STRUCTURAL EQUATION MODELING, 7(4), 557–595

The Performance of ML, GLS, and WLS Estimation in Structural Equation Modeling Under Conditions of Misspecification and Nonnormality

Ulf Henning Olsson
Department of Economics
Norwegian School of Management

Tron Foss
Department of Economics
Norwegian School of Management

Sigurd V. Troye
Institute of Strategy and Management
Norwegian School of Economics and Business Administration

Roy D. Howell
College of Business Administration
Texas Tech University

This simulation study demonstrates how the choice of estimation method affects indexes of fit and parameter bias for different sample sizes when nested models vary in terms of specification error and the data demonstrate different levels of kurtosis. Using a fully crossed design, data were generated for 11 conditions of peakedness, 3 conditions of misspecification, and 5 different sample sizes. Three estimation methods (maximum likelihood [ML], generalized least squares [GLS], and weighted least squares [WLS]) were compared in terms of overall fit and the discrepancy between estimated parameter values and the true parameter values used to generate the data. Consistent with earlier findings, the results show that ML compared to GLS under conditions of misspecification provides more realistic indexes of overall fit and less biased parameter values for paths that overlap with the true model. However, despite

Requests for reprints should be sent to Ulf H. Olsson, Department of Economics, Norwegian School of Management BI, Box 580, N–1301 Sandvika, Norway. E-mail: ulf.h.olsson@bi.no

recommendations found in the literature that WLS should be used when data are not normally distributed, we find that WLS under no conditions was preferable to the 2 other estimation procedures in terms of parameter bias and fit. In fact, only for large sample sizes (N = 1,000 and 2,000) and mildly misspecified models did WLS provide estimates and fit indexes close to the ones obtained for ML and GLS. For wrongly specified models WLS tended to give unreliable estimates and over-optimistic values of fit.

The degree to which models are correctly specified and data are multivariate normal are two important issues in structural equation modeling (SEM). When the hypothesized model is correctly specified and observed variables are multivariate normal, it can be analytically derived that different estimation procedures such as maximum likelihood (ML), generalized least squares (GLS), and weighted least squares (WLS) will produce estimates that converge to the same optimum and have similar asymptotic properties (Browne, 1974, 1984). Under ideal conditions the choice between methods is thus arbitrary. Under the more realistic assumption of misspecified models and data that are not multivariate normally distributed, the different procedures may not converge to the same optimum. The question can then be raised: Which estimation procedure, if any, is preferable when ideal conditions are not met? To answer this question we address alternative criteria that can be used for judging the adequacy of structural equation models and in a simulation study apply these criteria to solutions provided by different estimation procedures.

Olsson, Troye, and Howell (1999) suggested a distinction between measures of "theoretical fit" and "empirical fit." Theoretical fit refers to the degree of isomorphism between *structure* and *parameter values* of a theoretical model and those of the "true" model that generates the data. Empirical fit refers to commonly used indexes of fit that reflect the discrepancy between observed covariance structure and the one implied by a theoretical model. Theoretical fit is the isomorphism between the theoretic model (M_{theory}) and its parameters (θ) on one hand and the true model (M_{true}) and actual parameters (θ_{true}) on the other. Empirical fit is the correspondence between the observed covariance matrix (S) and the one implied by the estimated model ($\Sigma(\hat{\theta})$). The correspondence between the population covariance matrix (Σ) and the covariance structure implied by the theoretical model ($\Sigma(\theta)$) is the "true" empirical fit. The true empirical fit is a population value (e.g., error of approximation [EA]). The distinction between empirical and theoretical fit is explicated in Figure 1. The figure reflects the idea that the purpose of research is to bridge the gap between three different domains: *Reality* refers to some unknown generative mechanism (M_{true}) that generates population states and variance–covariance structures (Σ). The *empirical domain* consists of the data and corresponding sample variance–covariance structures (S). Finally, the *theoretical domain* consists of the theoretical model (M_{theory}) of the unknown generative mechanism. In substantive research, *theoretical fit,* or the correspondence between

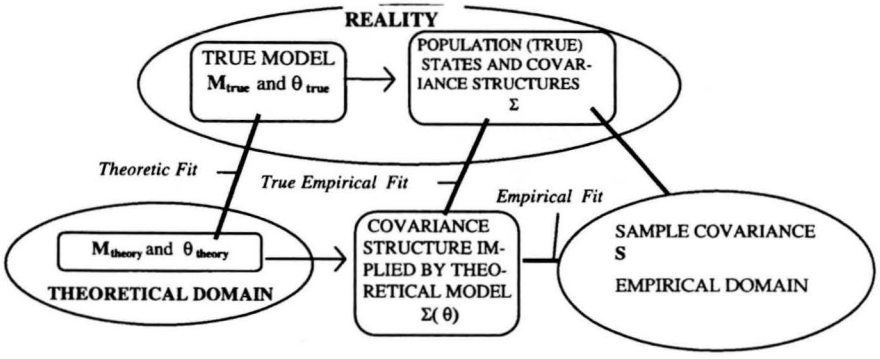

FIGURE 1 Three different domains of fit.

M_{true} and M_{theory}, can never be proved, but in SEM it is typically assessed from the degree of *empirical fit*, that is, the congruence between the observed variance–covariance (S) matrix and the one implied by the theoretical model $\Sigma(\theta)$. To the extent data are properly sampled from the population and the variables are properly measured, empirical measures of fit (such as chi-square, root mean squared error of approximation [RMSEA], etc.) will be reasonable estimates of *true empirical fit*. Unfortunately, empirical fit does not guarantee theoretical fit as demonstrated by Olsson et al. (1999). In fact, a method that performs well in terms of empirical fit may have a poor theoretical fit.

Although theoretical fit is the objective of research projects, the true model is not known in substantive research and there is no direct evidence for judging the correspondence of the theoretical model to the true model. In fact, our ignorance of the true model in substantive research is the rationale for empirical research. Simulated data can be used to assess how models that depart from a known "true" model affect the performance of various estimation methods when data vary in nonnormality.

Table 1 summarizes previous studies that have addressed issues relevant to those in this study. Cell 1 represents the case when both ideal criteria are satisfied. As pointed out earlier, Browne (1974, 1984) showed that for this case different estimation methods will produce converging results both in terms of theoretical and empirical fit. For correctly specified models Finch, West, and MacKinnon (1997) found negligible effects of nonnormality on parameter estimates for both ML and ADF (Asymptotic Distribution of Fit). This was also found in earlier studies by Bollen (1989), Boomsma (1983), Browne (1987), and Jöreskog and Sörbom (1988). Cell 2 represents studies that have addressed violations of the normality criterion. Chou, Bentler, and Satorra (1991), among others, concluded that the performance of GLS and ML with respect to empirical fit was reasonably robust to moderate deviations from multivariate normality. But Bentler (1989), Bollen

TABLE 1
Overview of Studies Addressing the Effects of Misspecification and Nonnormal Distributions

	Multivariate Normal Distribution	
Specification Error	*Yes*	*No*
No		
Empirical fit	1a, B1	2a, CBS, YB, Be,Bo,J&S
Theoretical fit	1b, B1	2b,YB, FWM, Bo, Bm, J&S, B2,*
Yes		
Empirical fit	3a, OTH	4a, CWF
Theoretical fit	3b, OTH	4b, *

Note. B1 = Browne, 1974; B2 = Browne, 1987; Be = Bentler, 1989; Bo = Bollen, 1989; Bm = Boomsma, 1983; J&S = Jöreskog and Sörbom, 1988; FWM = Finch, West, and MacKinnon, 1997; OTH = Olsson, Troye, and Howell, 1999; CWF = Curran, West, and Finch, 1996; CBS = Chow, Bentler, and Satorra, 1991; YB = Yuan and Bentler, 1997; * = this study.

(1989), and Jöreskog and Sörbom (1988) reported that model fit and significance tests may be affected by deviations from normality. The study by Olsson et al. (1999) fits in the third cell. They investigated the performance of GLS and ML both in terms of empirical fit and theoretical fit under varying conditions of misspecification. Whereas it has been observed in some studies (Browne, 1974; Ding, Velicer, & Harlow 1995; Olsson et al., 1999) that GLS tends to produce better empirical fit than ML, the Olsson et al. study showed that the superiority in terms of empirical fit for GLS appears to come at the cost of lower theoretical fit when models contain specification error. Parameter estimates for correctly specified paths within a partly misspecified model were found to be significantly more biased for GLS than for ML. Curran, West, and Finch (1996) examined the impact on empirical fit of GLS and ADF (or WLS) with both nonnormality and specification error (Cell 4). Yuan and Bentler (1997) found that ML is much less biased than ADF estimators for all distributions and sample sizes, but when the sample size is large ADF can compete with ML (Cell 2).

This study extends knowledge in two directions: (a) It extends the issues investigated in Olsson et al. (1999) by adding the effects of nonnormality (see Cell 2; because we relax the assumption of nonnormality it becomes natural to also include asymptotically distribution free estimators or weighted least squares in LISREL terminology) and (b) it extends the Curran et al. (1996) study by investigating performance in terms of theoretical fit.

The rationale for addressing both the nonnormality and the misspecification issue in the same study is that there might be a tradeoff between the estimation methods' ability to handle the two types of violations. The apparent benefits obtained with ML in terms of parameter estimates for partly misspecified models may not

hold when the data are not multivariate normal. On the other hand, the assumed superiority of ADF (WLS) when data are highly peaked or skewed may not hold when models are misspecified.

In the next section we show how ML, GLS, and WLS can be expected to produce different results when the model does not hold in the population and/or the data are not normally distributed. Hence, discrepant results provide an indication that a model is not correctly specified. We then provide the rationale for some criteria that can be used to choose between methods when discrepant results are obtained. A simulation study is then presented for a particular class of misspecified models.

WHY ESTIMATION METHODS MAY PRODUCE DISCREPANT EMPIRICAL FIT AND PARAMETER ESTIMATES

This section illustrates that when models are misspecified, but the data are multivariate normal, the fit function and parameter estimates for ML should be expected to differ from those provided by GLS and WLS. On the other hand, when models are not misspecified (or only mildly misspecified), but data are not normally distributed, the WLS solutions should differ from those obtained with GLS and ML.

In SEM the covariance structure is fitted to the sample covariance matrix S by minimizing a discrepancy function $F(S,\Sigma(\theta))$ with respect to the parameter vector θ. The fit function for ML can be expressed as

$$F_{ML}(\theta) = \log|\Sigma(\theta)| + \mathrm{tr}\{S\Sigma(\theta)^{-1}\} - \log|S| - q,$$

q being the number of observed variables. For GLS as $F_{GLS}(\theta) = \frac{1}{2}\mathrm{tr}\{(S-\Sigma(\theta))S^{-1}\}^2$ or equivalently as $F_{GLS}(\theta) = (s-\sigma_\theta)' U_{GLS}^{-1} (s-\sigma_\theta)$, and for WLS as $F_{WLS}(\theta) = (s-\sigma_\theta)' U_{WLS}^{-1} (s-\sigma_\theta)$.

The reason why the various fit functions do not converge when models are misspecified or data are not multivariate normal is that the nature of the weight element differs. In the following we first show why WLS will provide solutions that differ from those obtained with GLS and ML when distributions are nonnormal. We then explain why ML on one side and GLS and WLS on the other do not provide equivalent results when models are misspecified.

The estimation theory of ML and GLS is based on the assumption of multivariate normality. Browne (1984) generalized this theory to any multivariate distribution for continuous variables satisfying very mild assumptions.

The WLS estimator uses a weight matrix with a typical element that is a combination of estimates of second- and fourth-order moments:

$$[U_{WLS}]_{ij,gh} = s_{ijgh} - s_{ij}s_{gh} \quad i \geq j \quad g \geq h$$

where

$$s_{ijgh} = \frac{\sum (x_i - \bar{x}_i)(x_j - \bar{x}_j)(x_g - \bar{x}_g)(x_h - \bar{x}_h)}{N}$$

is an estimate of

$$\sigma_{ijgh} = E\{(x_i - Ex_i)(x_j - Ex_j)(x_g - Ex_g)(x_h - Ex_h)\}$$

E being the expectation symbol, and s_{ij} is an estimate of the covariance σ_{ij}.

