1998) presented findings from a simulation study of a variety of model specifications across a number of sample sizes and multivariate distributions. Consistent with the results of Sugawara and MacCallum (1993), they found the sample estimate of the RMSEA to be far less influenced by the type of estimation (ML vs. GLS) compared to other incremental fit indices, particularly at large sample sizes. Furthermore, they found the sample estimate of the RMSEA to be somewhat overestimated at smaller sample sizes, especially for proper model specifications; this overestimation diminished with increasing sample size and increasing misspecification. Finally, results indicated that, relative to other baseline fit indices, the sample estimate of the RMSEA was moderately sensitive to

minor model misspecifications and highly sensitive to more severe misspecifications. These results shed important initial light on the finite sampling behavior of the RMSEA estimate under conditions commonly encountered in applied research. However, certain limitations of this study included lack of consideration of severely misspecified models and no evaluation of the appropriateness of the corresponding confidence intervals.

Fan, Thompson, and Wang (1999) also presented findings from a simulation study that, in part, examined the RMSEA across several experimental conditions. As with Hu and Bentler (1998), Fan et al. primarily focused on assessing bias in the mean of the sample RMSEA estimates. Results suggested that, on average, the RMSEA estimates were overestimated at smaller sample sizes for properly specified models, and this bias decreased with increasing sample size and increasing misspecification. They also found that the sample estimates of the RMSEA were largely unaffected by the method of estimation (ML vs. GLS). Although important, these findings are similarly limited in that severely misspecified models were not examined, nor was accuracy in the calculation of the corresponding confidence intervals.

Finally, in an unpublished dissertation thesis, Hammervold (1998) presented detailed results from an extensive simulation study that examined a large number of fit indices across a variety of model parameterizations, sample sizes, and multivariate distributions. Consistent with prior research, she found that the sample RMSEA estimates were generally inflated at smaller sample sizes, but this inflation became negligible at larger sample sizes. Furthermore, she found that the RMSEA also tended to be inflated in the presence of multivariate nonnormality. However, as with both Hu and Bentler (1998) and Fan et al. (1999), Hammervold did not examine the sampling characteristics of the RMSEA confidence intervals across these same experimental conditions.

Despite the importance of these initial findings, many characteristics of the RMSEA have yet to be closely examined. For example, although initial simulations have studied the mean of the simulated RMSEA sampling distribution, little is known about the sampling variability of the RMSEA across various experimental conditions. Although the mean may be unbiased, the variability may not. More

important, although one of the key advantages of the RMSEA is the ability to estimate confidence intervals, and strong recommendations are made for the widespread use of these intervals in model evaluation, we are aware of no studies that have empirically examined the accuracy of the RMSEA CIs across various finite sampling conditions. If the CIs are estimated with differential accuracy across sample size and model misspecification, this is critical information to consider when using these CIs in practice. Finally, although Hu and Bentler (1998), Fan et al. (1999), and Hammervold (1998) all considered misspecified models, none considered models that were severely misspecified to the degree at which T may no longer follow the referenced noncentral chi-square distribution. This is an important condition to consider given that the estimation of the noncentrality parameter assumes that errors of approximation are "not great" relative to errors of estimation (Steiger et al. 1985).

PROPOSED HYPOTHESES

Drawing on both the statistical theory underlying the estimation of the RMSEA and the empirical simulation results from prior research, we propose three specific research questions about the sample estimation of the RMSEA.

1. *Are the sample estimates of the RMSEA based on the ML fitting function biased as a function of sample size under proper and improper model specification?* Statistical theory predicts that the sample estimates of the RMSEA will be unrelated to sample size at or above some minimum sample size and that this independence should hold for both properly specified and misspecified models given that the misspecification is not excessive. Drawing on both statistical theory and limited prior empirical results, we predict that the mean of the simulated RMSEA sample estimates will not be significantly biased at moderate to large sample sizes for both properly specified and moderately misspecified models, but we expect biased estimation at smaller sample sizes and for the most severely misspecified models.
2. *Does the coverage accuracy of the RMSEA confidence intervals depend on sample size under proper or improper specification?* As with point estimation, statistical theory predicts that at some minimum sample size, the confidence intervals will cover the known population value at the expected nominal rate across sample size and

misspecification, again assuming that the misspecification is not excessive. We are aware of no prior empirical studies of the CIs to help inform our research hypotheses. We thus expect that the confidence intervals will be accurate across all sample sizes and model specifications, but bias of unknown magnitude is expected for the more severely misspecified models at smaller sample sizes.

3. *Is the accuracy of the confidence intervals equivalent regardless of the percent coverage?* It is common to report 90 percent confidence intervals for the RMSEA, primarily because of the resulting direct link to hypothesis testing based on the model test statistic (for further details, see Browne and Cudeck 1993:145). However, there may be situations in which we want to consider intervals that represent 80 or 95 percent coverage, and it is important to examine the accuracy of estimation at these coverage levels as well. Again, we are aware of no prior empirical or theoretical work to help inform us about this question. We thus expect that the confidence intervals will be comparably accurate regardless of the percentage coverage.

METHOD

MODEL TYPES AND EXPERIMENTAL CONDITIONS

One of our guiding goals for this study was to identify population models that would allow us to maximize the external validity of resulting findings (for further details, see Curran et al. 2002; Paxton et al. 2001). To accomplish this, we reviewed five years of key journals within several areas of social science research to catalog the most common types of SEM applications. Using this information in combination with our own modeling experience, we selected three general model types for study: Model 1 (see Figure 1) contains 9 measured variables and three latent factors with three to four indicators per factor, Model 2 (see Figure 2) has 15 measured variables and three latent factors with five to six indicators per factor, and Model 3 (see Figure 3) consists of 13 measured variables with the same form as Model 1 but with the addition of four measured and correlated exogenous variables. We designed these models to represent features that are commonly encountered in social science research. Furthermore, for each model, we use one correct and four incorrect specifications, resulting in a total of 15 individual models.

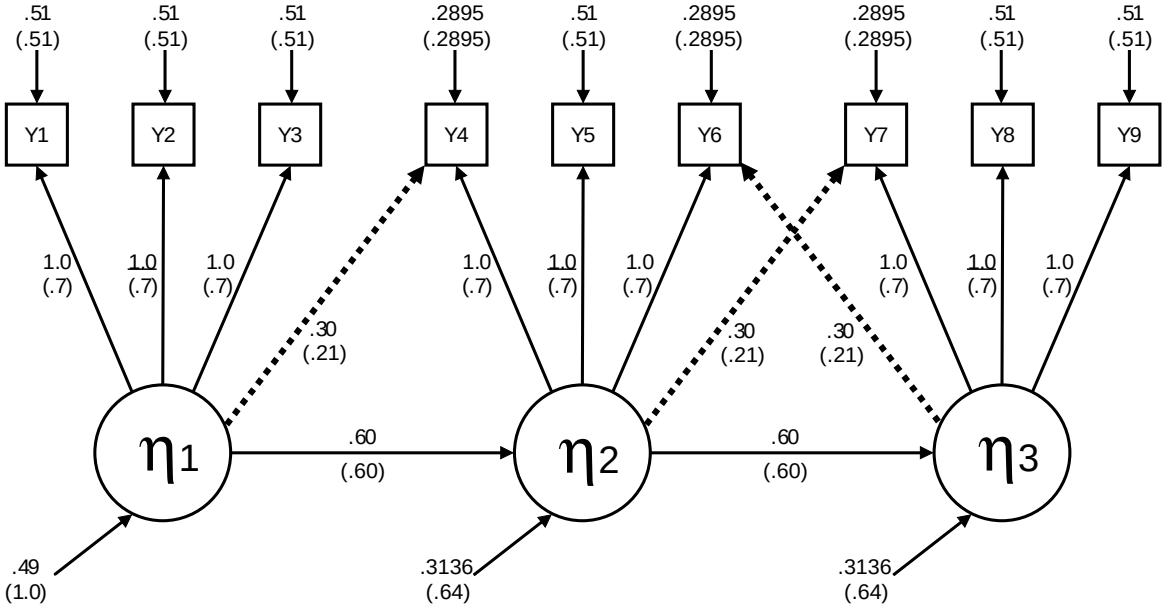

Figure 1: Target Population Model 1

NOTE: Numbers shown are unstandardized parameter values with standardized values in parentheses; solid and dashed lines represent the population model structure, and dashed lines represent omitted parameters under model misspecification.

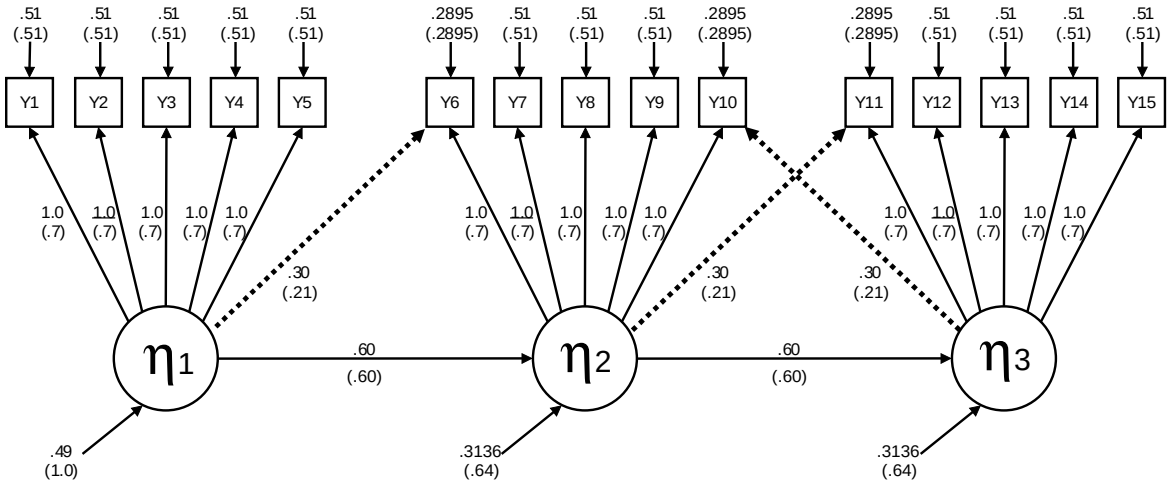

Figure 2: Target Population Model 2

NOTE: Numbers shown are unstandardized parameter values with standardized values in parentheses; solid and dashed lines represent the population model structure, and dashed lines represent omitted parameters under model misspecification.

