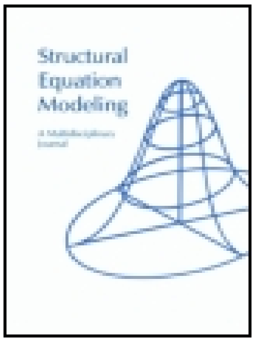
Effects of Sample Size and Nonnormality on the Estimation of Mediated Effects in Latent Variable Models

Article · January 1997

DOI: 10.1080/10705519709540063

David MacKinnon
Arizona State University



Routledge
Taylor & Francis Group

Structural Equation Modeling: A Multidisciplinary Journal

ISSN: 1070-5511 (Print) 1532-8007 (Online) Journal homepage: http://www.tandfonline.com/loi/hsem20

Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models

John F. Finch , Stephen G. West & David P. MacKinnon

STRUCTURAL EQUATION MODELING, 4(2), 87–107

Effects of Sample Size and Nonnormality on the Estimation of Mediated Effects in Latent Variable Models

John F. Finch
Department of Psychology
Texas A&M University

Stephen G. West and David P. MacKinnon
Department of Psychology
Arizona State University

A Monte Carlo approach was used to examine bias in the estimation of indirect effects and their associated standard errors. In the simulation design, (a) sample size, (b) the level of nonnormality characterizing the data, (c) the population values of the model parameters, and (d) the type of estimator were systematically varied. Estimates of model parameters were generally unaffected by either nonnormality or small sample size. Under severely nonnormal conditions, normal theory maximum likelihood estimates of the standard error of the mediated effect exhibited less bias (approximately 10% to 20% too small) compared to the standard errors of the structural regression coefficients (20% to 45% too small). Asymptotically distribution free standard errors of both the mediated effect and the structural parameters were substantially affected by sample size, but not nonnormality. Robust standard errors consistently yielded the most accurate estimates of sampling variability.

In structural equation models, causal relations among variables can be divided into two types. *Direct* effects involve only direct connections between variables, whereas *indirect* (mediated) effects are transmitted via one or more intervening variables (Alwin & Hauser, 1975; Duncan, 1975). Consideration of both types of effects is important both in theory testing research (e.g., Bentler & Speckart, 1981; Fiske, Kenny, & Taylor, 1982) and applied research that attempts to explain the

Requests for reprints should be sent to John F. Finch, Department of Psychology, Texas A&M University, College Station, TX 77843–4235. E-mail: jff@psyc.tamu.edu

mechanisms through which treatment effects operate (MacKinnon, 1994; Mac-Kinnon et al., 1991; West, Aiken, & Todd, 1993). The failure to consider both types of effects can obscure the true nature of a causal process and lead to incorrect causal inferences. Despite the importance of mediational analyses, which partition the total effect into direct and indirect effects, few studies have formally tested mediated effects. This has occurred even though several discussions in the psychological research literature have highlighted the importance of mediational analysis (Baron & Kenny, 1986; James & Brett, 1984; Judd & Kenny, 1981; MacKinnon & Dwyer, 1993).

The lack of attention to mediated effects may be due, in part, to the relatively recent development of statistical procedures for testing mediation. Historically, potential mediated effects were either ignored or were discussed without an explicit test of statistical significance (MacKinnon & Dwyer, 1993). More recently, Sobel (1982, 1986) developed a large sample test of mediation using the multivariate delta method to derive the standard error of the indirect effect. The multivariate delta standard error (Sobel 1982, 1986), however, is based on asymptotic statistical theory and assumes multivariate normality of the observed variables (see Rice, 1988, pp. 142–147, for a general presentation of the delta method). Guidelines concerning the minimum sample size required for application of the technique are only now beginning to appear. Under multivariate normality, Stone and Sobel (1990) suggested that a sample size of at least 400 is required for accurate estimation of the standard errors of mediated effects in large nonrecursive latent variable (LV) models. For a manifest variable path model with six observed variables, Stone and Sobel (1990) suggested a sample size of at least 200, although simulation studies by MacKinnon and his colleagues (MacKinnon & Dwyer, 1993; MacKinnon, Warsi, & Dwyer, 1995) suggested that the sample size requirements for some manifest variable models may be more modest. Using a manifest variable model with three observed variables and multivariate normal data, MacKinnon and Dwyer (1993) examined the accuracy of several different formulas for estimating indirect effect standard errors, including the multivariate delta method (Sobel, 1982). These investigators found little bias and close agreement across estimation methods at sample sizes as small as 50. Sample size requirements for the accurate estimation of the standard errors of mediated effects in LV models are less well understood. Furthermore, it remains to be determined whether indirect effect standard errors in LV models are accurate in the presence of nonnormal data.

This study extends previous research by Mackinnon and his colleagues (MacKinnon & Dwyer, 1993; MacKinnon et al., 1995) and Stone and Sobel (1990) in three important respects. In this study, we examine indirect effect estimation in the context of (a) nonnormally distributed data, (b) a LV model, and (c) multiple methods of estimation. Many studies based on confirmatory factor models have shown maximum likelihood (ML) standard errors to be negatively biased when the data are nonnormal (see West, Finch, & Curran, 1995, for a review). To date, no studies have explicitly examined bias in the standard errors of mediated effects in

LV models as a result of nonnormality and small sample size. Here, we compare multiple methods of estimation and systematically vary characteristics of the model and the data to determine the extent to which the estimated standard errors of mediated effects are adversely affected.

ESTIMATING INDIRECT EFFECTS

Figure 1 illustrates a basic mediational model involving both a direct and an indirect effect. In this model, the effect of the independent variable (η_1) on the final outcome variable (η_3) can be decomposed into a direct effect and an indirect effect through a mediating variable (η_2). For the basic model presented in Figure 1, the value of the indirect effect is estimated as b_1b_2. For this model, Sobel's (1982, 1986) asymptotic standard error[1] of the indirect effect based on the multivariate delta method is equal to $\sqrt{b_1^2\sigma_{b2}^2 + b_2^2\sigma_{b1}^2}$.

[1]The exact variance of the product of two independent random variables is discussed in several mathematical statistics texts (e.g., Rice, 1988), and for the b_1b_2 product this variance equals $b_1^2\sigma_{b2}^2 + b_2^2\sigma_{b1}^2 + \sigma_{b1}^2\sigma_{b2}^2$. Sobel's (1982, 1986) variance based on the multivariate delta method omits the typically negligible $\sigma_{b1}^2\sigma_{b2}^2$ term.