The matrix U_{WLS} is of order $u \times u$ where $u = q(q+1)/2$. For the GLS fit function, however, these elements only consist of second-order moments:

$$[U_{GLS}]_{ij,gh} = s_{ig}\, s_{jh} + s_{ih}\, s_{jg} \quad i \geq j \quad g \geq h$$

The matrix U_{GLS} is of order $u \times u$ where $u = q(q+1)/2$. When the model is correct under normality the elements of U_{WLS} and U_{GLS} will converge in probability to the population values $\sigma_{ig}\,\sigma_{jh} + \sigma_{ih}\,\sigma_{jg}$ although their variability is different (Hu, Bentler, & Kano 1992). For the ML estimator $\log|\Sigma(\theta)| + \mathrm{tr}\{S\Sigma(\theta)^{-1}\} - \log|S| - q$ Browne (1974) showed that if $\hat{\theta}_1$ is an ML estimate for θ_{true} the GLS estimate $\hat{\theta}_2$ for $V = \Sigma^{-1}(\hat{\theta}_1)$ in $F_{GLS}(\Sigma(\theta)|V) = \frac{1}{2}\mathrm{tr}\{(S - \Sigma(\theta))V\}^2$ will be a Best Generalized Least Squares (B.G.L.S.) estimator and $\lim_{N\to\infty} P(|\hat{\theta}_1 - \hat{\theta}_2| < \delta) = 1$, for any $\delta > 0$. This implies that the ML estimator can be seen as a GLS estimator, and consequently it will have the same asymptotic properties given that the model is correct.

The reason why some findings indicate that GLS provides better empirical fit than ML (Browne, 1974; Ding et al., 1995; Olsson et al., 1999) is that the estimators are no longer equivalent when models are misspecified.

The weight matrix for ML $(U_{ML}^{-1} = K_p^{-}(\Sigma(\hat{\theta}_1)^{-1} \otimes \Sigma(\hat{\theta}_1)^{-1})K_p^{-1}$, where K_p and K_p^- are defined in Browne, 1974) is a function of the model, but for GLS and WLS the elements in the weight matrix are only functions of second- and fourth-order moments of the observed variables. The elements of U_{GLS}^{-1} for GLS do not depend on the model, only on S, and because S will converge to Σ even if the model is poor, the result can be quite different from the ML situation in which $\Sigma(\hat{\theta}_1)$ (where $\hat{\theta}_1$ is an ML – estimate), approximates Σ only when the model fits very well or exactly. For the WLS method the elements of U_{GLS}^{-1} also depend on S

but include in addition the fourth-order moments that are estimated from the sample. Without any specification error we know that both S and $\Sigma(\hat{\theta}_1)$ will converge to Σ. In this situation the differences between the ML and GLS are due only to sampling error. For small deviations from normality and little or no specification error, the weight matrix for WLS should not differ systematically from the weight matrix of GLS or ML.

Consequently, if $S \approx \Sigma(\hat{\theta})$ then $s_{ij} \approx \hat{\sigma}_{ij}$ implying that min(ML) ≈ min(GLS) ≈ 0. Whether min(WLS) will be close to min(ML) and min(GLS) will depend on the distributional properties. For negligible kurtosis it is reasonable that the three are close. WLS is in the literature recommended when data are nonnormal. The assumed superiority of WLS when data are peaked or skewed may not hold when models are misspecified because WLS (a quadratic form estimator) belongs to the same family of estimators as GLS and therefore may have the same problems in terms of parameter bias (Olsson et al., 1999).

In summary, the reason why ML provides results that differ from those obtained with GLS and WLS is that misspecification will cause the model-based estimated covariances in the weight matrix for ML to be different from the observed covariances entered in the weight matrices of GLS and WLS. The discrepancy between WLS and the two other estimators when data are nonnormal is that only the weight matrix of the former contains elements (the fourth moments) that reflect this violation. As shown in Table 2, when the models are incorrectly specified and the data are not multivariate normal (Cell 4), the methods should give different results. With multivariate normal data but a misspecified model, WLS and GLS will be equivalent (Cell 2).

Theoretical Fit Versus Empirical Fit

A desirable characteristic of an estimation method is that it provide realistic indications of the theoretical model's correspondence with the true model and provide nonbiased estimates of parameters of paths that overlap with the true model. In most realistic situations the researcher will not know the structure of the true model (M_{true}) or the parameter values (θ_{true}). The researcher therefore has to resort to indi-

TABLE 2
When Are ML, GLS, and WLS Equivalent?

Model/Distribution	*Normal*	*Nonnormal*
Correct model	ML ⇔ GLS ⇔ WLS asymptotically	ML ⇔ GLS asymptotically, but $(N-1)F$ is not χ^2 distributed
Misspecified model	GLS ⇔ WLS asymptotically	No equivalence

cations of empirical fit, that is the model's ability to recapture the covariances, as a basis for judging theoretical fit. The typical procedure is to base the acceptance of a theoretical model—if not a proof of its verisimilitude or truth—on the degree of correspondence between a structure implied by the model and the structure in the actual data (as measured by fit indexes) in addition to the statistical significance of the parameter estimates. The better the empirical fit, and the more statistically significant the parameter estimates, the more faith one has in the theoretical model. Moreover, modification indexes in combination with theoretical considerations provide the basis for improvements of the original model.

This mode of theory testing appears to be justifiable as long as it can be safely assumed that theoretical fit and empirical fit are perfectly related. If the goal of attaining empirical fit is attained at the cost of lower theoretical fit, the procedure is problematic and may lead to TYPE- II errors. Several factors preclude good empirical fit from guaranteeing that the structure of M_{true} has been identified and the parameter estimates are accurate. For example, because models that are theoretically distinct may produce the same variance–covariance matrix (i.e., may be empirically equivalent; see Luijben, 1991; MacCallum, Wegener, Uchino, & Gabrigar, 1993), empirical support for one theoretic model does not preclude the possibility that the data are generated by another model. Also, in a simulation study of model respecification based on modification indexes, MacCallum (1986) showed that under certain circumstances data can appear to be well-accounted for by a wrongly respecified model when a substantially different model was used to generate the data. Therefore, estimation procedures that appear to successfully fit models to the data may do so both at the expense of identifying the wrong model and providing incorrect parameter estimates. Olsson et al. (1999) further demonstrated that the correspondence between theoretical fit and empirical fit cannot be taken for granted. Their simulation study showed that the empirical fit of misspecified models (in the sense of being parsimonious representations of the true model) can easily be overestimated. ML and GLS were found to differ with respect to theoretical and empirical fit. An analysis of variance (ANOVA) of theoretical fit (parameter bias) and empirical fit (RMSEA) as a function of sample size, degree of misspecification, and choice of estimator showed a very strong method-by-misspecification effect for both types of fit. The apparent superiority of GLS relative to ML in terms of empirical fit was achieved at the expense of higher parameter distortion, that is, lower theoretic fit. Moreover, the potential benefits of modification indexes as a criterion for correctly modifying the tested model to include omitted paths could not be taken for granted. The rather optimistic estimates of empirical fit obtained with GLS lead apparently to complacency and no apparent need for improvements. The results of the Olsson et al. (1999) study have several implications for a number of decisions (e.g., choice of estimation procedure, acceptance/rejection of single models, choice between competing models, and modification of exist-

ing models) and suggest that if the pursuit of empirical fit leads to the choice of GLS, the cost can lower theoretical fit with respect to both model structure and parameter bias.

Parsimony As Specification Error

A theoretical model will seldom perfectly represent real world phenomena (MacCallum & Tucker, 1991; Meehl, 1990). Lack of theoretical fit can occur for a number of reasons: inadequate operational procedures and estimation methods, distributional properties of data and sampling error, in addition to incorrectly specified models (Olsson et al., 1999). Often, less than perfect theoretical fit results from theories being parsimonious representations of reality (Browne & Cudeck, 1992; MacCallum & Tucker, 1991; Meehl, 1990). Parsimony can be due to ignorance of all the real-world complexities that result in the pattern of population variances and covariances, or can reflect a conscious choice of models that are restricted to only a part of the dimensions and causal structures in the domain. Certain factors are ignored or omitted, whereas factors that are known and considered important for theoretic or pragmatic reasons are included. The higher the parsimony relative to the true model, the lower the empirical fit of the parsimonious model (Mulaik et al., 1989) unless the parsimony is specifically taken into account as is the case for several fit indexes (e.g., RMSEA, Browne & Cudeck, 1992; PNFI, Mulaik et al., 1989; AIC, Akaike, 1974; SAIC, Schwartz, 1978).

Although all of these considerations threaten theoretical fit as well as empirical fit, this simulation study primarily focuses on parsimony as a source of error, that is, the omission of factors and relations that impinge on the phenomena of interest. When theoretical models under-represent the complexity of the true model, it is well known that this will impair empirical fit in the sense that the theoretic model cannot account for the portion of the variance in the data that is due to the omitted variables. Thus parsimony is attained at the cost of lower explanatory power. However, the researcher can still hope for theoretical fit in the sense that the parameters associated with the variables included in the simplified model are approximately correct. Thus, although empirical fit may not be perfect, and the theoretical model is known to be a simplified and therefore less-than-accurate reflection of the true model (M_{true}), the researcher should still be interested in parameter estimates ($\hat{\theta}$) that are close to the true parameters (θ_{true}) for the part of the model that coincides with the true model. To the extent that empirical fit does not imply theoretic fit, one should prefer estimation methods that provide more accurate parameter estimates even in the presence of simplified and less-than-perfect theoretical models, and estimation methods whose degree of empirical fit provides a closer indication of theoretical fit.

Nonnormality

In contrast to misspecification, the other departure from the ideal condition, nonnormality can be assessed and screening (continuous) variables for normality prior to conducting the testing is an important step in every multivariate analysis. Normality of variables is assessed by either statistical or graphical methods. Two important aspects of nonnormality are skewness and kurtosis. Skewness, defined to be $\alpha_3 = E(x-\mu)^3/\sigma^3$ can be estimated by $\frac{\sum (x-\bar{x})^3 / N}{(\sum (x-\bar{x})^2 / N)^{3/2}}$, and has to do with the symmetry of the distribution. Kurtosis, defined as $\alpha_4 = E(x-\mu)^4/\sigma^4$, which can be estimated by $\frac{\sum (x-\bar{x})^4 / N}{(\sum (x-\bar{x})^2 / N)^2}$ has to do with the peakedness of the distribution. With a normal distribution skewness is 0 and kurtosis is 3 (although, for the reference normal distribution to have a kurtosis of zero, 3 is subtracted from α_4). Consequences of nonzero skewness and kurtosis have been widely studied with results indicating that tests of mean differences appear more affected by skewness than kurtosis, whereas tests for variances and covariances are more affected by kurtosis than skewness (Mardia, Kent, & Bibby, 1979). Browne (1982, 1984) and Finch et al. (1997) noted that kurtosis can have considerable impact on significance tests and standard errors of parameter estimates. Univariate normality is a necessary but not sufficient condition for multivariate normality. As a consequence, multivariate nonnormal data can be ascertained by creating nonnormal univariate data. This study concentrates on departures from nonnormality in terms of peakedness or kurtosis. The combined impact of misspecification and peakedness on theoretical fit has not been addressed previously in SEM.

Design Methodology

ML, GLS, and WLS were used to estimate four models varying from none to severe misspecification. Simulated data sets had 11 different levels of kurtosis ($\alpha_4 - 3$) in the latent variables ranging from mild (–1.2) to severe peakedness (+25.45), keeping the distribution symmetric. The sample sizes were N= 100, 250, 500, 1,000, and 2,000.

A total of four models were designed: a "true model" and three misspecified models nested within the true model. The "true model" (M_{true}) used to generate the data contained three major and four minor factors and 12 observed variables (Figure 2). The M_{true} model used was the same as the one in Olsson et al. (1999). This model is also quite similar to the one used by MacCallum and Tucker (1991) in their simulation study. The first partially true model (M_1) contained all observed variables and three latent variables (the four minor factors were omitted). The model is partially true in the sense that all paths for the remaining factors and variables overlap with

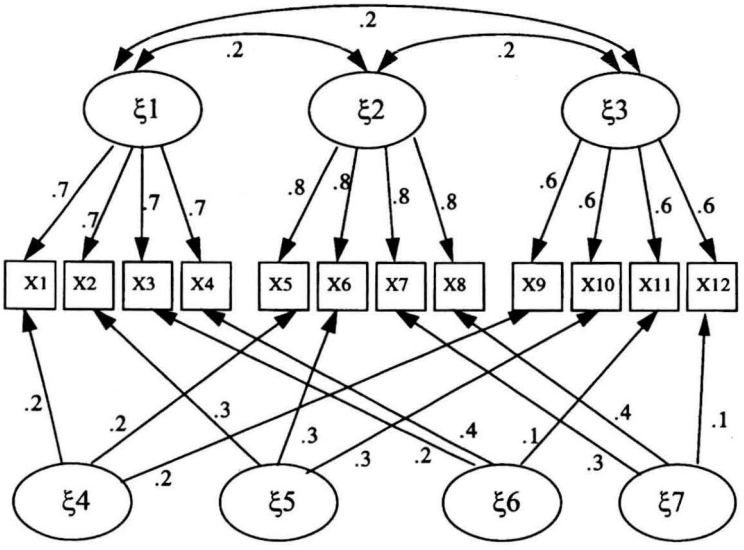

FIGURE 2 Population model (the error terms are not included in the figure).

those of the true model. The next partially true (M_2) but more parsimonious model is an orthogonal version of M_1 in which the correlations between the included factors were omitted. Finally, the most severely misspecified model (M_3) excluded paths between two factors and the respective observed variables.