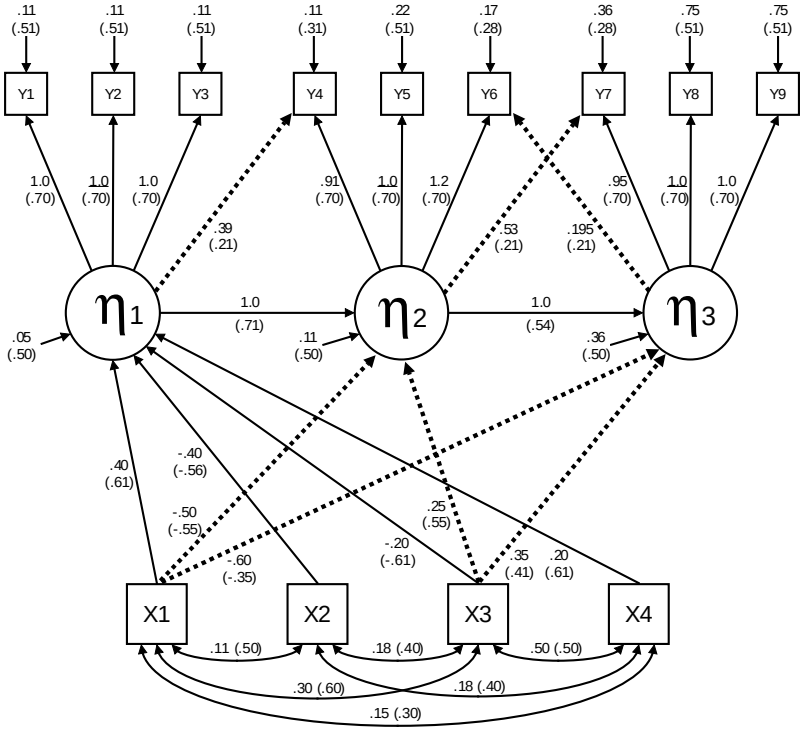

Figure 3: Target Population Model 3

NOTE: Numbers shown are unstandardized parameter values with standardized values in parentheses; solid and dashed lines represent the population model structure, and dashed lines represent omitted parameters under model misspecification.

Model 1. Specification 1 is a *properly* specified model such that the estimated model matches the population model; Specification 2 omits the complex loading linking Item 7 with Factor 2, Specification 3 also omits the complex loading linking Item 6 with Factor 3, Specification 4 also removes the complex loading linking Item 4 with Factor 1, and Specification 5 is the standard uncorrelated variables model in which variances are estimated but all covariances are fixed at zero.

Model 2. Specification 1 is properly specified, Specification 2 omits the complex loading linking Item 11 with Factor 2, Specification 3 also

omits the complex loading linking Item 10 with Factor 3, Specification 4 also removes the complex loading linking Item 6 with Factor 1, and Specification 5 is the standard uncorrelated variables model.

Model 3. Specification 1 is properly specified, Specification 2 jointly omits the set of three complex factor loadings (Item 7 with Factor 2, Item 6 with Factor 3, and Item 4 with Factor 1), Specification 3 jointly omits the set of four regression parameters (Factor 2 regressed on Predictor 1, Factor 3 regressed on Predictor 1, Factor 2 regressed on Predictor 3, and Factor 3 regressed on Predictor 3), Specification 4 jointly combines the omissions of Specifications 2 and 3 (omission of the set of three factor loadings and the set of four regression parameters), and Specification 5 is the standard uncorrelated variables model.

Model parameterization. For all three model types, parameter values were carefully selected to result in a range of effect sizes (e.g., communalities and R^2 values ranging from 49 to 72 percent) and, for the misspecified conditions, to lead to both a wide range of power to detect the misspecifications (e.g., power estimates computed using the method of Satorra and Saris [1985] ranged from .07 to 1.0 across all sample sizes) and a range of bias in parameter estimates (e.g., absolute bias ranged from 0 to 37 percent). See Paxton et al. (2001) for a comprehensive description of our model parameterization. We believe this parameterization reflects values commonly encountered in applied research and that the omission of one or more parameters would result in meaningful impacts on parameter estimation and overall model fit.

Sample size. We chose seven sample sizes to represent those commonly encountered in applied research, and these range from very small to large: 50, 75, 100, 200, 400, 800, and 1,000.

Data generation and estimation. We used the simulation feature in Version 5.7*b* of EQS (Bentler 1995) to generate the raw data and EQS's maximum likelihood estimation to fit the sample models to the resulting covariance matrices. Population values for each parameter were used as initial start values, and a maximum of 100 iterations was allowed to achieve convergence.

Distribution. We generated data from a multivariate normal distribution.

Replications. There were a total of 105 experimental conditions (three models, five specifications, and seven sample sizes), and we generated up to 500 replications for each condition.

Convergence. We eliminated any replication that failed to converge within 100 iterations or did converge but resulted in an out-of-bounds parameter estimate (e.g., "Heywood Case"). We adopted this strategy because the research hypotheses were directly related to proper solutions in SEM, and the external validity of findings would be threatened with the inclusion of improper solutions.[2] To maintain 500 replications per condition, we generated an initial set of up to 650 replications. We then fit the models to the generated data and selected the first 500 proper solutions or selected as many proper solutions as existed when the total number of replications was reached. This resulted in 500 proper solutions for all properly specified and most misspecified experimental conditions, but several misspecified conditions resulted in fewer than 500 proper solutions. Of the 105 experimental conditions, 82 (78 percent) contained 500 replications, and 23 (22 percent) contained fewer than 500 replications. Of those 23 conditions containing fewer than 500 replications, the number of replications ranged from 443 and 499 with a median of 492, and the smallest number of 443 replications was for Model 3, Specification 4, $n = 50$.

Outcome measures. The outcome measures studied here are the RMSEA point estimates and associated 80, 90, and 95 percent confidence intervals. These values were computed in SAS Version 8 (SAS 2000) using the computational formulas presented in Browne and Cudeck (1993) based on the maximum likelihood fit function minima computed by EQS Version 5.7*b*. We cross-validated the accuracy of these computations by comparing the computed values with those produced by EQS, and all point estimates and all 90 percent confidence intervals were equal for all experimental conditions.[3]

ANALYTIC PLAN

As with any large simulation study, there were far too many empirical results to present in full detail within the scope of a single article. We thus present the information in three different ways. First, detailed results are presented for findings that are particularly critical in our evaluation of the research hypotheses. Second, less critical

information is presented in summarized form, either in tables or in figures. Finally, full and comprehensive results are presented in a technical appendix that may be obtained from the first author or may be downloaded directly from www.unc.edu/∼curran.

RESULTS

The general organization of our findings is as follows. First, we report the population values for the ML fitting function, the noncentrality parameter, and the population RMSEA across all model specifications under study. Next, we examine $\hat{\varepsilon}$, the sample estimates of ε from the simulations. We consider the means, standard deviations, relative bias, and root mean square error (denoted root MSE to avoid confusion with RMSEA) of $\hat{\varepsilon}$ across all design factors. Finally, we evaluated the accuracy of the simulated 80, 90, and 95 percent confidence intervals in terms of the percentage coverage of ε (the population RMSEA) relative to the expected nominal coverage rate.

POPULATION VALUES

Table 1 summarizes the population values of the maximum likelihood fitting function (F_{ML}), the noncentrality parameter (λ), and the RMSEA (ε); the number of manifest variables within each model type (9 for Model 1, 15 for Model 2, and 13 for Model 3); and the model df for each specification for each model type (dfs ranging from 22 to 105). Several characteristics of ε are immediately evident. First, ε is zero for all three properly specified models. This is because the noncentrality parameter λ is zero under proper model specification, thus making the numerator term of ε zero (see equation (2) above). Second, for the misspecified conditions, ε does not vary as a function of sample size *within* model type. For example, for Specification 3 of Model 1, ε is equal to .04 regardless of sample size. Note that this characteristic differs from the expected value of the likelihood ratio test statistic T for misspecified models in which the expected value does vary as a function of sample size.