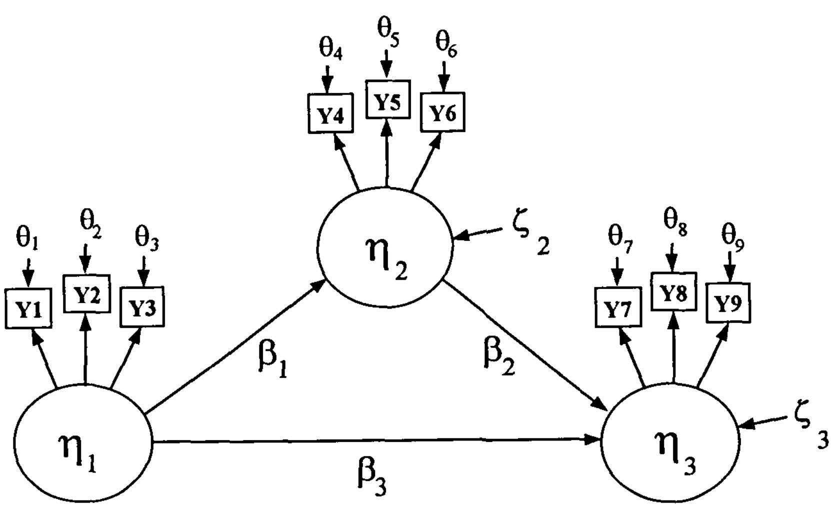


FIGURE 1 Latent variable model used in simulations. Consistent with LISREL all-y notation, latent constructs are denoted by η and are enclosed in circles, manifest variables (Y_i) are enclosed in boxes, errors in equations are denoted by ζ, and errors of measurement in the observed variables are denoted by θ.

STATISTICAL THEORY

Normal theory estimation procedures, notably ML and generalized least squares (GLS), are currently the most widely used techniques for estimating the parameters of LV structural equation models. Assuming a correctly specified model, ML and GLS produce similar parameter estimates and standard errors in large samples. Both procedures, however, operate under the assumption that the manifest variables are multinormally distributed. Because real data often violate this assumption (Micceri, 1989), there has been growing interest in determining the robustness of normal theory estimators. Studies examining bias in ML parameter estimates have shown ML to be accurate at sample sizes of 100 under conditions of multivariate normality (Boomsma, 1982) and to exhibit negligible bias even under marked nonnormality. In contrast, simulation studies suggest that ML and GLS χ^2 test statistics are substantially overestimated, and ML and GLS standard errors are substantially attenuated when the data are nonnormal (e.g., Harlow, 1985; see West et al., 1995, for a review).

Two different approaches are most frequently used by researchers attempting to address these problems with normal theory estimation. The first approach is Browne's (1982, 1984) asymptotically distribution free (ADF) estimation method. Browne has demonstrated that by using higher order moment information, a distribution free estimator can be developed that possesses the same desirable asymptotic properties as ML and GLS, but requires milder distributional assumptions. The ADF estimator, however, has been shown to be affected by both model size and sample size (Hu, Bentler, & Kano, 1992; Muthén & Kaplan, 1992). Because ADF estimation requires the storage and inversion of a $p^* \times p^*$ weight matrix at each iteration (where p^* represents the $p(p + 1)/2$ nonredundant elements in the covariance matrix of the p observed variables), moderate to large models become computationally problematic. In addition, large sample sizes are required to obtain stable estimates of the higher order elements contained in the ADF weight matrix (see Appendix).

The second approach is to adjust the normal theory χ^2 and standard errors for the presence of nonzero kurtosis. Although the normal theory χ^2 test statistic does not follow the expected χ^2 distribution under conditions of nonnormality, this test statistic may be corrected or rescaled to more closely approximate the referenced χ^2 distribution (Browne, 1982, 1984). Satorra and Bentler (1988; Satorra, 1990, 1991) developed a variant of this rescaled test statistic that is currently implemented in the EQS program (Bentler, 1989). This rescaled χ^2 test statistic appears to be less sensitive to model size and sample size than the χ^2 produced by the ADF estimator (Chou, Bentler, & Satorra, 1991; Hu et al., 1992). However, the adjustment procedure is theoretically not as asymptotically efficient as the ADF estimator.

With respect to standard errors, a correction to normal theory standard errors in the form of a robust covariance matrix of the parameter estimates has also been proposed by Browne (1982, 1984) and a variant of this correction procedure (Bentler, 1983; Bentler & Dijkstra, 1985; Satorra, 1990, 1991) is currently available in EQS. This robust covariance matrix, which allows for deviations from multinormality in the manifest variables, may be employed in conjunction with normal theory ML. Robust standard error estimates, which are theoretically valid in large samples even in the presence of nonzero kurtosis, can be computed from this robust matrix as the square root of the diagonal elements divided by n.

Studies comparing ML, ML–robust (ML–r), and ADF standard errors in confirmatory factor analysis (CFA) models have shown ML standard errors to be accurate when data are multinormal, but biased when the data are nonnormal (Chou et al., 1991). ADF, by contrast, yields correct standard errors, but only at large sample sizes (Hu et al., 1992; Yung & Bentler, 1994).

STUDY 1

Study 1 focuses on the robustness of ML and ADF parameter estimates of direct and indirect effects and their associated standard errors varying (a) sample size, (b) the population values for the model parameters, and (c) the degree of nonnormality characterizing the data. In addition to normal theory ML standard errors, ML–r standard errors were also examined to determine the range of nonnormality conditions under which these standard errors are accurate.

Method

Model Specification

The model considered in this study was comprised of three latent factors with three manifest indicators per factor (see Figure 1). Four sets of parameter values, four sample sizes (150, 250, 500, 1,000), and three distributional conditions were considered. Following the precedent of Hu et al. (1992), 200 replications were run in each of the 48 conditions.

The first distributional condition corresponded to the ideal situation in which all measured variables were specified to be normally distributed in the population (all univariate skewness and kurtosis[2] coefficients equal to 0). The second and third distributional conditions reflected moderate (skewness = 2 and kurtosis = 7) and

[2]Following the convention of a number of statistical software packages, the univariate kurtosis coefficient in EQS is rescaled so that 0 corresponds to that of the normal distribution. Positive values reflect leptokurtic distributions and negative values reflect platykurtic distributions.

extreme (skewness = 3 and kurtosis = 21) departures from multinormality in the population. Univariate frequency distributions (based on 10,000 observations) corresponding to the three distributional conditions created in this study are presented in Figure 2. These levels of skewness and kurtosis were chosen to represent normal, moderately nonnormal, and severely nonnormal distributions based on the examination of the distributions in several large community data sets.