The true model (M_{true}) is an oblique CFA mode, $X = \Lambda\xi + \delta$, with the structure shown graphically in Figure 2 and mathematically in Appendix B. The loadings from the domain factors were given the following values: $\lambda_{i1} = .7$ for $i = 1, 2, 3$, and 4; $\lambda_{i2} = .8$ for $i = 5, 6, 7$, and 8; and $\lambda_{i3} = .6$ for $i = 9, 10, 11$, and 12. The loadings from the minor factors (λ_{i4}) were set to .2 for $i = 1, 5$, and 9. The values of the other loadings were set as follows: $\lambda_{i5} = .3$ for $i = 2, 6$, and 10; $\lambda_{36} = .2$; $\lambda_{46} = .4$; $\lambda_{11,6} = .1$; $\lambda_{77} = .3$; $\lambda_{87} = .4$; and $\lambda_{12,7} = .1$. For the variance–covariance matrix of the latent variables (ϕ matrix) the variances were fixed to one ($\phi_{ii} = 1.0$) and the correlations between the three major factors were set to .2 ($\phi_{21} = \phi_{31} = \phi_{32} = .2$). The covariance matrix Θ_δ is diagonal and var(δ_i) is a function of the model and the fact that var(x_i) = 1 for $i = 1, 2, \ldots, 12$.

The misspecified models were three nested simplifications of the true model. For M_1 we excluded the minor factors (i.e., all λ_{ij}s for $j > 3$ were set to zero). The rest remained unchanged. The next model was an orthogonal version of M_1 (i.e., the ϕ_{ij} $i \neq$ j were set to zero). The factor pattern was unchanged. Because the true model is of the form, $x_i = \sum_{j=1}^{3} \lambda_{ij}\xi_j + \sum_{l=4}^{7} \lambda_{il}\xi_l + \delta_i$ for $i = 1, 2, \ldots, 12$ and the correlation between ξ_j and ξ_l for $j = 1, 2, 3$ and $l = 4, 5, 6, 7$ are all zero, no biased

parameter estimates were expected for these two models (M_1 and M_2) because bias for an included parameter is a function of the correlation between the included and the excluded variables. The third model was an even more parsimonious model in which λ_{52} and $\lambda_{12,3}$ were also fixed to zero.

Measures of Theoretic Fit

As indicated by Olsson et al. (1999), the various indexes should ideally accomplish the following: (a) give a numerical value of each model's degree of misspecification, (b) identify models that are truly acceptable, (c) give a comparative basis for choosing and ranking models according to their empirical and theoretic fit, and (d) give valid indications for how to modify misspecified models by adding or deleting paths between included variables. Because the true parameter values used to generate the data sets are known, the discrepancy between the true values and the estimated parameter values can be calculated. The Mean Squared Parameter Bias (SPB) was defined as a measure of parameter distortion and given by

$$SPB = \frac{1}{r}\sum_{i=1}^{r}\left(\frac{1}{a}\sum_{j=1}^{a}(\pi_{est_{ij}} - \pi_j^*)^2\right)$$

where π_j^* = true parameter value, π_{estij} = estimated parameter value, a = number of free parameters to be estimated, and r = number of replications where the estimation converged.

This index gives a basis for selecting among solutions in terms of theoretical fit. The index is of course of reduced utility for comparing models that are not of the same metric in the parameters.

Indexes of Misspecification

To get a misspecification index that is independent of estimation and sampling procedure, Olsson et al. (1999) suggested the fixed root mean squared residual (FRMR), defined as

$$\text{FRMR} = \left[2\sum_{i=1}^{q}\sum_{j=1}^{i}(\sigma_{ij} - \sigma_{Fij})^2 \,/\, q(q+1)\right]^{0.5}$$

where $\sigma_F \in \Sigma(\theta^*)$.

In contrast to the s_{ij}' elements in the Root Mean Squared Residual index that are based on the sample variance–covariance matrix, S, and the estimated parameters $\hat{\Theta}$, the σ_F elements in the FRMR formula are based on θ^*, the subset of θ_{true} values

for the subset of the theoretical model (M_{theory}) that overlaps with the true model (M_{true}), and the corresponding population variance–covariance matrix, S. The FRMR value will thus be zero for theoretic models that are correctly specified, that is, for models that overlap completely with the true model. The FRMR index can serve as a measure of misspecification when the true model is known. Models ranked according to the FRMR index can therefore be conveniently compared to the ranking based on empirical fit indexes to be discussed next. Although this index cannot be used when the true model is not known, it answers the call by Gerbing and Anderson (1993) for a method to characterize model lack of fit that is independent of the estimation method used. Ideally, empirical fit indexes should give the same ranking as the FRMR index if the goal is to identify the right models and the associated parameter values.

Measures of Empirical Fit

Steiger's (1990) RMSEA was employed as an indicator of empirical fit (Browne & Cudeck, 1992). It is based on the conceptualization that error of empirical fit can be partitioned into two types: EA and Error of Estimation (EE):

$$E(\text{Total Error}) \approx \text{EA} + E(\text{EE})$$

$$E(\text{F}(\Sigma, \Sigma(\hat{\theta}))) \approx \text{F}(\Sigma, \Sigma(\theta_0)) + E(\text{F}(\Sigma(\theta_0), \Sigma(\hat{\theta}))$$

$$\approx F_0 + (N-1)^{-1} t;$$

where $\Sigma(\theta_0)$ is the matrix we get when the model is fitted to the population, $\Sigma(\hat{\theta})$is the matrix we get when the model is fitted to the sample, t is the number of parameters to be estimated, and E is the expectation operator.

Because F_0 generally decreases when parameters are added, EA can be adjusted by the degrees of freedom to meet the desire for parsimony. The resulting RMSEA thus measures EA as discrepancy per degree of freedom (Browne & Cudeck, 1992). Using EA_{df} as the notation for EA per degrees of freedom, $EA_{df} = \sqrt{\frac{F_0}{df}}$. EA_{df} is an unknown population value "measuring the True Empirical Fit" (see Figure 1). The RMSEA index

$$(\text{i.e., } RMSEA = \sqrt{Max(\frac{\hat{F}}{df} - \frac{1}{N-1}; 0)} = \sqrt{Max(\frac{\chi^2}{(N-1)df} - \frac{1}{N-1}; 0)})$$

is an estimate of this quantity.

Data Generating Process

Simulations were used to study the effects of structural error and distributional error on empirical and theoretical fit using the true model presented earlier. The simulations were performed with C++, PRELIS 2, and LISREL 8 (also used for estimation; Jöreskog & Sörbom, 1993). The distributions of the population data were divided into three different categories of kurtosis holding the skewness constant equal to zero: For the distributions with negligible kurtosis the alpha values (α_4) were set at 1.8, 2.0, and 2.4. Distributions of medium kurtosis were generated with the following values: $\alpha_4 = 5, 6, 8, 8.5$, and 9.35 while the three most highly peaked distributions were generated based on the following values: $\alpha_4 = 18.14, 22.21$, and 28.45. Because kurtosis was defined with respect to the latent variables, a procedure based on Mattson (1997) and Ramberg, Tadikamalla, Dudewicz, and Mykytka (1979) was developed to produce observed variables with the desired distribution properties. Following we give just a brief overview of the procedure (further details are given in Appendix A).

The aim is to generate observations randomly from a population that may not be normally distributed. To do this we use the Lambda distribution and, assuming that p is taken randomly from a uniform distribution over the interval [0, 1], then the distribution for $R(p) = \lambda_1 + [p^{\lambda 3} - (1 - p)^{\lambda 4}] / \lambda_2$ will depend on the values we chose for λ_1, λ_2, λ_3, and λ_4. For this study we chose lambdas (λ_i) so that the expectation of $R(p)$ is zero ($\mu_1 = 0$), the variance of $R(p)$ is 1 ($\mu_2 = 1$), and the distribution of $R(p)$ is symmetric ($\mu_3 = 0$; see Ramberg et al., 1979). At the same time we can choose the levels of kurtosis (μ_4, the fourth-order moment is equal to α_4 because $\mu_2 = 1$) for the distribution of $R(p)$ (see appendix A for details).

The generating model is given by the equation: $X = \Lambda_x \xi + \delta$; where X is a $q \times 1$ matrix, Λ_x is a $q \times k$ matrix, ξ is a $k \times 1$ matrix of stochastic variables with the covariance matrix Φ of dimension $k \times k$, and δ is a $q \times 1$ matrix of uncorrelated error terms with the variance–covariance matrix θ_δ. The $q \times 1$ matrix X of "observed" variables was generated based on the following 5 steps:

1. By the Lambda distribution we draw a $k \times 1$ matrix ω_1 of uncorrelated values with expected kurtosis equal to μ_4.
2. We then calculate $\xi = P \omega_1$, where P is a $k \times k$ matrix so that $PP' = \Phi$. ξ then has the covariance matrix Φ.
3. A $q \times 1$ matrix ω_2 of uncorrelated values is drawn with expected kurtosis equal to μ_4.
4. We then calculate $\delta = D\omega_2$; where D is a $q \times q$ diagonal matrix so that $\text{Var}(\delta) = \theta_\delta$.
5. The observations are now easily calculated by the formula $X = \Lambda_x \xi + \delta$.

These observations have an expected covariance matrix Σ (the population covariance matrix), and the kurtosis depends on μ_4 (see Table A5). The observations are then used as input to PRELIS for the simulation process and then LISREL is used for estimation.

RESULTS

In the following we first demonstrate how ML, GLS, and WLS differ in terms of empirical fit as a function of misspecification, sample size, and kurtosis. Second, we show how estimates of empirical fit are biased and further demonstrate how this bias varies as a function of the same factors. Finally, we investigate how estimation method, misspecification, sample size, and kurtosis affect theoretical fit. The three methods were applied to 16,500 randomly selected replications each, for a total of 49,500 estimations.

Empirical Fit

Empirical fit was measured in terms of RMSEA, the degree of misspecification was captured by the FRMR index, and nonnormality was represented by the fourth-order moment (kurtosis), $\alpha 4$. Because the RMSEA values are method-specific (see Appendix A) and thus not quite comparable,[1] ANOVA was used to test the following model for each estimation method separately:

$$\text{RMSEA} = \mu + \text{FRMR} + N + \alpha_4 + \text{FRMR}*\alpha_4 + N*\alpha_4$$

where N = sample size, FRMR = index of specification error, and α_4 = the fourth-order moment.

Diagnostics suggest no substantial violation of the distributional assumption or the homogeneity of variance assumption of ANOVA. As can be seen in Tables 3A–3C, sample size (N) and misspecification (FRMR) all have a highly significant main effect for the three methods. This was quite expected. It is also desirable that measures of empirical fit reflect the actual level of misspecification. For ML and GLS the main effect of kurtosis, α_4, is slightly significant ($F = 5.23$ and 2.38), but not very strong ($\eta^2 < .001$ for both) whereas for WLS this effect is remarkably strong ($F = 52.66$ and $\eta^2 = .0448$). Another difference can be observed: The inter-

[1]We know that $\min F_{ML}(S,\Sigma(\theta)) \neq \min F_{GLS}(S,\Sigma(\theta)) \neq \min F_{WLS}(S,\Sigma(\theta))$ when the model is not correct. So empirical fit values (RMSEA, etc.) will be method-specific (i.e., not in the same metric) because the minimum of the fit function value enters into the empirical fit indexes.