ε values range from a minimum of zero for the properly specified models up to a maximum of .315 for the null uncorrelated variable specification of Model 1. The ε values range from .021 to .097 for the

TABLE 1: Population Minima (F_{ML}), Noncentrality Parameter (λ), Population RMSEA (ε), and Sample RMSEA ($\hat{\varepsilon}$)

n	*Number of Estimated Parameters*	df	*Population* F_{ML}	*Noncentrality* λ	*Population RMSEA*	*Sample RMSEA*		*Root MSE*	*Relative Bias*
						Mean	(SD)		
Model 1 (9 manifest variables): Specification 1									
50	23	22	0.000	0.000	0.000	0.039	(0.044)	0.059	–
75	23	22	0.000	0.000	0.000	0.030	(0.033)	0.045	–
100	23	22	0.000	0.000	0.000	0.027	(0.030)	0.040	–
200	23	22	0.000	0.000	0.000	0.017	(0.020)	0.026	–
400	23	22	0.000	0.000	0.000	0.011	(0.014)	0.017	–
800	23	22	0.000	0.000	0.000	0.007	(0.009)	0.012	–
1,000	23	22	0.000	0.000	0.000	0.007	(0.008)	0.011	–
Model 1 (9 manifest variables): Specification 2									
50	22	23	0.017	0.810	0.027	0.045	(0.045)	0.049	65.849
75	22	23	0.017	1.230	0.027	0.035	(0.035)	0.035	29.750
100	22	23	0.017	1.640	0.027	0.033	(0.032)	0.032	21.342
200	22	23	0.017	3.300	0.027	0.025	(0.023)	0.023	−8.261
400	22	23	0.017	6.610	0.027	0.024	(0.017)	0.017	−8.976
800	22	23	0.017	13.230	0.027	0.025	(0.011)	0.011	−6.824
1,000	22	23	0.017	16.540	0.027	0.024	(0.010)	0.011	−10.904

TABLE 1: Continued

n	*Number of Estimated Parameters*	df	*Population* F_{ML}	*Noncentrality* λ	*Population RMSEA*	*Sample RMSEA*		*Root MSE*	*Relative Bias*
						Mean	(SD)		
Model 1 (9 manifest variables): Specification 3									
50	21	24	0.038	1.840	0.040	0.049	(0.045)	0.046	23.010
75	21	24	0.038	2.780	0.040	0.040	(0.035)	0.035	1.592
100	21	24	0.038	3.720	0.040	0.040	(0.033)	0.033	0.818
200	21	24	0.038	7.490	0.040	0.035	(0.023)	0.024	−11.183
400	21	24	0.038	15.010	0.040	0.037	(0.016)	0.016	−6.035
800	21	24	0.038	30.060	0.040	0.039	(0.009)	0.009	−2.410
1,000	21	24	0.038	37.590	0.040	0.038	(0.008)	0.008	−3.961
Model 1 (9 manifest variables): Specification 4									
50	20	25	0.094	4.630	0.061	0.064	(0.045)	0.045	3.683
75	20	25	0.094	6.990	0.061	0.056	(0.037)	0.038	−9.207
100	20	25	0.094	9.350	0.061	0.059	(0.033)	0.033	−3.833
200	20	25	0.094	18.790	0.061	0.059	(0.020)	0.020	−4.176
400	20	25	0.094	37.670	0.061	0.061	(0.013)	0.013	−1.019
800	20	25	0.094	75.440	0.061	0.061	(0.008)	0.008	−1.090
1,000	20	25	0.094	94.320	0.061	0.060	(0.007)	0.007	−1.611

TABLE 1: Continued

n	*Number of Estimated Parameters*	df	*Population* F_{ML}	*Noncentrality* λ	*Population RMSEA*	*Sample RMSEA*		*Root MSE*	*Relative Bias*
						Mean	(SD)		
Model 1 (9 manifest variables): Specification 5									
50	9	36	3.569	174.860	0.315	0.315	(0.032)	0.032	−0.027
75	9	36	3.569	264.070	0.315	0.315	(0.026)	0.026	−0.079
100	9	36	3.569	353.280	0.315	0.315	(0.022)	0.022	0.008
200	9	36	3.569	710.130	0.315	0.315	(0.016)	0.016	−0.069
400	9	36	3.569	1423.830	0.315	0.314	(0.011)	0.011	−0.226
800	9	36	3.569	2851.240	0.315	0.314	(0.008)	0.008	−0.151
1,000	9	36	3.569	3564.940	0.315	0.314	(0.007)	0.007	−0.117
Model 2 (15 manifest variables): Specification 1									
50	35	85	0.000	0.000	0.000	0.051	(0.031)	0.060	–
75	35	85	0.000	0.000	0.000	0.032	(0.026)	0.041	–
100	35	85	0.000	0.000	0.000	0.025	(0.022)	0.033	–
200	35	85	0.000	0.000	0.000	0.014	(0.015)	0.021	–
400	35	85	0.000	0.000	0.000	0.008	(0.010)	0.013	–
800	35	85	0.000	0.000	0.000	0.006	(0.007)	0.009	–
1,000	35	85	0.000	0.000	0.000	0.005	(0.006)	0.008	–

TABLE 1: Continued

n	*Number of Estimated Parameters*	df	*Population* F_{ML}	*Noncentrality* λ	*Population RMSEA*	*Sample RMSEA*		*Root MSE*	*Relative Bias*
						Mean	(SD)		
Model 2 (15 manifest variables): Specification 2									
50	34	86	0.040	1.940	0.021	0.055	(0.031)	0.046	156.540
75	34	86	0.040	2.930	0.021	0.037	(0.026)	0.030	70.199
100	34	86	0.040	3.920	0.021	0.030	(0.023)	0.024	41.106
200	34	86	0.040	7.880	0.021	0.022	(0.016)	0.016	4.032
400	34	86	0.040	15.810	0.021	0.020	(0.011)	0.011	−8.225
800	34	86	0.040	31.650	0.021	0.021	(0.007)	0.007	−3.165
1,000	34	86	0.040	39.580	0.021	0.021	(0.005)	0.005	−2.265
Model 2 (15 manifest variables): Specification 3									
50	33	87	0.083	4.050	0.031	0.059	(0.030)	0.041	91.722
75	33	87	0.083	6.120	0.031	0.041	(0.026)	0.028	34.022
100	33	87	0.083	8.190	0.031	0.037	(0.022)	0.023	19.240
200	33	87	0.083	16.470	0.031	0.031	(0.015)	0.015	−0.170
400	33	87	0.083	33.020	0.031	0.029	(0.010)	0.010	−4.521
800	33	87	0.083	66.120	0.031	0.031	(0.005)	0.005	−0.585
1,000	33	87	0.083	82.670	0.031	0.031	(0.004)	0.004	−0.912

TABLE 1: Continued

n	*Number of Estimated Parameters*	*df*	*Population* F_{ML}	*Noncentrality* λ	*Population RMSEA*	*Sample RMSEA*		*Root MSE*	*Relative Bias*
						Mean	(*SD*)		
Model 2 (15 manifest variables): Specification 4									
50	32	88	0.141	6.900	0.040	0.064	(0.029)	0.038	59.460
75	32	88	0.141	10.420	0.040	0.048	(0.025)	0.026	20.307
100	32	88	0.141	13.940	0.040	0.044	(0.022)	0.022	9.833
200	32	88	0.141	28.020	0.040	0.040	(0.013)	0.013	−0.620
400	32	88	0.141	56.180	0.040	0.039	(0.008)	0.008	−1.505
800	32	88	0.141	112.500	0.040	0.040	(0.004)	0.004	−0.288
1,000	32	88	0.141	140.670	0.040	0.040	(0.004)	0.004	−0.226
Model 2 (15 manifest variables): Specification 5									
50	15	105	6.694	328.000	0.252	0.258	(0.022)	0.023	2.089
75	15	105	6.694	495.340	0.252	0.253	(0.018)	0.018	0.373
100	15	105	6.694	662.690	0.252	0.253	(0.015)	0.015	0.310
200	15	105	6.694	1332.070	0.252	0.252	(0.011)	0.011	−0.086
400	15	105	6.694	2670.840	0.252	0.253	(0.008)	0.008	0.013
800	15	105	6.694	5348.380	0.252	0.252	(0.006)	0.006	−0.153
1,000	15	105	6.694	6687.150	0.252	0.252	(0.005)	0.005	−0.239

TABLE 1: Continued

n	*Number of Estimated Parameters*	*df*	*Population* F_{ML}	*Noncentrality* λ	*Population RMSEA*	*Sample RMSEA*		*Root MSE*	*Relative Bias*
						Mean	(*SD*)		
Model 3 (13 manifest variables): Specification 1									
50	41	50	0.000	0.000	0.000	0.048	(0.038)	0.061	–
75	41	50	0.000	0.000	0.000	0.033	(0.029)	0.043	–
100	41	50	0.000	0.000	0.000	0.023	(0.024)	0.033	–
200	41	50	0.000	0.000	0.000	0.015	(0.017)	0.023	–
400	41	50	0.000	0.000	0.000	0.009	(0.011)	0.014	–
800	41	50	0.000	0.000	0.000	0.006	(0.008)	0.010	–
1,000	41	50	0.000	0.000	0.000	0.006	(0.007)	0.009	–
Model 3 (13 manifest variables): Specification 2									
50	38	53	0.126	6.190	0.049	0.067	(0.037)	0.041	36.612
75	38	53	0.126	9.340	0.049	0.054	(0.029)	0.029	9.846
100	38	53	0.126	12.500	0.049	0.049	(0.024)	0.024	0.878
200	38	53	0.126	25.120	0.049	0.048	(0.016)	0.016	−1.080
400	38	53	0.126	50.370	0.049	0.048	(0.009)	0.009	−2.515
800	38	53	0.126	100.870	0.049	0.049	(0.005)	0.005	0.295
1,000	38	53	0.126	126.120	0.049	0.048	(0.005)	0.005	−1.098