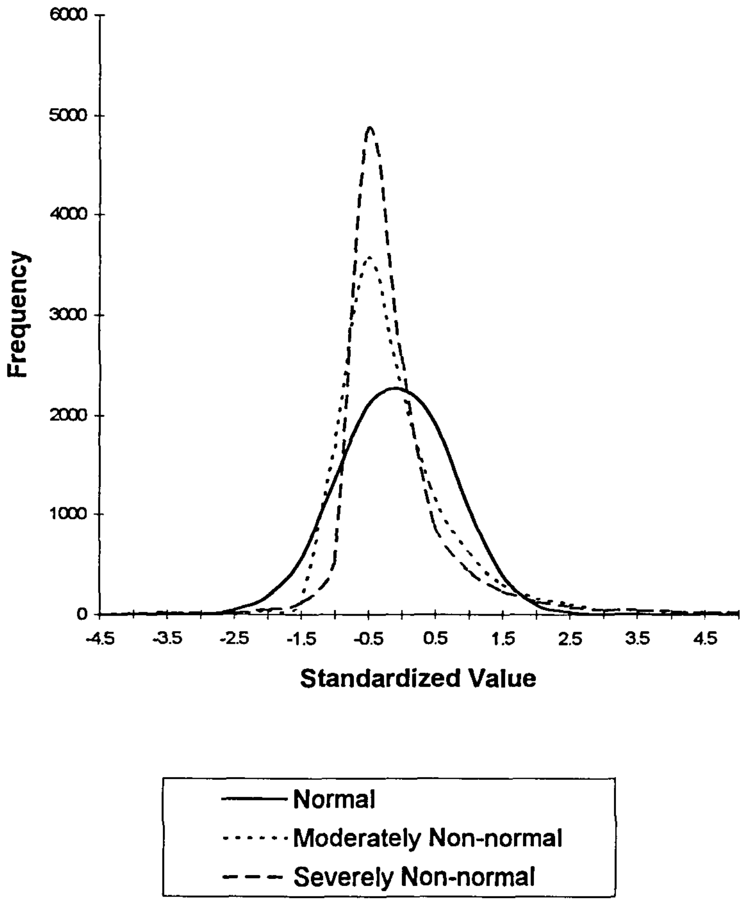


FIGURE 2 Univariate frequency distributions corresponding to the normal (skew = 0, kurtosis = 0), moderately nonnormal (skew = 2, kurtosis = 7), and severely nonnormal (skew = 3, kurtosis = 21) distributional conditions created in Study 1. Each distribution was generated in EQS based on 10,000 standardized observations. From "The Robustness of Test Statistics to Nonnormality and Specification Error in Confirmatory Factor Analysis," by P. J. Curran, S. G. West, and J. F. Finch, 1996, *Psychological Methods, 1,* pp. 16–29. Copyright 1996 by the American Psychological Association. Reprinted by permission of the publisher.

Specification 1. In the first model specification, the population values for the factor loadings were uniform and moderately high (standardized loadings = .70), whereas all error variances (θs) were specified to be .51. The population values for the structural regression coefficients were specified to be $\beta_1 = .60$, $\beta_2 = .20$, and $\beta_3 = .12$. In this specification, the magnitude of the direct effect of η_1 (latent exogenous variable) on η_3 (latent endogenous variable) was specified to be equal to the magnitude of the indirect effect of η_1 on η_3 via η_2.

Specification 2. Specification 2 differed from Specification 1 only in that the population values for the structural regression (path) coefficients were specified to be $\beta_1 = .30$, $\beta_2 = .40$, and $\beta_3 = .12$. The magnitude of the direct effect of η_1 on η_3 was again specified equal the magnitude of the indirect effect of η_1 on η_3 via η_2.

Specification 3. Specification 3 differed from Specifications 1 and 2 only in that the population values for the structural regression coefficients were specified to be $\beta_1 = .30$, $\beta_2 = .40$, and $\beta_3 = .36$. In this specification, the magnitude of the direct effect of η_1 on η_3 was specified to be three times as large as the indirect effect of η_1 on η_3 via η_2.

Specification 4. In Specification 4, the population values for the structural regression coefficients were again specified to be $\beta_1 = .30$, $\beta_2 = .40$, and $\beta_3 = .36$, and the population values of the factor loadings were varied. Specifically, the population value of the loading of the first manifest variable was specified as .40, the second loading was specified as .60, and the third loading was specified as .80 for each of the three LVs.

Data Generation

When the simulation option in EQS is selected, the program generates raw data with user-specified values of population skewness and kurtosis based on formulas developed by Fleischman (1978) and using procedures described by Vale and Maurelli (1983). EQS uses a random number generator based on work by Lewis, Goodman, and Miller (1969). In this study, for each of the four model specifications, the raw data were generated based on the covariance matrix implied by the model parameters in the following manner. After specifying population values for the 21 free parameters in the three-factor model under investigation, the population covariance matrix implied by these values was computed. This model-implied matrix was then used as the population covariance matrix, and simple random samples of 150, 250, 500, and 1,000 were drawn repeatedly by EQS based on this population matrix.

Method of Evaluating the Results

To facilitate comparison across the different conditions in the design, relative bias in the parameter estimates and their standard errors was computed. For each model parameter in each of the 48 conditions, relative bias was computed by subtracting the true parameter value from the mean of the 200 parameter estimates computed by EQS and dividing this difference by the true value of the parameter. Nonzero relative bias values indicated the degree to which parameter estimates failed to approximate their true population values. In this study, an empirical standard error was computed as the standard deviation of the set of 200 estimates of a single parameter. For each model parameter, relative bias in the estimated standard error was computed by subtracting the empirical standard error from the mean of the 200 estimated standard errors computed by EQS and dividing this difference by the empirical standard error.

Results

Parameter Estimates

For Specifications 1–4, negligible effects of sample size and nonnormality were observed on relative bias in the structural coefficients and indirect effect estimates for both ML or ADF estimation. In all four specifications, relative bias in the parameter estimates ranged between –4% and +4%, seldom exceeding ±3%.