TABLE 3A
Analysis of Variance of RMSEA for ML

Factor	*df*	*Adjusted SS*	*Sequential SS*	*F*	*p*	η^2
FRMR	2	0.294479	0.294479	66000	0.000	0.9978
N	4	0.000123	0.000123	13.75	0.000	0.0004
α4	10	0.000117	0.000117	5.23	0.000	0.0000
FRMR*α4	20	0.000042	0.000042	0.93	0.549	0.0000
*N*α4	40	0.000157	0.000157	1.76	0.150	0.0005
Error	88	0.000197	0.000197			
Total	164		0.295116			

Note. RMSEA = root mean squared error of approximation; ML = maximum likelihood; SS = sum of squares; FRMR = fixed root mean squared residual.

TABLE 3B
Analysis of Variance of RMSEA for GLS

Factor	*df*	*Adjusted SS*	*Sequential SS*	*F*	*p*	η^2
FRMR	2	0.106711	0.106711	35000	0.000	0.9375
N	4	0.006846	0.006846	1128.16	0.000	0.0601
α_4	10	0.000036	0.000036	2.38	0.015	0.0000
FRMR*α_4	20	0.000024	0.000024	0.80	0.712	0.0000
Nα_4	40	0.000074	0.000074	1.22	0.216	0.0006
Error	88	0.000134	0.000134			
Total	164		0.113825			

Note. RMSEA = root mean squared error of approximation; GLS = generalized least squares; SS = sum of squares; FRMR = fixed root mean squared residual.

TABLE 3C
Analysis of Variance of RMSEA for WLS

Factor	*df*	*Adjusted SS*	*Sequential SS*	*F*	*p*	η^2
FRMR	2	0.083715	0.083715	1928.52	0.000	0.3278
N	4	0.151353	0.151353	1743.35	0.000	0.5928
α_4	10	0.011429	0.011429	52.66	0.000	0.0448
FRMR*α_4	20	0.006554	0.006554	15.10	0.000	0.0257
Nα_4	40	0.000360	0.000360	0.42	0.998	0.0014
Error	88	0.001910	0.001910			
Total	164		0.255327			

Note. RMSEA = root mean squared error of approximation; WLS = weighted least squares; SS = sum of squares; FRMR = fixed root mean squared residual.

action effect of misspecification and kurtosis, FRMR*α_4, is only significant for WLS ($F = 15.10$ and $\eta^2 = .0257$). For ML and GLS this interaction effect is absent ($p = .55$ and .71). Because WLS is designed to adjust for violations of normality, the strong effect of α_4, kurtosis, for WLS should be expected. It is more surprising that the positive interaction effect of kurtosis and misspecification reflects that empirical fit actually increases (becomes better) with kurtosis. The interaction effect of sample size and kurtosis, $N^*\alpha_4$, is not significant at 5% level for either of the three methods.

Referring to the results presented in Tables B1–B7 (Appendix B; the mean values for RMSEA for each sample size based on 100 replications), the tables are organized so that the ML results are at the top, GLS results are in the middle, and WLS results are at the bottom in each cell. It is quite clear that RMSEA for WLS tends to drop relatively more (as a function of α_4) for severely misspecified models than for less misspecified models. As an example, for $N = 2{,}000$ (Table B5), RMSEA drops from .0454 to .0402 (11.2%) for M_1, from .0509 to .0421 for M_2 (17.2%), and from .0864 to .0669 (22.5%) for M_3.

Empirical Bias

Although the aforementioned analyses show the impact of sample size, kurtosis, and misspecification of empirical fit, we do not know the degree to which the estimates of empirical fit (RMSEA) are biased (i.e., depart from their population values). In this section we therefore address how the bias of RMSEA, $RMSEA_{bias}$, defined as $\overline{RMSEA} - EA_{df}$, where EA_{df} is the population EA, varies as a function of the same factors that affect empirical fit. The reason for addressing bias as a separate issue is that the two components of $RMSEA_{bias}$ are both method-specific. To analyze the bias, we thus need to calculate the EA_{df}, which is a population quantity that reflects the inadequacy of the model given the estimation method (Browne & Cudeck, 1992).

$EA_{df} = \sqrt{\frac{F_0}{df}}$; where $F_0 = \min F(\Sigma, \Sigma(\theta)) = F(\Sigma, \Sigma(\theta_0))$, thus varies across methods and must be calculated for each estimation method separately. $F(\Sigma, \Sigma(\theta))$ is often referred to as the population fit function.

For ML and GLS, EA_{df} is independent of the fourth-order moment and can easily be derived. The procedure for deriving EA_{df} for WLS is not so straightforward because it contains the fourth-order population moments and the products of second-order population moments. It therefore needs to be calculated for each distribution. In Appendix A the procedure for calculating EA_{df} values is shown. The EA_{df} values that are derived are shown in Tables A1–A4 and demonstrate the following. The EA_{df} values for GLS are lower than those for ML. WLS ap-

pears to reward the researcher for using highly peaked data: For a given model, the EA_{df} for WLS decreases with increasing values of α_4. As pointed out earlier, the decrease is substantially larger for severely misspecified models than for less misspecified models. When we compare the EA_{df} values (which are the true values, see tables A1–A4) with the estimated values ($\overline{RMSEA}$ in Tables B1–B7) for each estimation method, the following pattern emerges: For the 165 comparisons (11 levels of kurtosis, 5 sample sizes, and 3 levels of misspecifications), ML produces 98 RMSEA values higher than the population value and thus has a slight tendency to provide conservative estimates of empirical fit. This is not the case with GLS, which in 158 out of the 165 comparisons of EA_{df} and RMSEA produces overoptimistic estimates of empirical fit. There is, however, a tendency for GLS to produce more realistic RMSEA values for larger sample sizes. For WLS the pattern is more complicated. WLS is highly sensitive to sample size with respect to its ability to produce realistic RMSEA values. For $N = 100$ the bias varies between .07 and .13. For $N = 250$ the bias is in the interval from .013 to .046, for $N = 500$ the bias is in the interval from .0028 to .028, for $N = 1{,}000$ the bias is in the interval from .001 to .024, and finally, for $N = 2{,}000$ the bias is in the interval from .0005 to .018. The RMSEA value thus appears to approximate the population value at larger sample sizes. This supports the findings made by Hu et al. (1992), who reported that the ADF (WLS) method had a tendency to overestimate (i.e., give too good estimates) the true goodness of fit value for almost all "normal" sample sizes. However, the bias is in particular affected by kurtosis and specification error.

Theoretical Fit

ANOVA was used in an overall test of the effect of misspecification, sample size, kurtosis, and estimation method on theoretic fit, measured by the SPB index introduced earlier. The following model was tested:

$$\begin{aligned} SPB = \mu + FRMR + N + \alpha_4 + METHOD + FRMR*\alpha_4 + METHOD*\alpha_4 + \\ FRMR*METHOD + N*METHOD + N*\alpha_4 + FRMR*\alpha_4*METHOD \end{aligned}$$

where N = sample size, FRMR = index of specification error, α_4 = the fourth-order moment (kurtosis), and METHOD = estimation method.

The main effects of misspecification, sample size, and method are all significant. The main effect of kurtosis is also significant, which can be attributed to the relatively large sample. The interaction between misspecification and method (FRMR*METHOD) and the interaction between sample size and method (N*METHOD) are highly significant, whereas the interaction between kurtosis

TABLE 4A
Analysis of Variance of SPB

Factor	*df*	*Adjusted SS*	*Sequential SS*	*F*	*p*	η^2
FRMR	2	0.37775	0.37775	8530.11	0.000	0.43638
N	4	0.06231	0.06231	703.55	0.000	0.07200
α_4	10	0.01078	0.01078	48.68	0.000	0.01246
Method	2	0.14754	0.14754	331.62	0.000	0.17049
FRMR*α_4	20	0.00905	0.00905	20.44	0.000	0.01045
α_4*Method	20	0.01047	0.01047	23.65	0.000	0.01209
FRMR*Method	4	0.20169	0.20169	2277.23	0.000	0.23307
*N**Method	8	0.01978	0.01978	111.66	0.000	0.02286
*N**α_4	40	0.00162	0.00162	1.83	0.002	0.00187
FRMR*α_4*Method	40	0.01674	0.01674	18.90	0.000	0.01934
Error	344	0.02902	0.02902			
Total	494		0.86536			

Note. SPB = mean squared parameter bias; SS = sum of squares; FRMR = fixed root mean squared residual.

and method (α_4*METHOD), misspecification and kurtosis (FRMR*α_4), and the third-order interaction between misspecification, kurtosis and method (FRMR*α_4*METHOD) are moderately significant. The interaction effect of sample size and kurtosis (*N**α_4) is not significant. Analysis of mean differences shows that ML (*M* SPB = .00662, *SD* = .00664) is by far superior to GLS (*M* SPB = .03233, *SD* = .0366), which on its side outperforms WLS (*M* SPB = .04855, *SD* = .0547) with respect to theoretic fit.

The various interaction effects where METHOD is a component of the interaction term suggest that the impact of factors such as misspecification and kurtosis is contingent on which estimation method is chosen. To get a more detailed impression of the effects on the SPB value we therefore tested the following model for the three methods separately:

$$\text{SPB} = \mu + \text{FRMR} + N + \alpha_4 + \text{FRMR}*\alpha_4 + N*\alpha_4$$

where *N* = sample size, FRMR = index of specification error, and α_4 = the fourth-order moment.

In contrast to the results obtained for empirical fit, only sample size has a substantial main effect on the parameter precision for ML. The effects of misspecification and kurtosis are significant, but not strong ($F = 5.14$ and 3.52; $\eta^2 = .015$ and $.051$) compared to what we find for GLS and WLS. As for ML, sample size also affects the theoretical fit of GLS. However, consistent with the findings reported by Olsson et al. (1999) for normal data, theoretical fit for GLS is also highly influenced by misspecification ($F = 4152.12$ and $\eta^2 = .926$) and sample

TABLE 4B
Analysis of Variance of SPB for ML

Factor	*df*	*Adjusted SS*	*Sequential SS*	*F*	*p*	η^2
FRMR	2	0.000108	0.000108	5.14	0.008	0.01492
N	4	0.005068	0.005068	120.02	0.000	0.70019
α_4	10	0.000371	0.000371	3.52	0.001	0.05125
FRMR*α_4	20	0.000175	0.000175	0.83	0.675	0.02417
*N**α_4	40	0.000586	0.000586	1.39	0.102	0.08096
Error	88	0.000929	0.001515			
Total	164		0.007238			

Note. SPB = mean squared parameter bias; ML = maximum likelihood; SS = sum of squares; FRMR = fixed root mean squared residual.

TABLE 4C
Analysis of Variance of SPB for GLS

Factor	*df*	*Adjusted SS*	*Sequential SS*	*F*	*p*	η^2
FRMR	2	0.203543	0.203543	4152.12	0.000	0.92674
N	4	0.011261	0.011261	114.86	0.000	0.05127
α_4	10	0.000941	0.000941	3.84	0.000	0.00428
FRMR*α_4	20	0.001303	0.001303	2.66	0.001	0.00059
*N**α_4	40	0.000428	0.000428	0.44	0.998	0.00195
Error	88	0.002156	0.002156			
Total	164		0.219633			

Note. SPB = mean squared parameter bias; GLS = generalized least squares; SS = sum of squares; FRMR = fixed root mean squared residual.

TABLE 4D
Analysis of Variance of SPB for WLS

Factor	*df*	*Adjusted SS*	*Sequential SS*	*F*	*p*	η^2
FRMR	2	0.375787	0.375787	4077.79	0.000	0.76543
N	4	0.065763	0.065763	356.81	0.000	0.13395
α_4	10	0.019942	0.019942	43.28	0.000	0.04062
FRMR*α_4	20	0.024317	0.024317	36.39	0.000	0.04953
*N**α_4	40	0.001082	0.001082	0.59	0.969	0.00220
Error	88	0.004055	0.004055			
Total	164		0.490946			

Note. SPB = mean squared parameter bias; WLS = weighted least squares; SS = sum of squares; FRMR = fixed root mean squared residual.

size. Theoretical fit increases with sample size, but decreases considerably with misspecification.

For WLS we find similar effects of misspecification and sample size as for GLS, but also find significant effects of kurtosis and the interaction of misspecification and kurtosis (FRMR*α_4). More detailed analyses show how the performance of the various estimation methods varies with misspecification and sample size. Let us first address the results for the least misspecified model, M_1 (FRMR = .026). As can be seen in Tables B8–B14, ML and GLS perform equally well and significantly better than WLS for M_1 when sample size is less than 500. For sample sizes in this range ML and GLS provide almost equal parameter estimates for all levels of kurtosis. To be precise, GLS performs somewhat better than ML. For the partially true model, the performance of WLS improves with larger sample sizes and provides fairly precise parameter estimates for sample sizes of 1,000 and 2,000 and, in fact, achieves higher theoretic fit than the other methods. For the sample sizes of 1,000 and 2,000, WLS also provides precise estimates. In fact, WLS is the most accurate for M_1 (the least misspecified model) of the three for N = 2,000, particularly for high values of kurtosis ($\alpha_4 \geq 9.35$).