TABLE 1: Continued

n	*Number of Estimated Parameters*	*df*	*Population* F_{ML}	*Noncentrality* λ	*Population RMSEA*	*Sample RMSEA*		*Root MSE*	*Relative Bias*
						Mean	(SD)		
Model 3 (13 manifest variables): Specification 3									
50	37	54	0.385	18.840	0.084	0.098	(0.030)	0.033	15.756
75	37	54	0.385	28.460	0.084	0.089	(0.023)	0.023	5.741
100	37	54	0.385	38.070	0.084	0.085	(0.019)	0.019	0.468
200	37	54	0.385	76.530	0.084	0.085	(0.011)	0.011	0.462
400	37	54	0.385	153.440	0.084	0.084	(0.007)	0.007	−0.212
800	37	54	0.385	307.270	0.084	0.084	(0.005)	0.005	−0.404
1,000	37	54	0.385	384.190	0.084	0.084	(0.004)	0.004	0.041
Model 3 (13 manifest variables): Specification 4									
50	34	57	0.537	26.310	0.097	0.108	(0.029)	0.031	11.709
75	34	57	0.537	39.730	0.097	0.101	(0.021)	0.021	3.732
100	34	57	0.537	53.150	0.097	0.098	(0.018)	0.018	0.521
200	34	57	0.537	106.840	0.097	0.097	(0.011)	0.011	0.118
400	34	57	0.537	214.220	0.097	0.097	(0.007)	0.007	−0.504
800	34	57	0.537	428.970	0.097	0.097	(0.005)	0.005	−0.056
1,000	34	57	0.537	536.340	0.097	0.097	(0.004)	0.004	−0.147

TABLE 1: Continued

n	*Number of Estimated Parameters*	df	*Population* F_{ML}	*Noncentrality* λ	*Population RMSEA*	*Sample RMSEA*		*Root MSE*	*Relative Bias*
						Mean	(SD)		
Model 3 (13 manifest variables): Specification 5									
50	13	78	5.486	268.810	0.265	0.271	(0.020)	0.021	2.368
75	13	78	5.486	405.960	0.265	0.267	(0.016)	0.016	0.826
100	13	78	5.486	543.110	0.265	0.265	(0.014)	0.014	0.103
200	13	78	5.486	1091.710	0.265	0.266	(0.010)	0.010	0.140
400	13	78	5.486	2188.910	0.265	0.265	(0.007)	0.007	0.073
800	13	78	5.486	4383.310	0.265	0.265	(0.005)	0.005	0.071
1,000	13	78	5.486	5480.510	0.265	0.265	(0.004)	0.004	0.067

nine misspecified conditions (not including the uncorrelated variables model) with a median of .040. These values represent a range of misspecifications from properly specified to minor, moderate, and severe misspecifications. It is these ε parameter values that are being estimated by the sample RMSEA statistics, $\hat{\varepsilon}$, to which we now turn.

SAMPLING DISTRIBUTION OF $\hat{\varepsilon}$

Table 1 also provides information on the sampling distributions of $\hat{\varepsilon}$, the sample RMSEA estimates, within each experimental condition and statistics that enable us to compare them to the corresponding ε, the population value of the RMSEA. Comparing the mean of $\hat{\varepsilon}$ to the corresponding ε enables us to estimate the degree of bias in the sample estimates of the RMSEA. Relative bias was computed as the empirical mean of $\hat{\varepsilon}$ minus ε divided by ε and multiplied by 100. This provides a measure of the percentage of bias of the sample estimate of the RMSEA relative to the population value within each condition. For the properly specified conditions (Specification 1 of Models 1, 2, and 3), ε was zero, and relative bias is undefined because of the division by zero, so relative bias is not reported, although the absolute bias is evident from comparing the mean of $\hat{\varepsilon}$ to the corresponding ε. Table 1 also reports the standard deviation of the simulated $\hat{\varepsilon}$ and the root mean squared error (or root MSE) that is the square root of the sum of the squared deviations between $\hat{\varepsilon}$ and ε within each experimental condition. Root MSE gives a summary measure of accuracy in that it takes account of both bias and variance of an estimate. To further aid the interpretation of these results, the box plots of the simulated RMSEAs are presented in Figures 4, 5, and 6.

Properly specified models. To begin, consider the pattern of means of $\hat{\varepsilon}$ for the three properly specified conditions across sample size (the specific values are presented in Table 1, and box plots of the simulated sampling distributions of $\hat{\varepsilon}$ are in Figures 4-6). Although the value of ε within each model type is equal to zero, the means of the simulated sampling distributions of $\hat{\varepsilon}$ clearly vary as a function of sample size. Consider Specification 1 of Model 1. Although ε is zero, the mean of $\hat{\varepsilon}$ for $n = 50$ is .039, with a large standard deviation of .044 and root MSE of .059. Indeed, 55 percent of the RMSEA sample estimates were greater than zero. Similarly, the sample mean

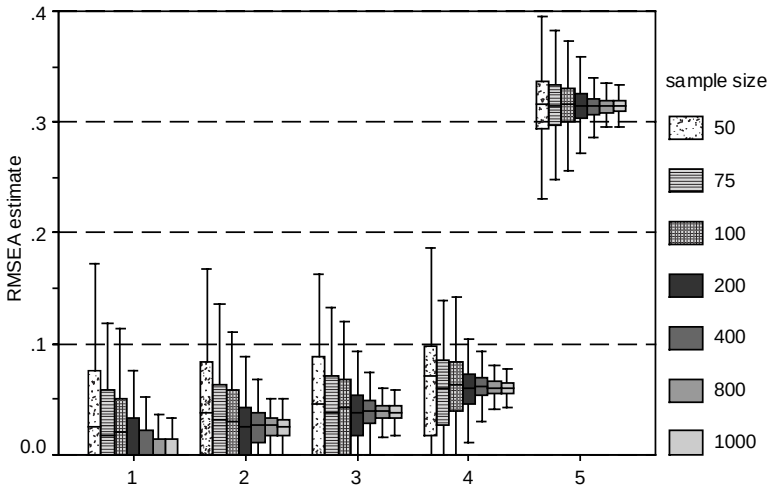

Figure 4: RMSEA Estimates, Model 1, Specifications 1-5, Sample Sizes 50-1,000

NOTE: The horizontal line within the box denote the sample median, the box length is the interquartile range, and the whiskers reflect the largest and smallest observed values that are less than 1.5 box lengths from either end of the distribution. RMSEA = root mean square error of approximation.

of $\hat{\varepsilon}$ at $n = 100$ is .027 ($SD = .030$; root MSE = .045; 54 percent > 0), at $n = 400$ is .011 ($SD = .014$; root MSE = .017; 48 percent > 0), and at $n = 1,000$ is .007 ($SD = .009$; root MSE = .011; 46 percent > 0). A highly similar pattern exists for Specification 1 of Models 2 and 3 as well, and these results will not be detailed here. Thus, consistent with limited prior research findings, under proper specification, $\hat{\varepsilon}$ tends to substantially overestimate ε at smaller sample sizes, and this overestimation decreases with increasing sample size. Specifically, the bias is present but negligible at sample sizes of $n = 200$ or greater.

Improperly specified models. Next consider the means of the simulated sampling distributions of $\hat{\varepsilon}$ for the misspecified conditions. For the misspecified models, we are able to compute bias and relative bias given the nonzero values of ε. Although the mean of $\hat{\varepsilon}$ continues to be inflated relative to ε at smaller sample sizes, the magnitude of this effect is less evident compared to the properly specified

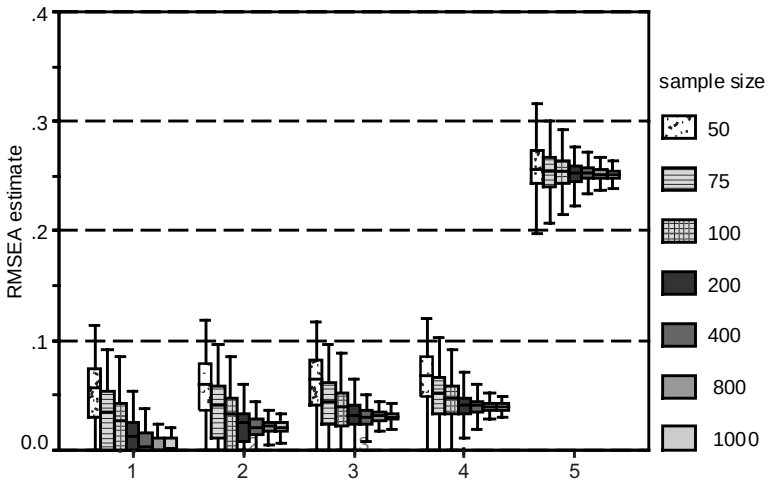

Figure 5: RMSEA Estimates, Model 2, Specifications 1-5, Sample Sizes 50-1,000

NOTE: The horizontal line within the box denote the sample median, the box length is the interquartile range, and the whiskers reflect the largest and smallest observed values that are less than 1.5 box lengths from either end of the distribution. RMSEA = root mean square error of approximation.

conditions described above. Across all three model types, there is clear evidence of significant overestimation at the smallest sample sizes. For example, for the minor misspecification (Specification 2) of Model 1, the ε is overestimated on average by 66 percent at $n = 50$ (mean = .045, SD = .04, root MSE = .049) and 21 percent at $n = 100$ (mean = .033, SD = .032, root MSE = .032). Similarly, for Specification 3 of Model 2, ε is overestimated on average by 92 percent at $n = 50$ (mean = .059, SD = .03, root MSE = .041) and 19 percent at $n = 100$ (mean = .037, SD = .022, root MSE = .023). However, this overestimation at smaller sample sizes quickly diminishes at moderate to large sample sizes. For example, no misspecified condition exceeded approximately 10 percent absolute relative value bias at sample sizes of $n = 200$ or above for any model type.