Standard Errors of Parameter Estimates

Normal theory ML. For Specifications 1–4, significant effects of nonnormality on relative bias in the indirect effect standard errors were observed. Larger sample sizes failed to result in any appreciable decrease in the relative bias of the ML indirect effect standard errors. Table 1 presents the full results for Specification 1. Because the results for Specifications 2–4 were very similar to those for Specification 1, they will not be reported later.

Inspection of Table 1 reveals that, under multivariate normality, relative bias in the indirect effect standard errors estimated via ML was negligible (M = 1.32%, range = –6.76% to 9.27%). Using Kaplan's (1988) criterion of 10% relative bias as representing potentially meaningful levels of bias, none of the estimates exceeds

TABLE 1
Standard Errors and Percentage Bias of Estimates for Specification 1

	Sample Size											
Parameter	*150*			*250*			*500*			*1,000*		
Normal												
b_1b_2	.1064	0.96%	0.09%	.0848	−6.76%	−8.31%	.0524	1.82%	1.11%	.0341	9.27%	3.33%
	.1341	−28.08%		.0815	−10.02%		.0590	−10.18%		.0387	−4.70%	
b_1	.1404	−5.16%	−6.74%	.0982	2.19%	1.63%	.0749	−6.21%	−6.06%	.0495	0.63%	0.40%
	.1473	−23.32%		.1191	−21.10%		.0747	−7.98%		.0479	1.81%	
b_2	.1579	2.30%	6.56%	.1400	−12.67%	−6.27%	.0875	−2.71%	−3.49%	.0558	6.42%	1.17%
	.2194	−29.48%		.1312	−11.61%		.0945	−12.49%		.0622	−5.10%	
b_3	.1758	−9.47%	−11.12%	.1345	−7.80%	−9.41%	.0924	−6.46%	−6.82%	.0570	−4.90%	5.26%
	.2334	−36.17%		.1317	−16.43%		.0970	−13.99%		.0555	5.22%	
Moderately nonnormal												
b_1b_2	.1200	−6.49%	0.77%	.0903	−8.06%	−0.09%	.0609	−9.44%	−0.83%	.0435	−6.23%	−5.95%
	.1802	−26.30%		.1016	−16.44%		.0683	−12.84%		.0387	−4.38%	
b_1	.2401	−34.82%	−17.91%	.1353	−24.53%	−0.99%	.0945	−25.74%	−0.30%	.0720	−31.72%	−6.52%
	.2200	−31.93%		.1469	−18.51%		.1315	−20.20%		.0592	−4.67%	
b_2	.1945	−17.71%	−14.02%	.1398	−10.45%	−5.55%	.1018	−15.73%	−10.04%	.0747	−21.54%	−3.84%
	.2581	−30.90%		.1721	−20.93%		.1137	−15.86%		.0581	−3.71%	
b_3	.1804	−9.46%	−1.81%	.1386	−9.01%	−0.00%	.0964	−8.79%	1.87%	.0733	−8.34%	−0.91%
	.2661	−32.06%		.1566	−17.42%		.1043	−14.54%		.0590	−2.32%	
Severely nonnormal												
b_1b_2	.1711	−23.42%	−14.71%	.1103	−21.44%	−9.95%	.0737	−24.66%	−9.60%	.0494	−22.03%	−3.84%
	.1601	−27.86%		.1099	−22.93%		.0642	−6.90%		.0487	4.38%	
b_1	.2377	−48.92%	−25.79%	.2130	−51.70%	−23.10%	.1281	−45.30%	−7.58%	.1031	−51.48%	−11.86%
	.2724	−32.39%		.2247	−38.65%		.1258	−12.06%		.0684	−4.67%	
b_2	.2135	−34.60%	−28.95%	.1774	−29.39%	−23.37%	.1126	−28.26%	−18.36%	.0835	−28.06%	−16.12%
	.2632	−18.92%		.1653	−19.97%		.1031	−10.76%		.0660	−0.47%	
b_3	.2257	−25.38%	−13.81%	.1570	−19.04%	−7.89%	.1191	−22.57%	−4.82%	.0777	−22.00%	−4.49%
	.2465	−33.56%		.1809	−26.61%		.1111	−17.24%		.0678	−6.31%	

Note. For each parameter at each sample size, the entries on the first line are, in order: the observed (empirical) standard error of the maximum likelihood (ML) parameter estimate, the percentage relative bias in the mean ML standard error estimate, and the percentage relative bias in the mean ML–r standard error estimate. The entries on the second line are, in order: the observed (empirical) standard error and the asymptotically distribution free (ADF) parameter estimate and the percentage relative bias in the mean ADF standard error estimate. All entries are based on 200 replications.

this threshold. In contrast, under severe nonnormality, ML estimates of the standard error of the indirect effect underestimated the empirical standard errors by an average of 22.89% (range = –21.44% to –24.66%).

For purposes of comparison, Table 1 also presents the ML standard errors for the three structural coefficients. Consistent with the results of previous Monte Carlo research using CFA models, relative bias in the standard errors of the direct effect structural coefficients was negligible under multivariate normality. These estimates became increasingly negatively biased as the manifest variables became increasingly nonnormal. Larger population values of the structural regression coefficients also appeared to be associated with greater downward bias in the ML standard errors under moderate and severe nonnormality.

Robust ML. For ML–r, weaker effects of nonnormality on the standard errors were observed (see Table 1). With regard to the robust estimates of the standard error of the indirect effect, there was minimal bias under normality (M = –0.95%, range = –8.31% to +3.33%), which increased under severe nonnormality (M = –9.53%, range = –3.84% to –14.71%). Under normality, the mean relative bias in the robust standard errors of the structural regression coefficients was –2.91% (range = –11.12% to +6.56%). Under severe nonnormality, the robust estimates of the standard errors of the structural regression coefficients exhibited an average relative bias of –15.51% across conditions. Weak effects of sample size were also observed, with some decrease in relative bias associated with larger sample sizes.

ADF. For ADF, practically significant effects of sample size, but not nonnormality, were found on relative bias in the estimated standard error of the indirect effect (see Table 1). The ADF estimator consistently became more accurate as sample size increased. At n = 150, the ADF standard errors of the direct and mediated effects had a mean bias of –29.86% (range = –18.92% to –36.17%) and –27.41% (range = –26.30% to –28.08%), respectively. Relative bias in the estimates of the standard errors became negligible for structural regression coefficients (M = –2.25%) and indirect effects (M = –1.57%) at n = 1,000.