For the intermediate model in terms of misspecification, M_2 (FRMR =.089), ML and GLS give almost identical SPB values for N = 100. For larger sample sizes, ML outperforms GLS across all levels of kurtosis. For the most severely misspecified model, M_3, the pattern observed for M_2 is even clearer: In 55 out of 55 comparisons ML provides lower SPB values (i.e., better) than GLS. The relative accuracy of ML is—as the ANOVA results suggest—very robust to violations of the normality and specification conditions. In contrast, the parameter bias of estimates obtained with GLS and WLS is strongly affected by misspecification and kurtosis.

A representative replication (Table 5) for a sample size of 250, which is a typical sample size in empirical studies, and the most severely misspecified model, M_3 (FRMR = .17), illustrates that the differences in theoretical fit between the three methods are quite substantial: For α_4 values from 5 to 28.45, SPB for ML varies from .0058 to .0095, whereas the corresponding value for GLS varies from .0715 to .0930. The SPB for GLS is almost 10 times larger than that for ML. However, the performance of WLS in terms of SPB is still worse as it varies from 0.0958 to .1527, and is for the most extreme case more than 16 times larger than for ML.

CONCLUSIONS, IMPLICATIONS, AND LIMITATIONS

The results can be summarized as follows: The performance in terms of empirical and theoretical fit of the three estimation methods is differentially affected by sample size, specification error, and kurtosis. Of the three methods, ML is

TABLE 5
Theoretical Fit of ML, GLS, and WLS as a Function of Kurtosis Relative to the Performance of ML for $\alpha_4 = 9.35$

α_4	*5*	*6*	*8*	*8.5*	*9.35*	*18.14*	*22.21*	*28.45*
ML	0.0061	0.0059	0.006	0.0069	0.0058	0.0091	0.0084	0.0095
	1.05	*1.01*	*1.03*	*1.18*	*1.00*	*1.56*	*1.45*	*1.64*
GLS	0.0745	0.0715	0.0734	0.0822	0.0745	0.0930	0.0909	0.0906
	12.84	*12.33*	*12.65*	*14.17*	*12.84*	*16.03*	*15.67*	*15.62*
WLS	0.1002	0.0958	0.1013	0.1250	0.1152	0.1527	0.1483	0.1486
	17.28	*16.52*	*17.46*	*21.55*	*19.86*	*26.33*	*25.57*	*25.62*

Note. The ratios SPB/0.0058 are in the second row. ML = maximum likelihood; GLS = generalized least squares; WLS = weighted least squares; SPB = mean squared parameter bias.

considerably more insensitive than the two others to variations in sample size and kurtosis. Only empirical fit is affected by specification error—as it should be. Moreover, ML tends in general not only to be more stable, but also demonstrates higher accuracy in terms of empirical and theoretical fit compared to the other estimators.

GLS requires well-specified models, but allows small sample sizes to do an acceptable job in terms of theoretical and empirical fit. As reported by Olsson et al. (1999), its appealing performance in terms of empirical fit can easily mislead researchers to retain and thus interpret parameters of a misspecified model.

WLS also requires well-specified models, but in contrast to GLS and ML it also requires large sample sizes to perform well. It not only allows peaked data, but actually appears to reward the researcher for using nonnormal data: The more highly peaked the data, the more encouraging its estimates of empirical fit. However, this apparent advantage is misleading: Our analyses suggest that the RMSEA values are inflated and that the apparent, but misleading advantage in terms of empirical fit is obtained at the cost of more inaccurate parameter estimates.

Our goal is not to declare a winner. Of course, within the constraints of this study, we admit that our choice of estimator would be ML. However, there is an alternative implication of our results.[2] Rather than choosing one estimation method, it could be argued that researchers may be well advised to employ a triangulation approach using all three. If the methods provide similar parameter estimates, one

[2]This idea is based on the Hausman Test, discussed in Arminger and Schoenber (1989) and in White (1994): "If the model is correctly specified, these different estimators should have similar values asymptotically. If these values are not sufficiently similar, then the model is not correctly specified" (p. 218). The obvious problem is that the methods are asymptotically similar, and might be problematic to use in practical situations where sample sizes often are quite small.

has an additional indication that the correct structure is identified and that the parameter estimates are accurate. When this is the case, additional precision in parameter estimates can be gained by using GLS when the sample is small. If the sample size is large and the kurtosis very high, improved theoretical fit in terms of parameter estimates can be achieved by choosing to rely on those provided by WLS. In all other cases (i.e., when the methods give discrepant results or sample sizes and kurtosis falls outside the ranges that suggest the choice of GLS and WLS) ML seems to be the natural choice.

Nonnormality can also affect significance tests. The reason for this is that standard errors can be underestimated when data are nonnormal. A comparison of LISREL estimates of standard errors and estimates of standard errors based on 100 replications in this study partly supports the findings reported by Browne (1984) and Finch et al. (1997) that standard errors of parameter estimates can be considerably underestimated by ML and GLS. However, preliminary analysis suggests that the amount of misspecification can affect the underestimation of standard errors of ML and GLS relative to WLS. For severely misspecified models, estimated standard errors for ML appear to be more realistic than is the case is for GLS and particularly for WLS. Future studies should address the generalizability of this tentative conclusion.

More research is also needed to investigate whether across-method variability of parameter estimates can be used to assess what parts of the tested model, if any, are close to the true model. Preliminary analyses suggest that the discrepancy of parameter estimates given by the three methods is higher for the misspecified parts of the model.

We admit that our results and implications may not be generalizable across all types of misspecification and models. In the current research we focused on a family of nested models where incorrect models represent higher parsimony. More research should be conducted to investigate other types of specification error. Nevertheless, our analyses have underscored the fact that indexes of estimated empirical fit (e.g., RMSEA) cannot be easily compared across methods because both the population value (here EA_{df}) and the estimated value are dependent on the estimation method. The reliance on fixed criterion values of acceptance (e.g., RMSEA $< .05$), irrespective of estimation method, may therefore not be appropriate. The rather over-optimistic RMSEA values provided by GLS and WLS suggest that more stringent threshold levels for acceptance may be in order for these methods.

REFERENCES

Akaike, H. (1974). A new look at statistical model identification. *IEEE Transaction and Automatic Control*, AC–19, 716–723.

Arminger, G., & Schoenberg, R. J. (1989). Pseudo maximum likelihood estimation and a test for misspecification in mean and covariance structure models. *Psychometrika, 54*(3), 409–425.

Bentler, P. M. (1989). *EQS structural equations program manual.* Los Angeles: BMDP Statistical Software.

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

Boomsma, A. (1983). *On the robustness of LISREL (Maximum Likelihood estimation) against small sample size and nonnormality.* Amsterdam: Sociometric Research Foundation.

Browne, M. W. (1974). Generalized least-squares estimators in the analysis of covariance structures. *South African Statistical Journal, 8*,1–24.

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

Browne, M. W. (1987). Robustness of statistical inference in factor analysis and related models. *Biometrica, 74,* 375–384.

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

Chou, C. P., Bentler, P., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology, 44*, 347–357.

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

Ding, L., Velicer, W. F., & Harlow, L. L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling, 2*, 119–141.

Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling, 2*, 87–105.

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

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

Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7: A guide to the program and applications* (2nd ed.). New York: SPSS.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the Simplis command language.* Chicago: Scientific Software International.

Luijben, T. C. W. (1991). Equivalent models in covariance structure analysis. *Psychometrika, 56*, 653–665.

MacCallum, R. C. (1986). Specification searches in covariance structure modeling. P*sychological Bulletin, 100*, 107–120.

MacCallum, R. C., & Tucker, L. R. (1991). Representing sources of error in the common-factor model: Implication for theory and practice. *Psychological Bulletin, 109*, 502–511.

MacCallum, R. C., Wegener, D. T., Uchino B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin, 114*, 185–199.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis.* New York: Academic.

Mattson, S. (1997). How to generate non-normal data for simulation of structural equation models. *Multivariate Behavioral Research, 32*(4), 355–373.

Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry, 1*, 108–141.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennet, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin, 105*, 430–445.

Olsson, U. H., Troye, S. V., & Howell, R. D. (1999). Theoretic fit and empirical fit: The performance of maximum likelihood versus generalized least squares in structural equation models. *Multivariate Behavioral Research, 34*, 31–59.

Ramberg, J. S., Tadikamalla, P. R., Dudewicz, E. J., & Mykytka, E. F. (1979). A probability distribution and its use in fitting data. *Technometrics, 21,* 201–215.

Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.

Steiger, J. H. (1990). Structural model evaluation modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173–180.

Tukey, J. W. (1960). *The practical relationship between the common transformations of percentages of counts and of amounts* (Tech. Rep. 36). Princeton, NJ: Statistical Techniques Research Group, Princeton University.

White, H. (1994). *Estimation, inference and specification analysis*. New York: Cambridge University Press.

Yuan, K. H., & Bentler, P. M. (1997). Improving parameter tests in covariance structure analysis. *Computational Statistics & Data Analysis, 26,* 177–198.

APPENDIX A

Error of Approximation (EA) for Maximum Likelihood (ML), Generalized Least Squares (GLS), and Weighted Least Squares (WLS)

For ML and GLS, EA_{df} is independent of the fourth-order moment. It only depends on elements and products of elements from $\Sigma(\theta)$ and S. The population fit functions are given by $F_{ML}(\Sigma, \Sigma(\theta)) = (\log|\Sigma(\theta)| + \mathrm{tr}\{\Sigma\Sigma(\theta)^{-1}\} - \log|\Sigma| - q)$; q being the number of observed variables and $F_{GLS}(\Sigma, \Sigma(\theta)) = 1/2\mathrm{tr}\{(I - \Sigma(\theta) \cdot \Sigma^{-1})^2\}$. As we know, GLS can equivalently be written as the quadratic form: $F_{GLS}(\delta) = (\sigma - \sigma_\theta)' U_{GLS}^{-1}(\sigma - \sigma_\theta)$, where the weight matrix now is given by $[U_{GLS}]_{ij,gh} = \sigma_{ig}\sigma_{jh} + \sigma_{ih}\sigma_{jg}$; $i \geq j$ and $g \geq h$.

Note that we now are using elements of Σ in the weight matrix instead of elements of S.

In Table A1 the EA values for ML and GLS are presented.

For the WLS fit function $F_{WLS}(\theta) = (\sigma - \sigma_\theta)' U_{WLS}^{-1} (\sigma - \sigma_\theta)$, the elements of the weight matrix are of the form: $[U_{WLS}]_{ij,gh} = \sigma_{ijgh} - \sigma_{ij}\sigma_{gh}$ where σ_{ijgh} is the fourth-order population moments and $\sigma_{ij}\sigma_{gh}$ is the product of second-order population moments. The derivation of the EA_{WLS} is therefore a much more complicated task than for ML and GLS. In the following section we give a brief discussion of the derivation of the population elements in the weight matrix for WLS.

TABLE A1
EA_{ML} and EA_{GLS}

FRMR/Method	*ML*	*GLS*
0.026	0.05047997	0.046075966
0.089	0.0621080	0.0555370
0.17	0.14442252	0.104266835

Note. EA = error of approximation; ML = maximum likelihood; GLS = generalized least squares; FRMR = fixed root mean squared residual.

We start with the true model

$$X = \Lambda_x \xi + \delta \tag{1}$$

and show that the population elements of U for WLS is a function of this model. To do this it is convenient to write the true model on a simpler form:

$$X = \Lambda_x \xi + \delta = A\omega \tag{2}$$

where A is a $q \times (q + k)$ matrix and ω is a $(q + k) \times 1$ vector of independent variables where $E(\omega)$ and $E(\omega\omega') = I$. We have assumed that X is a $q \times 1$ vector of observed variables, Λ_x is a $q \times k$ matrix of factor loadings, ξ is a $k \times 1$ vector of latent variables, and δ is a $q \times 1$ vector of stochastic error terms. Following the tradition we assume that $E(\xi) = 0$ and $\mathrm{Var}(\xi) = 1$. The covariance of ξ, $E(\xi\,\xi') = \Phi$ will therefore be a $k \times k$ matrix with 1s along the diagonal. We also assume that $E(\delta) = 0$ and $\mathrm{Var}(\delta) = \Theta_\delta$. The argument for writing $X = \Lambda_x \xi + \delta = A\omega$ is as follows:

Because Φ is positive definite there will exist a $k \times k$ matrix P so that $\Phi = PP'$.