As we expected based on limited prior research on the RMSEA, it is clear that the sample overestimation of the corresponding ε not only decreases with increasing sample size but also decreases with

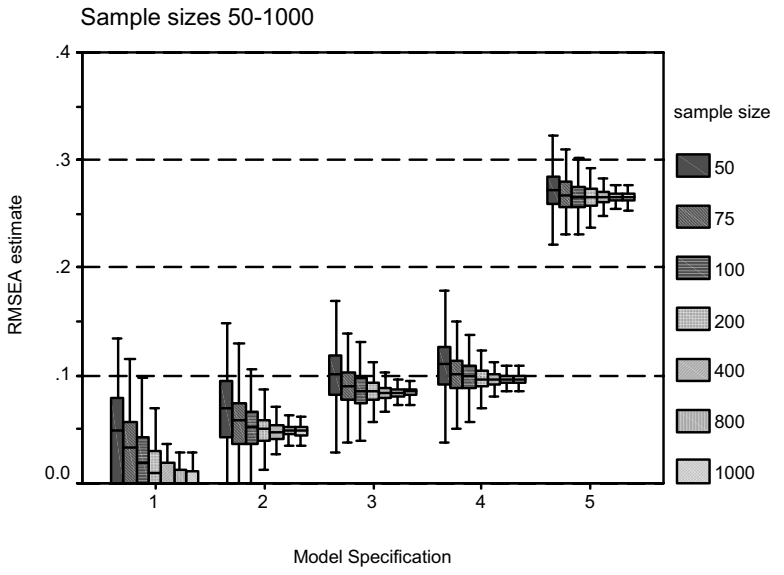

Figure 6: RMSEA Estimates, Model 3, Specifications 1-5, Sample Sizes 50-1,000

NOTE: The horizontal line within the box denote the sample median, the box length is the interquartile range, and the whiskers reflect the largest and smallest observed values that are less than 1.5 box lengths from either end of the distribution. RMSEA = root mean square error of approximation.

increasing misspecification. This is also evident in the reduction in root MSE with increasing sample size. To observe this effect graphically, consider the first panel of Figure 4 in which the horizontal line within the box denotes the sample median, the box length is the interquartile range, and the whiskers reflect the largest and smallest observed values that are less than 1.5 box lengths from either end of the distribution. It can be seen that the sample mean of $\hat{\varepsilon}$ (which is quite close to the median value shown in the plot) converges to ε at approximately $n = 200$ for a minor misspecification and $n = 75$ for a moderate misspecification, and it appears unbiased even at $n = 50$ for the severe misspecification. What is important to note is that although the sample means do demonstrate convergence on the corresponding expected value, the sampling variability is still quite high, especially at smaller sample sizes. For example, the sample

mean of $\hat{\varepsilon}$ for the severely misspecified condition of Model 1 (Specification 4) at $n = 50$ shows little mean bias (sample mean = .064, parameter value = .061, relative bias = 3.7 percent), but the sample standard deviation is .045 with a point estimate range of 0 up to .187. Thus, although the means of the RMSEA estimates at smaller sample sizes are unbiased under increasing model misspecification, there remains substantial variability in the sample estimates.

The standard deviations of the simulated sampling distributions of $\hat{\varepsilon}$ decrease for each model and specification as the sample size increases. This expected pattern reflects the well-known pattern of decreased variability as the sample size increases. It also is interesting to note that the growth in the mean value in $\hat{\varepsilon}$ as the degree of misspecification increases is not accompanied by a growth in its standard deviation for a given sample size. For instance, in Model 2, the standard deviations of $\hat{\varepsilon}$ for Specifications 1 to 4 at $n = 100$ are .022, .023, .022, and .022, respectively. The main exception is in the case of the most severe misspecification, in which the standard deviation of $\hat{\varepsilon}$ tends to be smaller at a given sample size.

The root MSE provides a summary of both the bias and variability of $\hat{\varepsilon}$ as an estimate of the population RMSEA (ε). Perhaps the most obvious trend in the results is the decrease in the root MSE of $\hat{\varepsilon}$ within each model and specification as the sample size grows. Given our observations on the decrease in bias and shrinkage in standard deviations as n is larger, this result for the root MSE is fully expected. Less expected is the tendency for the root MSE of $\hat{\varepsilon}$ to decrease as the degree of misspecification increases. In Model 2 at $n = 100$, for instance, the root MSE of $\hat{\varepsilon}$ is .033 for the correct model and .023 for Specification 3. This is likely to result from the drop in relative bias as the degree of specification error increases.

Uncorrelated variable baseline model. Finally, the results in Table 1 indicate that the mean of the simulated sampling distribution of $\hat{\varepsilon}$ for the uncorrelated variables model is very close to the corresponding ε across all sample sizes for all three model types. For example, for the uncorrelated variable baseline specification of Model 1 at $n = 100$, ε is .315, and the sample mean of $\hat{\varepsilon}$ is .315; for Model 2 at $n = 200$, ε is .252, and the sample mean of $\hat{\varepsilon}$ is .252; finally, for Model 3 at $n = 400$, ε is .265, and the sample mean of $\hat{\varepsilon}$ is .265. These results indicate that for the uncorrelated variable baseline model, the

mean of the simulated $\hat{\varepsilon}$ is an unbiased estimate of the corresponding ε. However, as was found for the misspecified models, the simulated sampling distribution of the RMSEA was also characterized by a rather large variance, especially at smaller sample sizes. We will explore this finding more thoroughly in a moment.

SUMMARY

Overall, the mean of the simulated sampling distribution of $\hat{\varepsilon}$ tended to overestimate the corresponding ε, particularly at smaller sample sizes and for the properly specified and less severe misspecified conditions. It is evident from the plots in Figure 4 that the overestimation at smaller sample sizes and under less severe misspecification is attributable to the higher number of sample estimates that were estimated to be negative but were set to zero due to the restricted lower bound of the RMSEA of 0. Although this overestimation decreased with increasing sample size and increasing misspecification, there remained substantial sampling variability in $\hat{\varepsilon}$. The sample standard deviations and the root MSE of $\hat{\varepsilon}$ decreased as the sample size grew within a given model and specification. Holding model and sample size constant, there was some tendency for the root MSE to be smaller as the degree of specification error grew. We will now examine the confidence intervals of $\hat{\varepsilon}$.

RMSEA CONFIDENCE INTERVALS

To evaluate the accuracy of the RMSEA confidence intervals, we computed the percentage coverage of ε by the simulated CIs within each experimental condition. If ε for a given specification fell within the bounds of the simulated CI for a given replication, then that CI was deemed "accurate," given that it properly covered the corresponding ε value. In the long run, it is expected that the percentage of CIs that cover the corresponding ε will converge on the nominal rate of the CI (i.e., 90 percent of the simulated 90 percent CIs are expected to contain the associated ε). The empirical coverage rates and mean interval widths are presented in Table 2.

TABLE 2: Coverage Rates for 80, 90, and 95 Percent Confidence Intervals (CIs) and Interval Width for 90 Percent CI

				90% *CI Interval Width*	
n	80% *Coverage Rate*	90% *Coverage Rate*	95% *Coverage Rate*	*Mean*	(*SD*)
Model 1 (9 manifest variables): Specification 1					
50	0.762	0.876	0.922	0.114	(0.038)
75	0.786	0.894	0.946	0.092	(0.034)
100	0.762	0.880	0.938	0.080	(0.028)
200	0.786	0.890	0.950	0.055	(0.020)
400	0.802	0.900	0.950	0.038	(0.014)
800	0.818	0.914	0.960	0.027	(0.010)
1,000	0.802	0.892	0.960	0.023	(0.009)
Model 1 (9 manifest variables): Specification 2					
50	0.754	0.881	0.938	0.117	(0.036)
75	0.785	0.888	0.945	0.095	(0.030)
100	0.769	0.878	0.938	0.083	(0.025)
200	0.814	0.900	0.958	0.061	(0.016)
400	0.802	0.906	0.944	0.045	(0.010)
800	0.810	0.904	0.950	0.033	(0.005)
1,000	0.774	0.884	0.956	0.029	(0.004)
Model 1 (9 manifest variables): Specification 3					
50	0.769	0.890	0.944	0.119	(0.035)
75	0.799	0.894	0.943	0.097	(0.029)
100	0.792	0.897	0.940	0.086	(0.024)
200	0.826	0.916	0.962	0.064	(0.012)
400	0.798	0.888	0.938	0.046	(0.007)
800	0.790	0.900	0.950	0.029	(0.003)
1,000	0.780	0.888	0.950	0.025	(0.002)
Model 1 (9 manifest variables): Specification 4					
50	0.779	0.878	0.940	0.125	(0.031)
75	0.809	0.894	0.933	0.103	(0.025)
100	0.780	0.889	0.946	0.091	(0.017)
200	0.810	0.918	0.958	0.064	(0.009)
400	0.772	0.882	0.946	0.039	(0.004)
800	0.788	0.910	0.954	0.025	(0.001)
1,000	0.774	0.888	0.934	0.022	(0.000)

TABLE 2: Continued

n	80% *Coverage Rate*	90% *Coverage Rate*	95% *Coverage Rate*	90% *CI Interval Width*	
				Mean	(*SD*)
Model 1 (9 manifest variables): Specification 5					
50	0.682	0.804	0.874	0.082	(0.001)
75	0.652	0.768	0.850	0.066	(0.000)
100	0.674	0.798	0.872	0.056	(0.000)
200	0.664	0.790	0.876	0.039	(0.000)
400	0.664	0.796	0.876	0.028	(0.000)
800	0.654	0.764	0.850	0.019	(0.000)
1,000	0.656	0.768	0.832	0.017	(0.000)
Model 2 (15 manifest variables): Specification 1					
50	0.584	0.730	0.816	0.088	(0.017)
75	0.706	0.838	0.890	0.068	(0.016)
100	0.734	0.848	0.922	0.057	(0.016)
200	0.796	0.880	0.926	0.038	(0.012)
400	0.816	0.896	0.944	0.026	(0.009)
800	0.824	0.912	0.958	0.018	(0.006)
1,000	0.810	0.914	0.962	0.016	(0.006)
Model 2 (15 manifest variables): Specification 2					
50	0.573	0.719	0.824	0.088	(0.016)
75	0.696	0.832	0.908	0.069	(0.015)
100	0.744	0.862	0.918	0.060	(0.014)
200	0.788	0.884	0.936	0.043	(0.009)
400	0.792	0.896	0.946	0.031	(0.006)
800	0.778	0.896	0.946	0.021	(0.004)
1,000	0.784	0.880	0.944	0.018	(0.003)
Model 2 (15 manifest variables): Specification 3					
50	0.598	0.725	0.831	0.088	(0.015)
75	0.718	0.846	0.916	0.071	(0.013)
100	0.750	0.848	0.916	0.061	(0.012)
200	0.808	0.880	0.940	0.044	(0.007)
400	0.782	0.890	0.930	0.029	(0.005)
800	0.810	0.900	0.950	0.017	(0.002)
1,000	0.802	0.910	0.950	0.014	(0.001)