Discussion

The results of Study 1 are consistent with existing CFA studies (e.g., Arminger & Schoenberg, 1989; Chou et al., 1991; Curran, West, & Finch, in press) in illustrating the adverse effects of nonnormality on the accuracy of significance tests in LV models estimated using normal theory methods. Although these earlier results were based exclusively on CFA models, this study found similar levels of bias in the estimated standard errors of both the direct and indirect effect estimates in a LV structural model.

In general, normal theory ML standard errors were too small when the normality assumption was violated. By contrast, ADF standard error estimates were unaffected by the distributional characteristics of the variables, but were substantially negatively biased in small samples. The practical effect of negatively biased standard errors would be rejection of the null hypothesis too frequently. Under nonnormality, the robust standard errors performed much better at all sample sizes in the present simulation. For all three methods of estimating standard errors, the magnitude of the observed bias varied little across differing ratios of direct to indirect effects. The pattern of bias in the standard errors of direct and indirect effects was also not influenced by variation in the population values of the factor loadings.

STUDY 2

The findings of Study 1 were all based on models in which the manifest variables were identically distributed in the population. Study 2 extended the generality of the findings by examining bias under conditions in which the degree of nonnormality differed across manifest variables. Based on an examination of several data sets from large substance abuse and mental health studies, three distributional conditions were constructed that approximated conditions observed in practice. The population values of skewness and kurtosis specified in these distributional conditions are described in Table 2. In the first distributional condition, the nine measured

TABLE 2
Population Skewness and Kurtosis Values for the Manifest Variables in the Three Distributional Conditions in Study 2

	Distributional Condition 1		*Distributional Condition 2*		*Distributional Condition 3*	
	Skewness	*Kurtosis*	*Skewness*	*Kurtosis*	*Skewness*	*Kurtosis*
Y1	0.33	1.00	1.00	6.00	0.00	0.00
Y2	0.66	2.00	1.50	9.00	0.00	0.00
Y3	1.00	3.00	2.00	12.00	0.00	0.00
Y4	0.50	1.50	1.00	3.00	0.00	0.00
Y5	1.00	3.00	2.00	7.00	0.00	0.00
Y6	1.50	4.50	3.00	21.00	0.00	0.00
Y7	1.00	3.00	0.75	2.00	3.00	21.00
Y8	1.50	4.50	1.50	4.00	3.00	21.00
Y9	2.00	6.00	2.25	16.00	3.00	21.00

variables were mildly to moderately nonnormal, in the second distributional condition the nine manifest variables were moderately to severely nonnormal, and in the third distributional condition the measured indicators of the final outcome construct were severely nonnormal, whereas the other six variables were normally distributed.

Method

Model Specification

Like the models in Study 1, the model considered in Study 2 was comprised of three latent factors with three manifest indicators per factor. The factor loadings were again uniformly .70 and population values for the structural regression coefficients were specified to be $\beta_1 = .30$, $\beta_2 = .40$, and $\beta_3 = .36$. As before, four sample sizes were considered: 150, 250, 500, and 1,000. In addition, the level of nonnormality characterizing each manifest variable was systematically varied. Because the distributions of the nine manifest variables differed within each distributional condition, 400 replications were conducted for each combination of sample size and distributional condition.

Results

Parameter Estimates

No practically significant effects of either sample size or nonnormality were observed on relative bias in the structural coefficients or the indirect effect estimates for either ML or ADF estimation. Relative bias in the parameter estimates for both the direct and indirect effects ranged between –4% and +4%, seldom exceeding ±3%.

Standard Errors of Parameter Estimates

Under the first distributional condition, there were no appreciable effects of sample size on the estimated standard errors of the indirect effect using normal theory ML or ML–r (see Table 3). Bias in the indirect effect standard error estimates was negligible, ranging from –3.37% to +0.91% (M = –0.81%) for the ML estimates and from –1.26% to +5.73% (M = 2.44%) for the ML–r estimates. Modest levels of relative bias were found for the normal theory ML standard errors of the structural coefficients (M = –8.65%) and the corresponding ML–r standard errors (M =

TABLE 3
Standard Errors and Percentage Bias of Estimates in Study 2

Parameter	*Sample Size* 150			250			500			1,000		
Distribution 1: Moderate, heterogenous nonnormality												
b_1b_2	.0627	−1.07%	−1.26%	.0493	−3.37%	−0.0%	.0319	0.91%	5.30%	.0224	0.31%	5.73%
	.0774	−28.46%		.0517	−16.88%		.0324	−0.84%		.0229	1.31%	
b_1	.1310	−10.24%	−9.12%	.1017	−10.96%	−8.65%	.0677	−6.62%	−4.21%	.0460	−2.94%	0.54%
	.1573	−31.21%		.1065	−19.10%		.0713	−10.88%		.0483	−4.98%	
b_2	.1305	−7.97%	−4.81%	.0991	−7.88%	−3.48%	.0723	−12.65%	−6.54%	.0524	−13.78%	−6.82%
	.1690	−30.73%		.1160	−21.29%		.0739	−7.69%		.0519	−3.35%	
b_3	.1323	−9.03%	−3.76%	.0999	−5.75%	2.88%	.0685	−5.63%	−3.09%	.0513	−10.29%	−0.15%
	.1602	−30.81%		.1135	−22.11%		.0763	−13.38%		.0492	−5.31%	
Distribution 2: Severe, heterogenous nonnormality												
b_1b_2	.0697	−7.86%	−1.06%	.0520	−7.72%	0.08%	.0368	−12.47%	−0.81%	.0237	−5.18%	8.16%
	.0821	−23.67%		.0543	−11.73%		.0333	0.20%		.0238	2.88%	
b_1	.1416	−17.16%	−12.61%	.1086	−17.23%	−10.97%	.0730	−13.91%	−4.48%	.0521	−14.57%	−3.91%
	.1782	−33.28%		.1116	−16.06%		.0705	−5.64%		.0534	−9.56%	
b_2	.1497	−19.49%	−12.33%	.1121	−18.35%	−9.64%	.0848	−24.71%	−13.19%	.0514	−12.66%	1.89%
	.1668	−25.60%		.1239	−20.18%		.0814	−12.31%		.0546	−4.44%	
b_3	.1498	−17.62%	−10.61%	.1135	−17.67%	−5.88%	.0737	−10.87%	−4.04%	.0511	−9.88%	6.65%
	.1909	−32.56%		.1167	−16.22%		.0793	−11.93%		.0517	−2.66%	
Distribution 3: Outcome severely nonnormal												
b_1b_2	.0639	−34.61%	−33.06%	.0536	−11.74%	−8.27%	.0314	−2.35%	−7.02%	.0234	−4.06%	1.48%
	.0862	−44.94%		.0434	−10.98%		.0306	−3.29%		.0211	−2.24%	
b_1	.1207	−2.35%	−4.00%	.0867	−4.73%	−4.05%	.0645	−1.87%	−2.41%	.0456	−1.64%	−1.78%
	.1535	−30.09%		.1008	−16.39%		.0726	−14.84%		.0447	−2.10%	
b_2	.1433	−55.11%	−52.51%	.1168	−21.34%	−14.18%	.0716	−11.41%	−0.99%	.0520	−13.19%	−0.57%
	.1763	−32.67%		.0998	−18.74%		.0640	−1.17%		.0523	−9.94%	
b_3	.1527	−47.23%	−42.01%	.1003	−17.67%	2.33%	.0691	−5.95%	7.12%	.0538	−14.82%	−1.66%
	.1510	−38.22%		.1052	−27.16%		.0736	−16.16%		.0488	−6.32%	