We can then write $\xi = P\omega_1$ where ω_1 is a $k \times 1$ vector of independent stochastic variables.

Let $D = \begin{bmatrix} \sqrt{\Theta_{\delta 1}} & 0 & .. & 0 \\ 0 & \sqrt{\Theta_{\delta 2}} & .. & 0 \\ 0 & .. & .. & .. \\ 0 & .. & .. & \sqrt{\Theta_{\delta q}} \end{bmatrix}$ be a $q \times q$ matrix where the standard error of δ is along the main diagonal, and zeros elsewhere.

Then $\delta = D\omega_2$ where ω_2 is a $q \times 1$ vector of independent stochastic variables where $E(\omega_2) = 0$ and $Var(\omega_2) = 1$. We can now write

$$X = \Lambda_x \xi + \delta = \Lambda_x P\omega_1 + D\omega_2 = (\Lambda_x P | D)\left(\frac{\omega_1}{\omega_2}\right) + A\omega$$

where A is composed of the $q \times k$ matrix $\Lambda_x P$ and the $q \times q$ matrix D. A is therefore a $q \times (q + k)$ matrix. The vector ω is composed of the $k + 1$ vector ω_1 and the $q \times 1$ vector ω_2. ω is therefore a $(q + k) \times 1$ vector of independent variables where $E(\omega) = 0$ and $E(\omega\, \omega') = I$.

The covariance matrix Σ can then be written as

$$E(XX') = E(A\omega\omega' A') = AE(\omega\omega')A' = AI\, A' = AA'. \tag{3}$$

Let U be the asymptotic covariance matrix.

A typical element of U is $\sigma_{ijgh} - \sigma_{ij}\sigma_{gh}$ where $\sigma_{ijgh} = E(X_i X_j X_g X_h)$ and σ_{rs} is the population covariance between X_r and X_s. The matrix $E((XX') \otimes (XX'))$ will have σ_{ijgh} as a typical element.

If we use $X = A\omega$ and the fact that $(EF) \otimes (GH) = (E \otimes G)(F \otimes H)$ (*), the Kroenecker product $(XX') \otimes (XX')$ can be written as

$$(XX') \otimes (XX') = (A\omega\omega' A') \otimes (A\omega\omega' A') = ((\mathrm{A}\omega) \otimes (A\omega))((\omega' A') \otimes (\omega' A')) \tag{4}$$

where we have set $E = G = A\omega$ and $F = H = \omega' A'$ from (*). In the same way we can write $(A\omega) \otimes (A\omega) = (A \otimes A)(\omega \otimes \omega)$ and $(\omega' A') \otimes (\mathrm{w}' A') = (\omega' \otimes \omega')(A' \otimes A')$.

Substituting this into the equation for $(XX') \otimes (XX')$ we get

$$(XX') \otimes (XX') = (A \otimes A)(\omega \otimes \omega)(\omega' \otimes \omega')(A' \otimes A') = (A \otimes A)(\omega\omega') \otimes (\omega\omega')(A \otimes A)'$$

The expectation of $(XX') \otimes (XX')$ is a $q^2 \times q^2$ matrix that we denote T:

$$T = E((XX') \otimes (XX')) = (A \otimes A)E((\omega\omega') \otimes (\omega\omega'))(A \otimes A)' \tag{5}$$

$A \otimes A$ is a $q^2 \times (q + k)^2$ matrix, $E((\omega\omega') \otimes (\omega\omega'))$ is a $(n + k)^2 \times (n + k)^2$ matrix and $(A \otimes A)'$ is a $(q + k)^2 \times q^2$ matrix.

A typical element in the matrix $E((\omega\omega') \otimes (\omega\omega'))$ is $E(\omega_i\, \omega_j\, \omega_g\, \omega_h)$. Because ω is a vector of independent stochastic variables this expression is easy to calculate:

$$E(\omega_i\, \omega_j\, \omega_g\, \omega_h) = E(\omega_i^4) = \mu_4 \text{ if } I = j = g = h$$

$$E(\omega_i\, \omega_j\, \omega_g\, \omega_h) = \mu_2{}^2 = 1 \text{ if } (i = j \text{ and } g = h \text{ and } i \neq g)$$
$$\text{or } (i = g \text{ and } j = h \text{ and } i \neq j)$$
$$\text{or } (i = h \text{ and } j = g \text{ and } i \neq j)$$

$$E(\omega_i\, \omega_j\, \omega_g\, \omega_h) = 0 \text{ elsewhere.}$$

With the earlier derivations we can calculate $E((\omega\omega') \otimes (\omega\omega'))$. T is calculated from Equation 5. We also wish to calculate the products of the second-order moments: $\sigma_{ij}\ \sigma_{gh}$.

This is done by letting $\Sigma = AA'$, then the matrix $\Sigma \otimes \Sigma = AA' \otimes AA' = (A \otimes A)(A \otimes A)'$ has $\sigma_{ij}\ \sigma_{gh}$ as a typical element.

For the WLS weight matrix we then have

$$U = T - \Sigma \otimes \Sigma = (A \otimes A)E((\omega\omega') \otimes (\omega\omega'))(A \otimes A)' - (A \otimes A)(A \otimes A)' \quad (6)$$

In Tables A2–A4 we have calculated the EA_{df} values for WLS for the three models with 11 different levels of kurtosis.

Thus we see that EA_{df} decreases significantly when α_4 goes up. This tendency is a result of the presence of the fourth-order moment in the weight matrix U. It is therefor also likely that the RMSEA will decrease as a function of α_4. Because RMSEA is an estimate of EA it will be affected by the sample size. As showed by Hu, Bentler, and Kano (1992) the WLS (ADF) almost never diagnosed the true model for sample sizes below 2,500 (i.e., it was positively biased). So it is reasonable that the RMSEA index estimated with WLS will show positive bias also for misspecified models, especially for small sample size.

Data Generating Process

The data were generated according to the following design: From Equation 2 we can write the true model

$$X = A\omega \quad (7)$$

TABLE A2
Negligible kurtosis: EA_{WLS}

RMR/α_4	*1.8*	*2*	*2.4*
0.026	0.047112	0.046931	0.046575
0.089	0.062358	0.060807	0.058303
0.17	0.121726	0.117990	0.111717

Note. EA = error of approximation; WLS = weighted least squares; RMR = root mean squared residual.

TABLE A3
Medium Kurtosis: EA_{WLS}

RMR/α_4	*5*	*6*	*8*	*8.5*	*9.35*
0.026	0.044560	0.043824	0.042485	0.041759	0.040867
0.089	0.050139	0.048363	0.045722	0.044463	0.043471
0.17	0.089244	0.084115	0.076561	0.073006	0.071038

Note. EA = error of approximation; WLS = weighted least squares; RMR = root mean squared residual.

TABLE A4
Extreme Kurtosis: EA_{WLS}

RMR/α_4	*18.14*	*22.21*	*28.45*
0.026	0.037405	0.035871	0.033890
0.089	0.038486	0.036661	0.034407
0.17	0.057802	0.053739	0.049044

Note. EA = error of approximation; WLS = weighted least squares; RMR = root mean squared residual.

Generating a vector ω will give us an X vector. Ramberg et al. (1979) showed how we can generalize Tukey's (1960) Lamda function to a four-parametric distribution given by the equation

$$R(p) = \lambda_1 + \left[p^{\lambda_3} - (1-p)^{\lambda_4} \right] / \lambda_2 \qquad 0 \le p \le 1$$

where λ_1 is a location parameter, λ_2 is a scale, and λ_3 and λ_4 are shape parameters (Ramberg, Tadikamalla, Dudewicz, and Mykytka, 1979). Drawing randomly an element p from the uniform distribution over the interval [0,1], $R(p)$, will be a random variable from a distribution of finite moments of first-, second-, third-, and fourth-order. The magnitude of these moments will depend on the values of λ_1 to λ_4.

Letting μ_1, μ_2, α_3, and α_4 be the expectation, variance, skewness, and kurtosis to $R(p)$, we can use the tables in Ramberg et al. (1979) to find the correct choice for the lambdas that will give $R(p)$ with the desired values of μ_1, μ_2, α_3, and α_4.

In our study we choose the lambdas so that $\mu_1 = 0$ and $\mu_2 = 1$ (i.e., the mean is zero and the variance is 1). This will imply that the skewness and kurtosis will be the same as the third- and fourth-order moments (or –3 if we shall follow the tradition of using the normal distribution as a reference). Using this approach we can generate the vector ω. ω is a vector of independent stochastic variables. $X = A\omega$ will give us the realizations of the observed variables. The fourth-order

moments for x_i, $E(x_i^4)$ will be functions of the first-, second-, third-, and fourth-order moments for ω_i, which are given by

$$E(x \otimes x' \otimes x \otimes x') = E((xx') \otimes (xx')) = (A \otimes A)(\omega\omega') \otimes (\omega\omega')(A \otimes A)' \quad (8)$$

Because a typical element in this matrix is $E(X_i X_j X_g X_h)$, the fourth-order moment element for x_i, is taken to be $E(X_i X_i X_i X_i)$ from this moment matrix.

Alternatively this can be done in a computationally less intensive way: Letting

$$A = \begin{pmatrix} a_1' \\ a_2' \\ . \\ . \\ a_q' \end{pmatrix}$$

where a_i' is row no. i. a_i will be a $(q \times 1)$ matrix and $E(x_i^4) = (a_i \otimes a_i)' (\omega\omega') \otimes (a_i \otimes a_i)$. Because the ω_i's are independent we can write

$$E(x_i^4) = \sum_{m=1}^{q+n} a_{im}^4 \mu_{4m} + 6 \sum_{m=2}^{q+n} \sum_{k}^{m-1} a_{im}^2 a_{ik}^2$$

The third-order moment of x is given by $E(x \otimes x' \otimes x)$ and

$$x \otimes x' \otimes x = (xx') \otimes x = (A\omega\omega' A') \otimes A\omega = (A\omega \otimes A\omega)\omega' A' = (A \otimes A)(\omega \otimes \omega)\omega' A' = (A \otimes A)(\omega\omega') \otimes \omega) A'.$$

TABLE A5
Expected Fourth-Order Moments $E(x_i^4)$

α_4	*x1*	*x2*	*x3*	*x4*	*x5*	*x6*	*x7*	*x8*	*x9*	*x10*	*x11*	*x12*
1.8	2.58	2.63	2.58	3.78	2.67	2.69	2.69	2.71	2.50	2.56	2.41	2.51
2	2.65	2.69	2.65	2.87	2.72	2.74	2.76	2.76	2.58	2.64	2.50	2.60
2.4	2.79	2.81	2.79	4.05	2.84	2.85	2.85	2.85	2.75	2.78	2.68	2.78
5	3.68	3.60	3.68	5.20	3.54	3.49	3.49	3.46	3.86	3.71	3.80	3.97
6	4.02	3.91	4.02	5.66	3.81	3.74	3.74	3.70	4.23	4.07	4.24	4.44
8	4.71	4.52	4.71	6.56	4.36	4.25	4.25	4.17	5.05	4.80	5.11	5.35
8.5	5.14	4.90	5.14	7.11	4.70	4.56	4.56	4.46	5.57	5.25	5.65	5.93
9.35	5.36	5.11	5.36	7.43	4.91	4.75	4.75	4.65	5.82	5.48	5.91	6.20
18.14	8.19	7.61	8.19	11.11	7.13	6.78	6.78	6.54	9.23	8.45	9.51	10.02
22.21	9.58	8.85	9.58	12.92	8.23	7.79	7.79	7.49	10.9	9.91	11.28	11.88
28.45	11.71	10.71	11.71	15.72	9.93	9.35	9.35	8.95	13.46	12.16	13.98	14.75

Because ω is a vector of stochastic independent variables we can easily calculate $E(\omega\omega'\otimes\omega)$. The third-order moment is therefore $E(x \otimes x'\otimes x) = (A \otimes A)E(\omega\omega'\otimes\omega)A'$. A typical element will be on the form $E(x_i\, x_j\, x_g)$.

The third-order moment $E(x_i^3)$ will be one element in the moment matrix.