TABLE 2: Continued

n	80% *Coverage Rate*	90% *Coverage Rate*	95% *Coverage Rate*	90% *CI Interval Width*	
				Mean	(*SD*)
Model 2 (15 manifest variables): Specification 4					
50	0.623	0.754	0.836	0.088	(0.015)
75	0.734	0.854	0.926	0.071	(0.012)
100	0.750	0.864	0.924	0.062	(0.011)
200	0.808	0.904	0.944	0.042	(0.007)
400	0.756	0.882	0.946	0.025	(0.004)
800	0.804	0.902	0.948	0.015	(0.001)
1,000	0.792	0.912	0.956	0.013	(0.000)
Model 2 (15 manifest variables): Specification 5					
50	0.572	0.698	0.788	0.049	(0.001)
75	0.588	0.724	0.822	0.039	(0.000)
100	0.638	0.726	0.814	0.034	(0.000)
200	0.606	0.726	0.804	0.023	(0.000)
400	0.606	0.706	0.792	0.016	(0.000)
800	0.546	0.658	0.760	0.011	(0.000)
1,000	0.582	0.694	0.766	0.010	(0.000)
Model 3 (13 manifest variables): Specification 1					
50	0.646	0.778	0.868	0.098	(0.023)
75	0.742	0.852	0.928	0.077	(0.022)
100	0.808	0.894	0.946	0.065	(0.018)
200	0.782	0.896	0.936	0.045	(0.014)
400	0.800	0.910	0.954	0.030	(0.011)
800	0.816	0.896	0.962	0.021	(0.008)
1,000	0.818	0.900	0.946	0.019	(0.007)
Model 3 (13 manifest variables): Specification 2					
50	0.662	0.797	0.871	0.101	(0.018)
75	0.752	0.859	0.927	0.083	(0.015)
100	0.803	0.902	0.958	0.073	(0.012)
200	0.798	0.892	0.952	0.049	(0.007)
400	0.802	0.908	0.954	0.029	(0.003)
800	0.782	0.898	0.960	0.018	(0.000)
1,000	0.792	0.876	0.934	0.016	(0.000)

TABLE 2: Continued

n	80% *Coverage Rate*	90% *Coverage Rate*	95% *Coverage Rate*	90% *CI Interval Width*	
				Mean	(*SD*)
Model 3 (13 manifest variables): Specification 3					
50	0.719	0.828	0.916	0.097	(0.015)
75	0.791	0.896	0.935	0.075	(0.012)
100	0.790	0.890	0.945	0.062	(0.008)
200	0.810	0.912	0.948	0.037	(0.002)
400	0.848	0.934	0.980	0.024	(0.000)
800	0.820	0.932	0.976	0.017	(0.000)
1,000	0.838	0.918	0.964	0.015	(0.000)
Model 3 (13 manifest variables): Specification 4					
50	0.727	0.833	0.903	0.090	(0.014)
75	0.762	0.867	0.928	0.069	(0.010)
100	0.765	0.879	0.943	0.056	(0.006)
200	0.813	0.908	0.948	0.035	(0.001)
400	0.818	0.912	0.966	0.023	(0.000)
800	0.820	0.916	0.962	0.016	(0.000)
1,000	0.804	0.908	0.958	0.014	(0.000)
Model 3 (13 manifest variables): Specification 5					
50	0.692	0.820	0.888	0.057	(0.001)
75	0.732	0.844	0.914	0.045	(0.000)
100	0.720	0.834	0.906	0.039	(0.000)
200	0.688	0.814	0.896	0.027	(0.000)
400	0.718	0.848	0.914	0.019	(0.000)
800	0.698	0.840	0.920	0.013	(0.000)
1,000	0.712	0.816	0.898	0.012	(0.000)

PROPERLY SPECIFIED MODELS: 90 PERCENT CONFIDENCE INTERVALS

For all three properly specified models, the coverage rates of the ε were remarkably accurate at moderate to large sample sizes. For example, at sample sizes of $n = 200$ and higher, the proportion of simulated CIs correctly containing ε were within 2 percent of the expected nominal 90 percent rate. Although the accuracy degraded somewhat at smaller sample sizes, this decrement in performance varied as a function of model complexity. For Model 1, the CIs were within 2 percent of the nominal rate at all sample sizes, even at the smallest

sample size of $n = 50$. However, this degree of accuracy at the smaller sample sizes did not hold for Models 2 and 3. For example, for the proper specification of Model 2, the ε was accurately covered only 73 percent of the time at $n = 50$, 84 percent at $n = 75$, and 85 percent at $n = 100$; for the proper specification of Model 3, the ε was accurately covered 78 percent of the time at $n = 50$, 85 percent at $n = 75$, and 89 percent at $n = 100$. Thus, the coverage rate converged on the nominal rate at $n = 50$ for Model 1, $n = 200$ for Model 2, and $n = 100$ for Model 3.

It is interesting to note that for those cases in which the simulated CI did not contain the associated ε, the expected value always fell below the *lower* boundary of the CI. There was not a single instance in any replication in which the simulated CI failed to contain the corresponding ε by falling below that value; this was true for all specifications of all models.

MISSPECIFIED MODELS: 90 PERCENT CONFIDENCE INTERVALS

For the three misspecified conditions for each of the three target models (but not yet considering the uncorrelated variable baseline model), the empirical coverage rates were quite similar to those found for the properly specified conditions. For all three misspecified conditions of Model 1, the empirical coverage rates were within 2 percent of the nominal rate across all sample sizes. The nominal rate was again achieved at $n = 200$ and $n = 100$ for all three misspecified conditions of Models 2 and 3, respectively.

However, as was found for the properly specified conditions of Models 2 and 3, the empirical rate fell well below the nominal rate at the smaller sample sizes. For example, for the moderately misspecified condition of Model 2, the empirical CIs only covered the corresponding ε value 73 percent at $n = 50$, 85 percent at $n = 75$ and $n = 100$, and 88 percent at $n = 200$. Similarly, for the moderately misspecified condition of Model 3, the empirical CIs only covered the corresponding ε value 83 percent at $n = 50$ but was within 2 percent of the nominal rates at $n = 75$ and above. Thus, in general, the CIs were quite accurate for all misspecified models at sample sizes of $n = 200$ and higher, and this accuracy was evident at smaller sample sizes but only for the less complex models.

UNCORRELATED VARIABLE BASELINE MODEL: 90 PERCENT CONFIDENCE INTERVALS

Of course, in applied research settings, it would rarely be of interest to compute the RMSEA and associated confidence intervals for the uncorrelated variable baseline model given that this is used primarily as a null model with which to compute the baseline fit indices. However, as described earlier, one of the key assumptions underlying the RMSEA is that the errors of approximation are "not great" relative to errors of estimation. It is thus of interest to examine the finite sampling behavior of the RMSEA CIs for the uncorrelated variable baseline model to examine whether this severe misspecification is too extreme in magnitude to reference the underlying noncentral chi-square distribution on which the computation of the CIs is based.

In general, there was evidence of poor accuracy in CI coverage of the expected value of the RMSEA for the uncorrelated baseline conditions of all three models across all sample sizes. For example, for the uncorrelated variable baseline model for Models 1 and 3, the percentage of 90 percent CIs that cover the known ε is approximately 80 percent across all sample sizes; for Model 2, the coverage rate drops to approximately 70 percent across all sample sizes. Thus, for the uncorrelated variables' specification of all three model types, the empirical coverage of the expected value of the RMSEA falls 10 to 20 percent below the nominal coverage rate. This suggests that, consistent with the results of Curran et al. (2002), the severe misspecification of the uncorrelated baseline model may indeed be too extreme to reference the corresponding noncentral chi-square distribution in the computation of the CIs.

RMSEA 80 AND 95 PERCENT CONFIDENCE INTERVALS

Widespread use of the 90 percent CI is based on the ability to link this CI to the use of the usual likelihood ratio test T. Specifically, if the lower bound of the 90 percent CI is equal to zero, this implies that the probability level of the test statistic T is greater than .05. Thus, the CI provides information about the likelihood ratio test statistic but also provides much more information about model fit beyond the standard p value (for further details, see MacCallum et al. 1996). Although this is a useful link on which to capitalize, there is no reason why any

other CI coverage rate cannot be considered. To better understand the accuracy of the CI estimates as a function of width of coverage, we repeated the above analyses for CIs with widths of 80 and 95 percent. As before, these analyses produced a tremendous amount of empirical results, and we only summarize the key findings here.