Note. For each parameter at each sample size, the entries on the first line are, in order: the observed (empirical) standard error of the maximum likelihood (ML) parameter estimate, the percentage relative bias in the mean ML standard error estimate, and the percentage relative bias in the mean ML–r standard error estimate. The entries on the second line are, in order: the observed (empirical) standard error of the asymptotically distribution free (ADF) parameter estimate and the percentage relative bias in the mean ADF standard error estimate. All entries are based on 400 replications.

–3.93%). Relative bias in the ADF standard errors decreased with increasing sample size, ranging from approximately –30% at n = 150 to approximately –3% at n = 1,000.

Under the second distributional condition, no consistent effects of sample size were observed for normal theory ML estimates of the standard error of the indirect effect. Normal theory ML estimates of the standard errors of the structural regression coefficients were negatively biased by approximately 10% to 20% at all sample sizes. For ML–r, estimates of the standard errors of the structural coefficients showed a general tendency to become more accurate with increasing sample size. For ADF, the magnitude of bias in the direct and indirect effects was similar to that observed in the first distributional condition.

Under the third distributional condition, there were large effects of nonnormality on the ML and ML–r standard errors of both b_2 and b_3 at n = 150. These effects tended to decrease as sample size became larger. By comparison, the standard error of the path from the exogenous construct (η_1) to the mediator construct (η_2), both of which had normally distributed indicator variables, was essentially unbiased at all sample sizes. Recall that only the outcome construct in this distributional condition had dramatically nonnormally distributed indicator variables. Like the bias observed in the standard errors of b_2 and b_3, the standard error of the indirect effect estimated via ML exhibited substantial negative bias (–34.61%) at n = 150. This bias decreased to –4.06% at n = 1,000. As can be seen in Table 3, the magnitude of ADF bias in the direct and indirect effects was very similar to that observed in distributional conditions one and two. With ADF, there were appreciable effects of sample size on the standard errors of the structural regression coefficients.

Discussion

The results of this study illustrate the adverse effects of nonnormality on the accuracy of significance tests in LV models estimated using normal theory ML. As found in previous CFA studies, normal theory ML estimates of the standard errors tended to be substantially negatively biased when the normality assumption was violated. ADF estimated standard errors of these coefficients also tended to be markedly negatively biased when the sample size was 500 or less. The ML–r standard errors of both the direct and indirect effects were comparable to those obtained using normal theory ML under conditions of multivariate normality. However, the ML–r standard errors provided more accurate estimates of sampling variability than the normal theory standard errors as nonnormality increased. The ML–r standard errors were also more accurate than the ADF standard errors under all distributional conditions at the smaller sample sizes (i.e., n = 150, n = 250) considered in this study.

The results of Study 2 suggest that when the distributions of the measured variables differ markedly, the pattern of nonnormality can influence the magnitude of bias in the estimated standard errors. In the third distributional condition in Study 2, the two standard errors that were most dramatically biased were those associated with the structural coefficients b_2 and b_3 (see Figure 1). Both of these coefficients reflected effects on the outcome construct (η_3), which in the third distributional condition had dramatically nonnormally distributed indicator variables. By comparison, the standard error of the path from the exogenous construct to the mediator construct, both of which had normally distributed indicator variables, was essentially unbiased at all sample sizes.

The general pattern of bias observed for the estimated standard error of the mediated effect was similar to that observed for the direct effects. Although Stone and Sobel (1990) suggested that a sample size of at least 400 is required for accurate estimation of the variance of indirect effects in LV models, the results of this study suggest that bias in the standard error of the indirect effect is negligible even at the smallest sample sizes examined when multivariate normality is present. This replicates the findings of MacKinnon and Dwyer (1993) and MacKinnon et al. (1995) for manifest variable models. The results also suggest that the estimated standard error of the indirect effect is generally more robust to nonnormality than the standard error of the direct effect. Using ML, bias in the former was typically negligible under mild to moderate violations of the multinormality assumption. The worst ML bias in the indirect effect standard error estimates (approximately –25%) occurred under severely nonnormal conditions (in Study 1) and when the outcome construct was measured with severely nonnormal variables in small samples (in Study 2).

In this study, the relative bias in the estimated standard error of the mediated effect was typically smaller than the bias in the standard errors of the constituent effects of which it was comprised. One explanation for this result is that estimates of the structural paths are included in the formula for the standard error of the indirect effect. Nonnormality does not affect these constituent parameter estimates, reducing the overall impact of nonnormality on the standard error of the indirect effect.

The minimal distributional assumptions on which the ADF method is based should make it the method of choice for models with nonnormally distributed variables. Although ADF yielded accurate estimated standard errors at the largest sample size considered (n = 1,000), the ADF standard errors exhibited substantial negative bias at smaller sample sizes. To understand the small sample breakdown of ADF, it is useful to compare ADF and normal theory approaches to estimation.