A computationally less intensive way of calculating $E(x_i^3)$ is by using

$$E(x_i^3) = (a_i \otimes a_i)'\, E((\omega\omega)'\otimes\omega)a_i$$

Because ω is a vector of stochastic independent variables, this can be reformulated as (Mattson, 1997).

By generating N realizations of ω we can generate N realizations of X. This raw data of xs is used as input in LISREL 8.14 for ML and GLS. For WLS the raw data is first used as input in PRELIS 2.12 for calculating the asymptotic covariance matrix.

The expectation values, $E(x_i^4) = \sum_{m=1}^{q+n} a_{im}^4 \mu_{4m} + 6\sum_{m=2}^{q+n}\sum_{k}^{m-1} a_{im}^2 a_{ik}^2$, of the fourth-order moments for the 12 observed variables under the 11 different α_4 values are given in Table A5.

APPENDIX B

In matrix notation the true model can be expressed as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \\ x_{11} \\ x_{12} \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 0 & 0 & \lambda_{14} & 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 & \lambda_{25} & 0 & 0 \\ \lambda_{31} & 0 & 0 & 0 & 0 & \lambda_{36} & 0 \\ \lambda_{41} & 0 & 0 & 0 & 0 & \lambda_{46} & 0 \\ 0 & \lambda_{52} & 0 & \lambda_{54} & 0 & 0 & 0 \\ 0 & \lambda_{62} & 0 & 0 & \lambda_{65} & 0 & 0 \\ 0 & \lambda_{72} & 0 & 0 & 0 & 0 & \lambda_{77} \\ 0 & \lambda_{82} & 0 & 0 & 0 & 0 & \lambda_{87} \\ 0 & 0 & \lambda_{93} & \lambda_{94} & 0 & 0 & 0 \\ 0 & 0 & \lambda_{103} & 0 & \lambda_{105} & 0 & 0 \\ 0 & 0 & \lambda_{113} & 0 & 0 & \lambda_{116} & 0 \\ 0 & 0 & \lambda_{123} & 0 & 0 & 0 & \lambda_{12,7} \end{pmatrix} \cdot \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \\ \xi_6 \\ \xi_7 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \\ \delta_{10} \\ \delta_{11} \\ \delta_{12} \end{pmatrix}$$

The φ matrix is given by

$$\begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} & 0 & 0 & 0 & 0 \\ \phi_{21} & \phi_{22} & \phi_{23} & 0 & 0 & 0 & 0 \\ \phi_{31} & \phi_{32} & \phi_{33} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \phi_{44} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \phi_{55} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \phi_{66} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \phi_{77} \end{pmatrix}$$

The loadings from the domain factors were given the following values:

$\lambda_{i1} = 0.7$ for $i = 1, 2, 3$, and 4; $\lambda_{i2} = 0.8$ for $i = 5, 6, 7$, and 8; and $\lambda_{i3} = 0.6$ for $i = 9, 10, 11$, and 12.

The loadings from the minor factors (λ_{i4}) were set to 0.2 for $i = 1, 5$, and 9. The values if the other loadings were set as follows: $\lambda_{i5} = 0.3$ for $i = 2, 6$, and 10. $\lambda_{36} = 0.2$, $\lambda_{46} = 0.4$, $\lambda_{11,6} = 0.1$ $\lambda_{77} = 0.3$, $\lambda_{87} = 0.4$, and $\lambda_{12,7} = 0.1$. For the variance–covariance matrix of the latent variables (φ matrix), the variances were fixed to one ($\phi_{ii} = 1.0$) and the correlations between the three major factors were set to 0.2 ($\phi_{21} = \phi_{31} = \phi_{32} = 0.2$). The covariance matrix Θ_δ is diagonal and var(δ_i) is a function of the model and the fact that var(x_i) = 1 for $i = 1, 2, \ldots, 12$.

TABLE B1
RMSEA for ML, GLS, and WLS ($N = 100$)

α_4	*5*	*6*	*8*	*8.5*	*9.35*	*18.14*	*22.21*	*28.45*
FRMR								
0.026	0.0517	0.0491	0.0552	0.0506	0.0527	0.0487	0.0520	0.0527
	(100)	(100)	(100)	(99)	(100)	(100)	(100)	(100)
	0.0298	0.0274	0.0309	0.0278	0.0288	0.0242	0.0296	0.0270
	(100)	(100)	(100)	(98)	(100)	(100)	(100)	(100)
	0.1120	0.1138	0.1164	0.1114	0.1153	0.1112	0.1103	0.1147
	(83)	(85)	(79)	(71)	(79)	(74)	(75)	(69)
0.089	0.0619	0.0652	0.0658	0.0677	0.0618	0.0648	0.0697	0.0702
	(100)	(100)	(99)	(100)	(100)	(100)	(99)	(100)
	0.0340	0.0377	0.0384	0.0410	0.0357	0.372	0.0402	0.0407
	(99)	(100)	(100)	(100)	(100)	(98)	(97)	(100)
	0.1303	0.1310	0.1267	0.1319	0.1291	0.1245	0.1269	0.1277
	(73)	(79)	(71)	(71)	(61)	(60)	(69)	(72)

(*continued*)

TABLE B1 (Continued)

α_4	*5*	*6*	*8*	*8.5*	*9.35*	*18.14*	*22.21*	*28.45*
0.17	0.1469	0.1455	0.1480	0.1495	0.1504	0.1529	0.1508	0.1555
	(99)	(98)	(100)	(99)	(100)	(99)	(100)	(99)
	0.0852	0.0877	0.0884	0.0893	0.0878	0.0893	0.0868	0.0917
	(46)	(55)	(53)	(34)	(46)	(56)	(45)	(47)
	0.1796	0.1749	0.1694	0.1620	0.1734	0.1596	0.1835	0.1647
	(19)	(21)	(12)	(12)	(17)	(9)	(8)	(11)

Note. RMSEA = root mean squared error of approximation; ML = maximum likelihood; GLS = generalized least squares; WLS = weighted least squares; FRMR = fixed root mean squared residual. A total of 500 repetitions gives ML = 0.0638; GLS = 0.0363, and WLS = 0.1297.

TABLE B2
RMSEA for ML, GLS, and WLS ($N = 250$)

α_4	5	6	8	8.5	9.35	18.14	22.21	28.45
FRMR								
0.026	0.0507	0.0502	0.0498	0.0512	0.0471	0.0510	0.0484	0.0510
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0412	0.0414	0.0405	0.0426	0.0379	0.0422	0.0387	0.0417
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0615	0.0606	0.0585	0.0586	0.0566	0.575	0.0554	0.0567
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(99)
0.089	0.0653	0.0604	0.0649	0.0636	0.0651	0.0668	0.0653	0.0653
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0526	0.0496	0.0521	0.0511	0.0523	0.0537	0.0524	0.0525
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0695	0.0651	0.0655	0.0644	0.0641	0.0632	0.0635	0.0623
	(100)	(100)	(100)	(100)	(99)	(100)	(100)	(99)
0.17	0.1452	0.1438	0.1451	0.1475	0.1464	0.1482	0.1450	0.1473
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0990	0.0980	0.0983	0.0997	0.0994	0.0994	0.0983	0.0983
	(89)	(89)	(83)	(79)	(88)	(89)	(80)	(79)
	0.1109	0.1070	0.1029	0.1030	0.1043	0.083	0.0962	0.0954
	(50)	(52)	(53)	(36)	(49)	(44)	(40)	(47)

Note. RMSEA = root mean squared error of approximation; ML = maximum likelihood; GLS = generalized least squares; WLS = weighted least squares; FRMR = fixed root mean squared residual.

TABLE B3
RMSEA for ML, GLS, and WLS ($N = 500$)

α_4	5	6	8	8.5	9.35	18.14	22.21	28.45
FRMR								
0.026	0.0503	0.0512	0.0505	0.0504	0.0493	0.0507	0.0490	0.0502
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0442	0.0451	0.0442	0.0437	0.0429	0.0443	0.0428	0.0440
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0494	0.0497	0.0478	0.0471	0.0473	0.466	0.0456	0.0453
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
0.089	0.0605	0.0633	0.0613	0.0624	0.0644	0.0654	0.0632	0.0644
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0522	0.0541	0.0526	0.0535	0.0554	0.0548	0.0547	0.0551
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0537	0.0545	0.0525	0.0515	0.0545	0.0494	0.0504	0.0489
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
0.17	0.1445	0.1457	0.1453	0.1446	0.1452	0.1459	0.1465	0.1445
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.1015	0.1017	0.1019	0.1015	0.1021	0.1015	0.1019	0.1011
	(96)	(96)	(98)	(97)	(93)	(96)	(92)	(94)
	0.0980	0.0947	0.0907	0.0889	0.0882	0.0824	0.0822	0.0808
	(75)	(73)	(69)	(74)	(60)	(55)	(50)	(58)

Note. RMSEA = root mean squared error of approximation; ML = maximum likelihood; GLS = generalized least squares; WLS = weighted least squares; FRMR = fixed root mean squared residual.

TABLE B4
RMSEA for ML, GLS, and WLS ($N = 1000$)

α_4	5	6	8	8.5	9.35	18.14	22.21	28.45
FRMR								
0.026	0.0504	0.0505	0.0502	0.0509	0.0504	0.0514	0.0502	0.0506
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0454	0.0454	0.0449	0.0455	0.0451	0.0460	0.0450	0.0452
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0468	0.0461	0.0452	0.0452	0.0444	0.429	0.0420	0.0414
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
0.089	0.0632	0.0631	0.0618	0.0621	0.0632	0.0639	0.0634	0.0626
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0556	0.0553	0.0545	0.0546	0.0553	0.0559	0.0554	0.0553
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0526	0.0513	0.0484	0.0484	0.0485	0.0462	0.0461	0.0415
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
0.17	0.1446	0.1447	0.1445	0.1450	0.1450	0.1455	0.1462	0.1450
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.1031	0.1032	0.1026	0.1029	0.1028	0.1029	0.1035	0.1030
	(97)	(99)	(98)	(99)	(99)	(100)	(99)	(100)
	0.0934	0.0894	0.0844	0.0830	0.0831	0.0770	0.0742	0.0727
	(91)	(93)	(85)	(79)	(88)	(74)	(63)	(74)

Note. RMSEA = root mean squared error of approximation; ML = maximum likelihood; GLS = generalized least squares; WLS = weighted least squares; FRMR = fixed root mean squared residual.

TABLE B5
RMSEA for ML, GLS, and WLS (N = 2,000)

α_4	*5*	*6*	*8*	*8.5*	*9.35*	*18.14*	*22.21*	*28.45*
FRMR								
0.026	0.0503	0.0501	0.0501	0.0503	0.0511	0.0513	0.0507	0.0501
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0457	0.0454	0.0454	0.0456	0.0462	0.0466	0.0458	0.0453
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0454	0.0447	0.0431	0.0432	0.0440	0.414	0.0409	0.0402
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
0.089	0.0618	0.0619	0.0626	0.0627	0.0621	0.0627	0.0623	0.0619
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0549	0.0549	0.0554	0.0555	0.0551	0.0555	0.0553	0.0549
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0509	0.0495	0.0472	0.0467	0.0465	0.0439	0.0432	0.0421
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
0.17	0.1444	0.1442	0.1445	0.1447	0.1449	0.1447	0.1450	0.1458
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.1032	0.1033	0.1036	0.1035	0.1037	0.1035	0.1038	0.1039
	(100)	(100)	(98)	(99)	(100)	(100)	(100)	(100)
	0.0864	0.0907	0.0810	0.0797	0.0785	0.0716	0.0696	0.0669
	(99)	(93)	(94)	(92)	(96)	(92)	(94)	(89)

Note. RMSEA = root mean squared error of approximation; ML = maximum likelihood; GLS = generalized least squares; WLS = weighted least squares; FRMR = fixed root mean squared residual.