In general, it is striking how similar the pattern of findings is for these alternative coverage rates, as were found for the 90 percent CIs presented in detail above. At sample sizes of $n = 200$ and higher, the coverage rates of the CIs were quite consistent with the corresponding expected nominal rate. This consistency was evident for the proper and three improper specifications of all three model types at $n = 200$ and higher. For example, for the moderate misspecification of Model 1 at $n = 200$, 82 percent of the sample 80 percent CIs and 96 percent of the 95 percent CIs covered ε. Similarly, for the severe misspecification of Model 3 at $n = 200$, 81 percent of the sample 80 percent CIs and 95 percent of the sample 95 percent CIs covered ε. At sample sizes less than $n = 200$, the expected nominal rates were uniformly underestimated by anywhere from 5 to 20 percent at the smallest sample size of $n = 50$. This underestimation was somewhat more pronounced for the 80 percent CIs compared to the 95 percent CIs, and this held across all three model types. Finally, as was found with the 90 percent CIs, none of the coverage rates of the CIs corresponded to the expected nominal coverage rates for the uncorrelated variables model for any of the three model types for either the 80 or 95 percent CIs.

DISCUSSION

The goal of our study was to empirically evaluate a set of theoretically generated research hypotheses about the finite sampling properties of the RMSEA. We used Monte Carlo computer simulations to generate data corresponding to three population model types across a variety of experimental conditions commonly encountered in applied social science research. Up to 500 replications were generated for each of 7 sample sizes and 15 model specifications, and we examined the sampling characteristics of both the RMSEA point estimates and the

associated confidence intervals. The results were largely consistent with the proposed hypotheses.

RMSEA POINT ESTIMATES

Consistent with Fan et al. (1999) and Hu and Bentler (1998), we found that for the properly specified conditions of all three model types, the mean of the simulated sampling distribution of $\hat{\varepsilon}$ substantially overestimated the corresponding ε at the smaller sample sizes, but this positive bias became negligible at sample sizes of $n = 200$ and higher. There are two likely sources of this overestimation. First, it is well known that the likelihood ratio test statistic is inflated at smaller sample sizes under correct (e.g., Boomsma 1983; Anderson and Gerbing 1984) and incorrect (e.g., Curran et al. 1996) model specification; this test statistic plays an integral role in the computation of $\hat{\varepsilon}$, and thus the inflation plays a role here as well. Second, ε under proper specification is zero, and $\hat{\varepsilon}$, the sample estimates of the RMSEA, vary around this value. However, given that the computation of the RMSEA is, by definition, fixed to zero when the sample estimate falls below zero, the mean of the estimates is positively biased due to the fixing to zero of these negative values. This explains why we found evidence for decreased bias in $\hat{\varepsilon}$ with increasing sample size, given the corresponding decrease in sampling variability; similar results were described by both Hu and Bentler (1998) and Fan et al. (1999).

Our results further indicated that the overestimation of the RMSEA relative to the corresponding population value decreased with increasing model misspecification. That is, the means of the simulated sampling distributions of $\hat{\varepsilon}$ were systematically larger than the corresponding population values at smaller sample sizes, but this overestimation was attenuated by increasing misspecification. This attenuation of bias was further highlighted in that the severely misspecified null independence model showed no overestimation even at the smallest sample size. In general, though, some evidence of positive bias was found for the less severe misspecifications at sample sizes of less than $n = 200$. As before, the decreasing bias associated with increasing sample size is due in part to the associated decrease in sampling variability of the point estimates. Similarly, the decreasing bias associated with increasing misspecification is due in part to the larger

corresponding population values of the RMSEA and subsequently fewer sample estimates that were negative and thus fixed to zero.

In sum, the means of the simulated sampling distributions of $\hat{\varepsilon}$ were generally unbiased across all model types and all model specifications at sample sizes of $n = 200$ and greater. Overestimation of the corresponding ε was evidenced at samples less than $n = 200$ across all but the null independence model, and this overestimation was oftentimes substantial. Finally, there was no evidence of overestimation for the uncorrelated variables model for any model type at any sample size. It is important to note, however, that all of the findings we have summarized thus far only relate to the sample estimates of the RMSEA. A second critical piece of information to consider is the accuracy of the associated confidence intervals.

RMSEA CONFIDENCE INTERVALS

Although computation of the 90 percent CIs is standard in commercial SEM packages, there is no reason to limit ourselves to this particular coverage rate. We thus examined CIs for the RMSEA for coverage rates of 80, 90, and 95 percent to examine potential departures in finite sampling behavior as a function of nominal coverage. Results were nearly identical across all three coverage rates, and we will thus focus on the 90 percent rate with the understanding that these results directly generalize to the other two rates as well.

In general, the CIs were remarkably accurate across many of the experimental conditions studied here. For example, the sample 90 percent CIs covered ε, the RMSEA parameter value, within 2 percent of the nominal rate across all sample sizes for the properly specified condition of Model 1, as well as at sample sizes of $n = 200$ and above for Models 2 and 3. This same degree of accuracy held for the three misspecified conditions as well, again at all sample sizes for Model 1 and at sample sizes of $n = 200$ and above for Models 2 and 3. This was impressive accuracy, especially given that we considered several of these conditions to be severely misspecified (but note that we have not yet discussed the null independence model). It thus appears that, at least at sample sizes of $n = 200$ or greater, the CIs are characterized by accurate coverage rates under the very conditions in which they are most needed: a moderately to severely misspecified model that we might consider to be theoretically tenable.

However, for sample sizes less than $n = 200$ for the properly specified and three misspecified conditions of Models 2 and 3, the CIs systematically covered the expected value of the RMSEA below the expected nominal rates, sometime substantially so. For several conditions at the smallest sample size, the estimated 90 percent CIs only covered ε 75 percent of the time. Closer examination of the data indicated that for every single case that did not cover ε, the CI was overestimated. That is, in every case, the lower bound of the CI was above ε, and under no condition was the error made as a result of the upper bound falling below this value. In general, the $\hat{\varepsilon}$ estimates were unbiased at sample sizes of $n = 200$ and above; similarly, the CIs were covering ε at the expected nominal rate at sample sizes of $n = 200$ and above. The only exception to this pattern of findings is the accurate coverage of the CIs at smaller sample sizes across all specifications for Model 1.

It is not immediately clear as to why the CIs showed little coverage bias across all sample sizes for Model 1, whereas this lack of coverage bias was only evident at sample sizes of $n = 200$ and greater for Models 2 and 3. One likely explanation for this relates to the magnitude of the noncentrality parameter for the misspecified conditions. Although ε values were rather similar within specification and across model type (e.g., the ε for Specification 2 of Model 1 was .027, and Specification 2 of Model 2 was .021), the noncentrality parameters were often more discrepant (e.g., the noncentrality parameter at $n = 1,000$ for Specification 2 of Model 1 was 16.54, and for Specification 2 of Model 2, it was 39.58). The reason for the differences in relative magnitude between the RMSEA and the corresponding noncentrality parameter is that the RMSEA makes an adjustment for df, and these varied greatly across the model (e.g., Specification 2 of Model 1 was characterized by 23 df compared to 86 df for Specification 2 of Model 2). Given that the CIs are computed based on the noncentral chi-square distribution and then transformed into the metric of RMSEA (see equation (3)), there may be greater overestimation of the sample noncentrality parameter for Models 2 and 3 at the smaller sample size given the larger value of the noncentrality parameter for those same conditions. That is, although several of the RMSEA values were comparable across conditions, there are larger differences in the associated noncentrality parameters, given

that an adjustment has yet to be made for the model df. This is only a hypothesis, and further work is necessary to better understand this effect.

Finally, although an applied researcher would rarely if ever desire a CI for an uncorrelated variables model, it was important to examine the CIs under this condition, given the prediction that the likelihood ratio test statistic will not follow the referenced noncentral chi-square distribution under severe specification error. Results indicated that the CIs for the uncorrelated variables model were substantially biased for all three model types across all sample sizes. For example, even at the largest sample size of $n = 1,000$ for the uncorrelated variable specification of Model 2, only 69 percent of the 90 percent CIs actually covered ε. Note that these inaccuracies in the CIs were present even though the mean of $\hat{\varepsilon}$ reflected no bias relative to the corresponding ε. This highlights the critical importance of examining both the mean and variance of a test statistic in simulation studies. (For a more detailed discussion of this, see Curran et al. 2002.) In conclusion, it appears that the degree of misspecification introduced in the uncorrelated independence model was sufficiently severe that the resulting test statistic no longer referenced the underlying noncentral chi-square distribution, and the resulting CIs were not valid.

POTENTIAL LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

The results of any Monte Carlo study are necessarily limited to the parameterization of the models and conditions under study, and care should be taken when generalizing our findings presented here. However, we took great care in the design of our simulation experiment to maximize external validity, and our selected models and conditions represented an array of SEM applications in the social sciences that ranged from properly specified to severely misspecified and were characterized by a small to large number of df. We thus feel that these findings generalize to many similar types of SEM applications, although future work will do well to consider additional model types as well. One set of models that might be of particular interest that we did not study here are models with a very small df. For example, a three–time point linear latent growth model is characterized by a single df, and we have found in our own applied work that resulting RMSEA values can be quite large for a model that otherwise

appears to fit the data well. Further examination of the RMSEA under conditions such as these could be quite interesting. Finally, we chose to focus on data drawn from a multivariate normal distribution, and it is well known that nonnormally distributed data, particularly data characterized by excess kurtosis, introduce further bias in normal theory ML estimation (Boomsma 1983; Browne 1984; Curran et al. 1996; Muthén and Kaplan 1985, 1992).

Of most importance, there has been much recent discussion of optimal methods for using the RMSEA in practice to evaluate model fit. For example, Hu and Bentler (1998) have explored using just the point estimates of the RMSEA to determine adequacy of model fit, and MacCallum et al. (1996) have proposed the use of tests of close fit and not close fit based on the confidence intervals of the RMSEA. Given that our results presented here provide empirical insight into the experimental conditions under which the point estimates and associated CIs are accurately estimated, future work can turn to a more comprehensive evaluation of how to optimally use these estimates to judge model fit in applied research. We hope that our findings help inform the thoughtful consideration of this important question.