ADF and normal theory standard errors both require the estimation of a weight matrix; however, the elements of the ADF weight matrix are based on the computation of fourth-order moments. Normal theory ML and GLS, by contrast, employ a weight matrix comprised of elements that can be expressed as products of more stable second-order moments. Because the estimates of the elements in the ADF and normal theory weight matrices will reflect the sampling variability of the terms of which they are comprised, their variability could be expected to differ. Under multivariate normality, the elements of the ADF weight matrix converge in probability to products of second-order moments as $n \rightarrow \infty$. Consequently, ADF estimators and normal theory GLS estimators are asymptotically equivalent under normality assumptions. In small samples, however, the variability of the ADF weight matrix elements is likely substantially larger than the variability of elements calculated using either ML or GLS. This is because stable estimates of the higher order moments used in ADF require large sample sizes. An appreciable reduction in sampling variability is attained when the weight matrix elements can be expressed as a function of more stable second-order moments, as is the case with normal theory ML and GLS estimators under multivariate normality. These differences in the sampling variability of the weight matrix elements would explain the relatively poor performance of the ADF estimator that was observed in this study under normality at sample sizes below 500 (see also Yung & Bentler, 1994).

The reason for the superior performance of the robust standard errors over the ADF standard errors may also be due to differences in the stability of the weight matrices employed by the two methods. The calculation of standard errors in both estimation methods involves the computation of a weight matrix of fourth-order moments. However, the robust standard errors employ this matrix directly, whereas ADF standard errors require also that this fourth-order moment matrix be stored and inverted. The variability of the ADF inverse estimated weight matrix is likely excessive in small to moderate samples and there may be accuracy problems associated with its computation. Because inversion of the fourth-order moment matrix is not required under the robust approach, robust standard errors would be expected to exhibit less small sample instability.

LIMITATIONS

Although this study evaluated the empirical behavior of parameter estimates under a range of distributional conditions, generalizations from the findings of Monte Carlo studies are limited by practical considerations related to the design of the simulation. In this study, only four combinations of parameter values and four sample sizes were considered. The six distributional conditions created represent only a tiny fraction of the universe of multivariate distributions. Nonetheless, care was taken to choose realistic values of the parameters and distributions investigated, at least in the context of community research in which the authors have their greatest experience.

Three limitations of this study are worthy of note. First, this study examined a relatively simple indirect effects model. The question of whether similar results would be obtained with more complex indirect paths involving multiple mediators has not been addressed and represents a topic for further research. Second, we did not address the situation in which indicator variables are skewed in opposite directions. Although rare in our experience (assuming that indicator variables have all been recoded to be oriented in the same direction), this situation may be theoretically important. Olsson (1979) reported the worst bias in CFA parameter estimates and the most dramatic inflation in χ^2 when indicator variables had only a small number of scale points and were skewed in opposite directions. It is possible that the relatively negligible bias in parameter estimates observed in this simulation might have been much more dramatic had distributional conditions involving oppositely skewed indicators been included in the simulation design. On the other hand, this problem appears to be far less serious when variables are measured with more rather than fewer scale points. Third, the present simulations examined only distributions involving positive values of kurtosis. Statistical theory has indicated that the bias in ML or GLS standard errors may be particularly marked when kurtosis is negative. However, the research to date investigating the effects of such distributions with CFA models (Chou et al., 1991; Hu et al., 1992) is largely consistent with the results of this study.

CONCLUSION

The results of this study replicate and extend the results of previous studies examining the impact of nonnormality and sample size on parameter estimation in structural equation models. When measured variables have skewness and kurtosis that differs from that of a normal distribution, the standard errors generated by normal theory ML and GLS are likely to be too small. The accuracy of ML and GLS parameter estimates, in contrast, is generally unaffected by departures from multinormality. In this study, indirect effect standard errors showed less bias than direct effect standard errors under nonnormality. In addition, the magnitude of bias in the standard errors of structural coefficients depended on the degree of distributional misspecification of the indicators of each factor. Standard error estimates for structural coefficients relating factors with normally distributed indicators remained unbiased.

ACKNOWLEDGMENTS

This research was funded in part by National Institute of Mental Health Grant P50–MH39246 and Public Health Service Grant DA 06211.

REFERENCES

Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review, 40,* 37–47.

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

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Bentler, P. M. (1983). Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. *Psychometrika, 48,* 493–517.

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

Bentler, P. M., & Dijkstra, T. (1985). Efficient estimation via linearization in structural models. In P. R. Krishnaiah (Ed.), *Multivariate analysis* (Vol. VI, pp. 9–42). Amsterdam: North-Holland.

Bentler, P. M., & Speckart, G. (1981). Attitudes "cause" behaviors: A structural equation analysis. *Journal of Personality and Social Psychology, 40,* 226–238.

Boomsma, A. (1982). The robustness of LISREL against small sample size in factor analysis models. In K. G. Joreskog & H. Wold (Eds.), *Systems under indirect observation, Part 1* (pp. 149–173). Amsterdam: North-Holland.

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

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

Chou, C.-P., Bentler, P. M., & 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. (in press). The robustness of test statistics to non-normality and specification error in confirmatory factor analysis. *Psychological Methods.*

Duncan, O. D. (1975). *Introduction to structural equation models.* New York: Academic.

Fiske, S. T., Kenny D. A., & Taylor, S. E. (1982). Structural models for the mediation of salience effects on attributions. *Journal of Experimental Social Psychology, 18,* 105–127.

Fleischman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika, 43,* 521–532.

Harlow, L. L. (1985). *Behavior of some elliptical theory estimators with non-normal data in a covariance structures framework: A Monte Carlo study.* Doctoral dissertation, University of California, Los Angeles.

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

James, L. R., & Brett, J. M. (1984). Mediators, moderators and tests for mediation. *Journal of Applied Psychology, 69,* 307–321.

Judd, C. M., & Kenny, D. T. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5,* 602–619.

Kaplan, D. (1988). The impact of specification error on the estimation, testing and improvement of structural equation models. *Multivariate Behavioral Research, 23,* 69–86.

Lewis, P. A. W., Goodman, A. S., & Miller, J. M. (1969). A pseudo-random number generator for the system 1360. *IBM Systems Journal, 8,* 136–146.

MacKinnon, D. P. (1994). Analysis of mediating variables in prevention and intervention studies [Monograph]. In A. Cezares & L. Beatty (Eds.), *Scientific methods for prevention intervention research* (pp. 127–153). Rockville, MD: National Institute on Drug Abuse.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review, 17,* 144–158.

MacKinnon, D. P., Johnson, C. A., Pentz, M. A., Dwyer, J. H., Hansen, W. B., Flay, B. R., & Wang, E. (1991). Mediating mechanisms in a school-based drug prevention program: First year effects of the Midwestern Prevention Project. *Health Psychology, 10,* 164–172

MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of indirect effect measures. *Multivariate Behavioral Research, 30,* 53–71.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105,* 155–166

Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology, 45,* 19–30.