TABLE B6
RMSEA for ML, GLS, and WLS: Negative Kurtosis

α_4	*1.8*[a]	*2.0*[a]	*2.4*[a]	*1.8*[b]	*2.0*[b]	*2.4*[b]	*1.8*[c]	*2.0*[c]	*2.4*[c]
FRMR									
0.026	0.0477	0.0518	0.0495	0.0506	0.0474	0.0477	0.0488	0.0499	0.0501
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0268	0.0278	0.0271	0.0414	0.0389	0.0388	0.0426	0.0439	0.0441
	(100)	(100)	(99)	(100)	(100)	(100)	(100)	(100)	(100)
	0.1164	0.1183	0.1198	0.0635	0.0599	0.0613	0.0499	0.0515	0.0507
	(86)	(86)	(80)	(100)	(100)	(100)	(100)	(100)	(100)
0.089	0.0614	0.0614	0.0626	0.0612	0.0602	0.0619	0.0613	0.0616	0.0622
	(100)	(100)	(100)	(100)	(100)	100	(100)	(100)	99
	0.0337	0.0341	0.0362	0.0504	0.0485	0.0505	0.0530	0.0529	0.0533
	(100)	(100)	(98)	(100)	(100)	100	(100)	(100)	100
	0.1414	0.1415	0.1405	0.0777	0.0750	0.0732	0.0673	0.0653	0.0629
	(86)	(82)	(79)	(100)	(100)	100	(100)	(100)	100
0.17	0.1446	0.1452	0.1452	0.1453	0.1444	0.1435	0.1435	0.1444	0.1440
	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)	(100)
	0.0862	0.0844	0.0859	0.0994	0.0982	0.0975	0.1013	0.1018	0.1014
	(58)	(37)	(59)	(78)	(100)	(81)	(94)	(98)	(96)
	0.2076	0.2070	0.1897	0.1421	0.1356	0.1296	0.1279	0.1258	0.1192
	(22)	(17)	(26)	(72)	(74)	(76)	(97)	(97)	(88)

Note. RMSEA = root mean squared error of approximation; ML = maximum likelihood; GLS = generalized least squares; WLS = weighted least squares; FRMR = fixed root mean squared residual.
[a]n = 100. [b]n = 250. [c]n = 500.

TABLE B7
RMSEA for ML, GLS, and WLS: Negative Kurtosis

α_4	*1.8*[a]	*2.0*[a]	*2.4*[a]	*1.8*[b]	*2.0*[b]	*2.4*[b]
FRMR						
0.026	0.0501	0.0502	0.0508	0.0502	0.0505	0.0502
	(100)	(100)	(100)	(100)	(100)	(100)
	0.0451	0.0452	0.0456	0.0456	0.0457	0.0456
	(100)	(100)	(100)	(100)	(100)	(100)
	0.0482	0.0485	0.0485	0.0476	0.0476	0.0470
	(100)	(100)	(100)	(100)	(100)	(100)
0.089	0.0625	0.0620	0.0622	0.0626	0.0619	0.0618
	(100)	(100)	(100)	(100)	(100)	(100)
	0.0547	0.0546	0.0547	0.0555	0.0549	0.0549
	(100)	(100)	(100)	(100)	(100)	(100)
	0.0643	0.0626	0.0601	0.0636	0.0615	0.0588
	(100)	(100)	(100)	(100)	(100)	(100)
0.17	0.1436	0.1447	0.1447	0.1444	0.1440	0.1443
	(100)	(100)	(100)	(100)	(100)	(100)
	0.1023	0.1028	0.1030	0.1036	0.1033	0.1035
	(98)	(100)	(100)	(100)	(100)	(100)
	0.1242	0.1208	0.1155	0.1233	0.1195	0.1133
	(100)	(100)	(95)	(100)	(100)	(100)

Note. RMSEA = root mean squared error of approximation; ML = maximum likelihood; GLS = generalized least squares; WLS = weighted least squares; FRMR = fixed root mean squared residual.
[a]n = 1,000. [b]n = 2,000.

TABLE B8
SPB: ML, GLS, and WLS (N = 100)

α_4	*5*	*6*	*8*	*8.5*	*9.35*	*18.14*	*22.21*	*28.45*
FRMR								
0.026	0.0117	0.0130	0.0138	0.0156	0.0141	0.0162	0.0174	0.0176
	0.0125	0.0143	0.0160	0.0187	0.0154	0.0171	0.0173	0.0183
	0.0382	0.0389	0.0400	0.0536	0.0488	0.0527	0.0489	0.0542
0.089	0.0277	0.0227	0.0248	0.0127	0.0159	0.0189	0.0478	0.0329
	0.0227	0.0205	0.0242	0.0230	0.0237	0.0303	0.0311	0.0258
	0.0551	0.0528	0.0588	0.0567	0.0603	0.0749	0.0702	0.0641
		[a]						
		0.0148						
		0.0226						
		0.0532						
0.17	0.0126	0.0141	0.0146	0.0160	0.0160	0.0224	0.0207	0.0182
	0.0928	0.0910	0.0959	0.1021	0.0938	0.1176	0.1201	0.1147
	0.1402	0.1449	0.1598	0.1636	0.1511	0.1883	0.2243	0.2053

Note. SPB = mean squared parameter bias; ML = maximum likelihood; GLS = generalized least squares; WLS = weighted least squares; FRMR = fixed root mean squared residual.
[a]Replications = 500.

TABLE B9
SPB: ML, GLS, and WLS ($N = 250$)

α_4	*5*	*6*	*8*	*8.5*	*9.35*	*18.14*	*22.21*	*28.45*
FRMR								
0.026	0.0056	0.0058	0.0060	0.0066	0.0083	0.0069	0.0068	0.0072
	0.0053	0.0057	0.0058	0.0064	0.0076	0.0067	0.0068	0.0073
	0.0080	0.0081	0.0090	0.0092	0.0095	0.0099	0.0099	0.0127
0.089	0.0051	0.0060	0.0067	0.0063	0.0072	0.0183	0.0094	0.0184
	0.0093	0.0100	0.0101	0.0098	0.0108	0.0125	0.0127	0.0119
	0.0131	0.0134	0.0155	0.0145	0.0156	0.0199	0.0198	0.0185
0.17	0.0061	0.0059	0.0060	0.0069	0.0058	0.0091	0.0084	0.0095
	0.0745	0.0715	0.0734	0.0811	0.0745	0.0930	0.0909	0.0906
	0.1002	0.0958	0.1013	0.1250	0.1152	0.1527	0.1483	0.1486

Note. SPB = mean squared parameter bias; ML = maximum likelihood; GLS = generalized least squares; WLS = weighted least squares; FRMR = fixed root mean squared residual.

TABLE B10
SPB: ML, GLS, and WLS ($N = 500$)

α_4	*5*	*6*	*8*	*8.5*	*9.35*	*18.14*	*22.21*	*28.45*
FRMR								
0.026	0.0033	0.0034	0.0035	0.0036	0.0038	0.0044	0.0048	0.0055
	0.0030	0.0031	0.0032	0.0035	0.0036	0.0040	0.0044	0.0050
	0.0036	0.0037	0.0037	0.0039	0.0040	0.0048	0.0056	0.0051
0.089	0.0034	0.0038	0.0039	0.0040	0.0041	0.0043	0.0043	0.0049
	0.0055	0.0058	0.0057	0.0060	0.0058	0.0070	0.0070	0.0072
	0.0067	0.0075	0.0074	0.0084	0.0074	0.0097	0.0107	0.0112
0.17	0.0037	0.0039	0.0040	0.0038	0.0045	0.0054	0.0053	0.0054
	0.0675	0.0688	0.0658	0.0676	0.0678	0.0806	0.0874	0.0821
	0.0902	0.0937	0.0978	0.1069	0.1046	0.1399	0.1415	0.1434

Note. SPB = mean squared parameter bias; ML = maximum likelihood; GLS = generalized least squares; WLS = weighted least squares; FRMR = fixed root mean squared residual.

TABLE B11
SPB: ML, GLS, and WLS (N = 1,000)

α_4	*5*	*6*	*8*	*8.5*	*9.35*	*18.14*	*22.21*	*28.45*
FRMR								
0.026	0.0022	0.0022	0.0025	0.0024	0.0026	0.0029	0.0032	0.0028
	0.0020	0.0020	0.0023	0.0021	0.0023	0.0026	0.0030	0.0026
	0.0020	0.0020	0.0023	0.0022	0.0022	0.0026	0.0028	0.0026
0.089	0.0023	0.0025	0.0025	0.0025	0.0028	0.0027	0.0029	0.0022
	0.0036	0.0037	0.0038	0.0039	0.0041	0.0047	0.0049	0.0038
	0.0042	0.0046	0.0047	0.0047	0.0050	0.0062	0.0067	0.0047
0.17	0.0033	0.0032	0.0034	0.0033	0.0032	0.0035	0.0038	0.0038
	0.0777	0.0775	0.0784	0.0782	0.0755	0.0775	0.0782	0.0785
	0.0980	0.1050	0.1119	0.1213	0.1154	0.1328	0.1399	0.1444

Note. SPB = mean squared parameter bias; ML = maximum likelihood; GLS = generalized least squares; WLS = weighted least squares; FRMR = fixed root mean squared residual.

TABLE B12
SPB: ML, GLS, and WLS (N = 2,000)

α_4	*5*	*6*	*8*	*8.5*	*9.35*	*18.14*	*22.21*	*28.45*
FRMR								
0.026	0.0017	0.0017	0.0017	0.0017	0.0019	0.0020	0.0019	0.0021
	0.0015	0.0015	0.0015	0.0015	0.0017	0.0018	0.0017	0.0019
	0.0015	0.0014	0.0014	0.0013	0.0014	0.0014	0.0014	0.0016
0.089	0.0018	0.0017	0.0019	0.0020	0.0020	0.0021	0.0020	0.0021
	0.0031	0.0030	0.0031	0.0035	0.0028	0.0034	0.0031	0.0035
	0.0033	0.0033	0.0036	0.0041	0.0032	0.0044	0.0040	0.0044
0.17	0.0021	0.0020	0.0021	0.0022	0.0025	0.0028	0.0029	0.0032
	0.0620	0.0625	0.0625	0.0633	0.0612	0.0754	0.0751	0.0756
	0.0867	0.0801	0.0961	0.0988	0.0987	0.1409	0.1429	0.1466

Note. SPB = mean squared parameter bias; ML = maximum likelihood; GLS = generalized least squares; WLS = weighted least squares; FRMR = fixed root mean squared residual.

TABLE B13
SPB: ML, GLS, and WLS: Negative Kurtosis

α_4	*1.8*[a]	*2.0*[a]	*2.4*[a]	*1.8*[b]	*2.0*[b]	*2.4*[b]	*1.8*[c]	*2.0*[c]	*2.4*[c]
FRMR									
0.026	0.0108	0.0108	0.0120	0.0052	0.0051	0.0051	0.0029	0.0034	0.0031
	0.0128	0.0121	0.0138	0.0050	0.0049	0.0051	0.0027	0.0031	0.0029
	0.0370	0.0362	0.0418	0.0070	0.0070	0.0068	0.0034	0.0038	0.0036
0.089	0.0113	0.0101	0.0111	0.0051	0.0052	0.0052	0.0033	0.0032	0.0031
	0.0189	0.0181	0.0192	0.0079	0.0082	0.0082	0.0051	0.0049	0.0052
	0.0456	0.0396	0.0429	0.0098	0.0101	0.0104	0.0057	0.0055	0.0057
0.17	0.0132	0.0147	0.0135	0.0060	0.0061	0.0065	0.0037	0.0040	0.0038
	0.1050	0.1151	0.1084	0.0908	0.0877	0.0849	0.0784	0.0803	0.0810
	0.1207	0.1400	0.1504	0.0761	0.0792	0.0810	0.0629	0.0704	0.0787

Note. SPB = mean squared parameter bias; ML = maximum likelihood; GLS = generalized least squares; WLS = weighted least squares; FRMR = fixed root mean squared residual.
[a]$n = 100$. [b]$n = 250$. [c]$n = 500$.

TABLE B14
SPB: ML, GLS, and WLS

α_4	*1.8*[a]	*2.0*[a]	*2.4*[a]	*1.8*[b]	*2.0*[b]	*2.4*[b]
FRMR						
0.026	0.0021	0.0020	0.0020	0.0016	0.0016	0.0016
	0.0018	0.0018	0.0018	0.0014	0.0014	0.0014
	0.0021	0.0020	0.0020	0.0016	0.0015	0.0015
0.089	0.0022	0.0021	0.0022	0.0017	0.0017	0.0018
	0.0036	0.0035	0.0036	0.0030	0.0029	0.0029
	0.0038	0.0036	0.0038	0.0031	0.0028	0.0029
0.17	0.0030	0.0029	0.0029	0.0024	0.0023	0.0025
	0.0752	0.0765	0.0763	0.0745	0.0741	0.0752
	0.0590	0.0633	0.0697	0.0574	0.0609	0.0667

Note. SPB = mean squared parameter bias; ML = maximum likelihood; GLS = generalized least squares; WLS = weighted least squares; FRMR = fixed root mean squared residual.
[a]$n = 1{,}000$. [b]$n = 2{,}000$.