NOTES

1. More generally, structural equation modeling (SEM) focuses on reproducing the population moments of the observed variable with a specified model. To simplify our discussion, we focus on the population covariance matrix here, recognizing that the discussion could be easily generalized to the population mean or other moments.

2. For completeness, all results reported here were reanalyzed for the conditions that were characterized by higher rates of nonconvergence, including both converged and nonconverged cases, and no substantive differences in any conclusions were found. Furthermore, although these nonconverged and improper solution cases were excluded from the present analyses, these cases are of great interest with respect to other related research questions, and we have examined these closely in other studies of these data. See Chen et al. (2001) for further details.

3. The 80 and 95 percent confidence intervals were not cross-validated in this way since all current commercial SEM packages only provide 90 percent CIs.

REFERENCES

Anderson, James C. and David W. Gerbing. 1984. “The Effect of Sampling Error on Convergence, Improper Solutions, and Goodness-of-Fit Indices for Maximum Likelihood Confirmatory Factor Analysis.” *Psychometrika* 49:155-73.

Bentler, Peter M. 1990. "Comparative Fit Indexes in Structural Models." *Psychological Bulletin* 107:238-46.

———. 1995. *EQS: Structural Equations Program Manual, Version 5.0*. Los Angeles: BMDP Statistical Software.

Bentler, Peter M. and Douglass G. Bonett. 1980. "Significance Tests and Goodness-of-Fit in the Analysis of Covariance Structures." *Psychological Bulletin* 88:588-606.

Bollen, Kenneth A. 1989a. *Structural Equations With Latent Variables*. New York: John Wiley.

———. 1989b. "A New Incremental Fit Index for General Structural Equation Models." *Sociological Methods & Research* 17:303-16.

Bollen, Kenneth A. and J. Scott Long. 1993. "Introduction." Pp. 1-9 in *Testing Structural Equation Models*, edited by K. A. Bollen and J. S. Long. Newbury Park, CA: Sage.

Boomsma, Anne. 1983. *On the Robustness of LISREL (Maximum Likelihood Estimation) Against Small Sample Size and Nonnormality*. Amsterdam: Sociometric Research Foundation.

Browne, Michael W. 1984. "Asymptotic Distribution Free Methods in the Analysis of Covariance Structures." *British Journal of Mathematical and Statistical Psychology* 37:127-41.

Browne, Michael W. and Robert Cudeck. 1993. "Alternative Ways of Assessing Model Fit." Pp. 136-62 in *Testing Structural Equation Models*, edited by K. A. Bollen and J. S. Long. Newbury Park, CA: Sage.

Chen, Feinian, Kenneth A. Bollen, Pamela Paxton, Patrick J. Curran, and James Kirby. 2001. "Improper Solutions in Structural Equation Models: Causes, Consequences, and Strategies." *Sociological Methods and Research* 29:468-508.

Curran, Patrick J., Kenneth A. Bollen, Pamela Paxton, James Kirby, and Feinian Chen. 2002. "The Noncentral Chi-Square Distribution in Misspecified Structural Equation Models: Finite Sample Results From a Monte Carlo Simulation." *Multivariate Behavioral Research* 37:1-36.

Curran, Patrick J., Stephen G. West, and John F. Finch. 1996. "The Robustness of Test Statistics to Nonnormality and Specification Error in Confirmatory Factor Analysis." *Psychological Methods* 1:16-29.

Fan, Xitao, B. Thompson, and L. Wang. 1999. "Effects of Sample Size, Estimation Methods, and Model Specification on Structural Equation Modeling Fit Indexes." *Structural Equation Modeling* 6:56-83.

Gerbing, David W. and James C. Anderson. 1993. "Monte Carlo Evaluations of Goodness-of-Fit Indices for Structural Equation Models." Pp. 40-65 in *Testing Structural Equation Models*, edited by K. A. Bollen and J. S. Long. Newbury Park, CA: Sage.

Hammervold, Randi. 1998. "Properties of Goodness of Fit Statistics for Structural Equation Models." Doctoral dissertation, NTNU Trondheim Norgest teknisknaturvitenskapelige Universitet.

Hu, Lit-ze and Peter M. Bentler. 1998. "Fit Indices in Covariance Structure Modeling: Sensitivity to Underparameterized Model Misspecification." *Psychological Methods* 3:424-53.

Jöreskog, Karl G. 1993. "Testing Structural Equation Models." Pp. 294-316 in *Testing Structural Equation Models*, edited by K. A. Bollen and J. S. Long. Newbury Park, CA: Sage.

Kaplan, David. 1990. "Evaluating and Modifying Covariance Structure Models: A Review and Recommendation." *Multivariate Behavioral Research* 25:285-305.

MacCallum, Robert C. 1990. "The Need for Alternative Measures of Fit in Covariance Structure Modeling." *Multivariate Behavioral Research* 25:157-62.

MacCallum, Robert C., Michael W. Browne, and H. M. Sugawara. 1996. "Power Analysis and Determination of Sample Size for Covariance Structure Modeling." *Psychological Methods* 1:130-49.

Mariano, Robert S. 1982. "Analytical Small-Sample Distribution Theory in Econometrics: The Simultaneous-Equations Case." *International Economic Review* 23:503-33.

McDonald, Roderick P. 1989. "An Index of Goodness of Fit Based on Noncentrality." *Journal of Classification* 6:97-103.

McDonald, Roderick P. and Herbert W. Marsh. 1990. "Choosing a Multivariate Model: Noncentrality and Goodness of Fit." *Psychological Bulletin* 107:247-55.

Muthén, Bengt O. and David Kaplan. 1985. "A Comparison of Some Methodologies for the Factor Analysis of Non-Normal Likert Variables." *British Journal of Mathematical and Statistical Psychology* 38:171-89.

———. 1992. "A Comparison of Some Methodologies for the Factor Analysis of Non-Normal Likert Variables: A Note on the Size of the Model." *British Journal of Mathematical and Statistical Psychology* 45:19-30.

Paxton, Pamela, Patrick J. Curran, Kenneth A. Bollen, James Kirby, and Feinian Chen. 2001. "Monte Carlo Experiments: Design and Implementation." *Structural Equation Modeling* 8:287-312.

Rigdon, Edward E. 1998. "The Equal Correlation Baseline Model for Comparative Fit Assessment in Structural Equation Modeling." *Structural Equation Modeling* 5:63-77.

SAS. 2000. *SAS/STAT User's Guide: Version 8*. Cary, NC: SAS Institute, Inc.

Satorra, Albert and Willem Saris. 1985. "The Power of the Likelihood Ratio Test in Covariance Structure Analysis." *Psychomotrika* 50:83-90.

Steiger, James H. 1989. *EzPath: A Supplementary Module for SYSTAT and SYGRAPH*. Evanston, IL: SYSTAT.

———. 2000. "Point Estimation, Hypothesis Testing, and Interval Estimation Using the RMSEA: Some Comments and a Reply to Hayduk and Glaser." *Structural Equation Modeling* 7:149-62.

Steiger, James H. and John C. Lind. 1980. "Statistically Based Tests for the Number of Common Factors." Paper presented at the annual meeting of the Psychometric Society, May, Iowa City, IA.

Steiger, James H., Alan Shapiro, and Michael W. Browne. 1985. "On the Multivariate Asymptotic Distribution of Sequential Chi-Square Statistics." *Psychometrika* 50:253-63.

Sugawara, H. M. and Robert C. MacCallum. 1993. "Effect of Estimation Method on Incremental Fit Indexes for Covariance Structure Models." *Applied Psychological Measurement* 17:365-77.

Tanaka, Jeff S. 1993. "Multifaceted Conceptions of Fit in Structural Equation Models." In *Testing Structural Equation Models*, edited K. A. Bollen and J. S. Long. Newbury Park, CA: Sage.

Tucker, Ledyard R. and C. Lewis. 1973. "A Reliability Coefficient for Maximum Likelihood Factor Analysis." *Psychometrika* 38:1-10.

Patrick J. Curran is an associate professor in the L. L. Thurstone Quantitative Laboratory in the Department of Psychology at the University of North Carolina at Chapel Hill. His primary area of research focuses on the development of new quantitative methods for longitudinal data analysis and the application of these methods to study drug and alcohol use in children and adolescents. He has recently published papers in Psychological Methods, Multivariate Behavioral Research, Structural Equation Modeling, Journal of Abnormal Psychology, *and* Development and Psychopathology.

Kenneth A. Bollen is director of the Odum Institute for Research in Social Science, Immerwahr Distinguished Professor of Sociology, and a fellow of the Carolina Population Center at the University of North Carolina at Chapel Hill. He is the Year 2000 recipient of the Lazarsfeld Award for Methodological Contributions in Sociology. His primary area of statistical research is structural equation modeling (SEM). He is the author of Structural Equation Models With Latent Variables *and coeditor (with J. S. Long) of* Testing Structural Equation Models. *Recent articles have been published in* Sociological Methodology, Sociology of Education, Studies in Comparative International Development, Annual Review of Sociology, *and* Annual Review of Psychology.

Feinian Chen is an assistant professor of sociology at Texas A&M University. Her research interests include family, demography, and quantitative methods. She has recently published papers in European Journal of Population *and* Population Development and Review. *She is currently investigating changes in family structures and familial relationship in China, Russia, and Vietnam.*

Pamela Paxton is an assistant professor of sociology at The Ohio State University. Her research interests include prosocial behavior, political sociology, and quantitative methods. Some of her recent publications have appeared in the American Journal of Sociology *and the* American Sociological Review.

James B. Kirby is a service fellow at the Agency for Healthcare Research and Quality. His research interests focus on family change, neighborhood characteristics, and health. He has recently published articles in the Journal of Health and Social Behavior *and* Medical Care.