Olsson, U. (1979). On the robustness of factor analysis against crude classification of the observations. *Multivariate Behavioral Research, 14,* 485–500.

Rice, J. A. (1988). *Mathematical statistics and data analysis.* Pacific Grove, CA: Wadsworth.

Satorra, A. (1990). Robustness issues in structural equation modeling: A review of recent developments. *Quality and Quantity, 24,* 367–386.

Satorra, A. (1991). *Asymptotic robust inferences in the analysis of mean and covariance structures.* Economics working paper, Universitat Pompeu Fabra.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *ASA: Proceedings of the Business and Economic Section,* 308–313.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290–313). San Francisco: Jossey-Bass.

Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. In N. B. Tuma (Ed.), *Sociological methodology* (pp. 159–186). San Francisco: Jossey-Bass.

Stone, M. E., & Sobel, M. E. (1990). The robustness of estimates of total indirect effects in covariance structure models estimated by maximum likelihood. *Psychometrika, 55,* 337–352.

Vale, C. D., & Maurelli, M. E., (1983). Simulating multivariate non-normal distributions. *Psychometrika, 48,* 465–471.

West, S. G., Aiken, L. S., & Todd, M. (1993). Probing the effects of individual components in multiple component programs. *American Journal of Community Psychology, 21,* 571–605.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equations with non-normal variables: Problems and remedies. In R. Hoyle (Ed.), *Structural equation modeling: Issues and applications* (pp. 56–75). Newbury Park, CA: Sage.

Yung, Y.-F., & Bentler, P. M. (1994). Bootstrap-corrected ADF test statistics in covariance structure analysis. *British Journal of Mathematical and Statistical Psychology, 47,* 63–84.

APPENDIX

All current methods of estimating the parameters of structural equation models involve minimizing a fitting function. For GLS and ML, the expression for the fitting function F is commonly written as

$$F = .5\mathrm{tr}\left(\{[\mathrm{S} - \Sigma(\theta)]\mathrm{V}^{-1}\}^2\right) \tag{1}$$

In this expression, tr denotes the trace of the expression in parentheses, S is the covariance matrix of the p measured variables obtained from a random sample of size N = (n + 1), θ is a vector of free parameters to be estimated, $\Sigma(\theta)$ is the model-implied covariance matrix, and V is a $p \times p$ weight matrix. For GLS, S is used to estimate V; in the ML approach, $\Sigma(\theta)$ is used to estimate V. Under the conditions of multivariate normality and a properly specified model, GLS and ML estimates are asymptotically equivalent.

Following Browne (1982, 1984), the ADF fitting function can be expressed as

$$F_{\mathrm{ADF}} = [\mathrm{s} - \sigma(\theta)]' \mathrm{W}^{-1}[\mathrm{s} - \sigma(\theta)] \tag{2}$$

where s is a column vector of the $p^* = (p)(p + 1)/2$ nonredundant elements of S, $\sigma(\theta)$ is the corresponding same-order vector of $\Sigma(\theta)$, the model-implied covariance matrix, and θ is a $t \times 1$ vector of free parameters. Values of θ are selected so as to minimize the sum of the weighted squared deviations of s from $\sigma(\theta)$. The optimal weight matrix, W, is a $p^* \times p^*$ covariance matrix of sample covariances. This covariance matrix of s is a matrix with typical element equal to

$$\mathrm{ACOV}(s_{\mathrm{ij}}, s_{\mathrm{kl}}) = N^{-1}(\sigma_{\mathrm{ijkl}} - \sigma_{\mathrm{ij}}\sigma_{\mathrm{kl}}) \tag{3}$$

where the fourth-order element, σ_{ijkl}, is equal to $E(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)(X_l - \mu_l)$, and σ_{ij} and σ_{kl} are the population covariances of X_i with X_j and X_k with X_l, respectively. In practice, s_{ij} is used to estimate σ_{ij}, and μ_i is estimated by the sample mean. If the variables are normally distributed, the asymptotic covariance between s_{ij} and s_{kl} can be expressed as

$$\mathrm{ACOV}(s_{\mathrm{ij}}, s_{\mathrm{kl}}) = N^{-1}(\sigma_{\mathrm{ik}}\sigma_{\mathrm{jl}} - \sigma_{\mathrm{il}}\sigma_{\mathrm{jk}}) \tag{4}$$

Under multinormality, the optimal weight matrix W in Equation 2 consists of products of covariances. Minimization of Equation 2 using Equation 4 to compute each of the elements in W results in GLS estimates. Thus, GLS can be seen to be a special case of the ADF approach.

To calculate the asymptotic standard errors, let D be a $p^* \times t$ matrix of derivatives of σ with respect to θ. Let W be a $p^* \times p^*$ optimal weight matrix, with elements defined by the estimation method employed. Standard errors can be obtained by inverting the information matrix $I(\theta) = (D'W^{-1}D)$.

The weight matrix W used in ML and GLS is the covariance matrix of s computed under the assumption of multinormality. This normal theory weight matrix can be expressed as

$$W = 2K'[\Sigma \otimes \Sigma]K \qquad (5)$$

where Σ is the population covariance matrix and K is a $p^2 \times p^*$ transition matrix that reduces the $p^2 \times p^2$ matrix in brackets to the appropriate $p^* \times p^*$ order (i.e., the effects of redundancy caused by the symmetry of S are eliminated). In the normal theory ML approach, $\Sigma(\theta)$ is used to estimate Σ, whereas estimation of Σ using S yields normal theory GLS estimates. The covariance matrix of the parameter estimators is obtained from $(D'W^{-1}D)^{-1} \times n^{-1}$; normal theory standard errors are given by the square root of the diagonal elements.

The W matrix used in ADF is a "distribution free" covariance matrix of the sample covariances. The elements of this fourth-order multivariate product moment matrix are given by Equation 3.

The robust approach to the computation of ML standard errors employs the matrix

$$(D'W_{ML}^{-1}D)^{-1}\ (D'W_{ML}^{-1}W_{ADF}W_{ML}^{-1}D)\ (D'W_{ML}^{-1}D)^{-1} \qquad (6)$$

Multiplication of this matrix by n^{-1} yields the robust covariance matrix of the ML estimator. Robust standard errors can be obtained as the square root of the diagonal elements.