Title

A Comparison of Maximum-Likelihood and Asymptotically Distribution-Free Methods of Treating Incomplete Non-Normal Data

Authors

Gold, Michael S.
Bentler, Peter M
Kim, Kevin H.

Publication Date

2002

Running head: TREATMENTS OF INCOMPLETE NON-NORMAL DATA

A Comparison of Maximum-Likelihood and Asymptotically Distribution-Free Methods of Treating Incomplete Non-Normal Data

Michael S. Gold, Peter M. Bentler, and Kevin H. Kim

University of California, Los Angeles

Michael S. Gold, Department of Psychology, University of California, Los Angeles; Peter M. Bentler, Department of Psychology, University of California, Los Angeles; Kevin H. Kim, Department of Psychology, University of California, Los Angeles.

Requests for reprints should be sent to Peter M. Bentler, UCLA Department of Psychology, Box 951563, Los Angeles, California 90095-1563. E-mail: bentler@ucla.edu

Abstract

This article describes a Monte Carlo study of 2 methods for treating incomplete non-normal data. Skewed, kurtotic data sets conforming to a single structured model, but varying in sample size, percentage of data missing, and missing-data mechanism, were produced. An asymptotically distribution-free available-case (ADFAC) method and structured-model expectation-maximization (EM) with non-normality corrections were applied to these data sets, and the 2 methods were then compared in terms of bias in parameter estimates, bias in standard-error estimates, efficiency of parameter estimates, and model chi-squares. The results favored the non-normality corrected EM over the ADFAC method in almost all respects, the only important exception involving bias in standard-error estimates with large samples.

A Comparison of Maximum-Likelihood and Asymptotically Distribution-Free Methods of Treating Incomplete Non-Normal Data

This article addresses the problem of incomplete data due to item nonresponse, in the context of structural equation modeling analyses. Increasingly, maximum-likelihood (ML) methods, particularly expectation-maximization (EM), are being seen as the methods of choice for dealing with analyses of incomplete data (e.g. Schafer, 1997), although their underlying assumption of normally distributed data would seem to limit their applicability (Graham, Hofer, & MacKinnon, 1996) in the absence of certain adjustments described below. Another method (analysis of available cases, also called *pairwise deletion*) would seem to be effective on intuitive grounds, but it does not consistently fare well (e.g. Gleason & Staelin, 1975; Brown, 1994; Graham et al., 1996; Marsh, 1998). This article explores a modification that might justify the resurrection of available-case analysis and make it especially useful with nonnormal data. Its performance is compared to that of a version of EM in which adjustments have been made to enhance its robustness to departures from normality. Both methods are evaluated in a structural equation modeling context.

Incomplete data: types and mechanisms of nonresponse

In a recent dissertation, Bernaards (2000) distinguished among three types of nonresponse resulting in incomplete data. One is called *unit nonresponse*, which refers to the outright failure of a sampled subject to participate in a study. A second is called *subject attrition* (or *subject mortality*), which occurs in a longitudinal study when a subject who has participated in earlier phases of the study ceases to be available for further participation. The third is called *item nonresponse*, which occurs in any kind of multivariate study (e.g. a survey) in which a subject responds to some items but not others. Incomplete data resulting from this third type of nonresponse will be the focus of this study.

Methods of analyzing incomplete data resulting from item nonresponse differ along several dimensions. One of these involves the underlying assumption of each method vis-à-vis the nonresponse mechanism. Three classes of mechanism have been distinguished in the existing literature: *missing completely at random* (MCAR), *missing at random* (MAR), and *not missing at random* (NMAR). A datum that is missing for reasons independent of the value of *any* variable is termed MCAR, while a datum that is

missing because of the value of some *observed* variable is called MAR; a datum is considered NMAR when it is missing because of the value of some *unobserved* variable, possibly including the value of the variable that is missing.

Solutions to incomplete data

Little and Rubin (1989/1990) distinguished three strategies for dealing with the incomplete-data problem. The first two (*imputation* and *weighting*) seek to restore the rectangular form of the data matrix for subsequent analysis. Imputation accomplishes this by substituting a seemingly reasonable estimate for each missing datum, whereas the weighting strategy deletes incomplete cases and then, in subsequent analysis, reweights the complete cases to compensate for the omitted ones. Imputed values are commonly chosen by regression (with mean imputation as a special case) or by sampling from other cases in the same data set (hot-deck imputation); among weighting approaches, the simplest form is complete-case analysis (also called *listwise deletion*), which assumes that the complete cases are a random subsample of the original sample (i.e. missingness is MCAR) and therefore weights all complete cases equally. Parameter estimation in the complete-case method, although unbiased for MCAR data, is necessarily inefficient due to the waste of measured data from incomplete cases. Moreover, even with a fixed number of cases in the data set and a fixed proportion of data missing, a greater proportion of cases will need to be deleted to conduct complete-case analysis as the number of variables in the data set increases (Little & Rubin, 1987). Single imputation (multiple imputation will not be addressed here) has not usually fared well in previous studies, even when the imputations were iterated to yield more accurate parameter estimates and/or a stochastic component was added to avoid overstated precision of those estimates (e.g. Gold & Bentler, 2000).

The third missing-data strategy identified by Little and Rubin (1989/1990) is *direct analysis of the incomplete data*, which is "direct" in the sense that analysis is undertaken without any attempt to restore the rectangular form of the data matrix. One exemplar of this strategy is available-case (also called *pairwise-deletion*) analysis, in which each quantity required for a statistical analysis is simply computed from all data available to calculate it: A variable's sum of squares, for example, is calculated from data provided by every case for which that variable was observed, regardless of whether that case can contribute an

observation to calculating some other sum of squares; similarly, the set of cases providing the paired observations needed for a sum of crossproducts might not be identical to the set of cases able to provide paired observations for some other sum of crossproducts.

Several disadvantages of available-case analysis have been identified. Available-case methods are based on the unlikely assumption that missing data are MCAR. In addition, available-case methods will often produce a covariance matrix of indeterminate sample size (Brown, 1994; Marsh, 1998), because the number of observations involved in calculations can vary from variable to variable (and from pair to pair of variables); this of course makes it impossible to appropriately calculate statistics that assess model fit. According to Little and Rubin (1987, 1989/1990), available-case methods are sometimes even less efficient than complete-case analysis and are otherwise unreliable: Depending on the analyst's choice of formula, R^2 values may be less than zero or may exceed unity, or elements of the covariance matrix may be mutually inconsistent (i.e. the covariance matrix may fail to be positive definite). Little and Rubin (1987) also noted that there are different, nonequivalent ways to calculate an available-case covariance matrix.

Most other exemplars of the direct analysis strategy involve estimation by maximum likelihood. Case-based ML for incomplete data was originated in structural equation modeling by Finkbeiner (1979) and Lee (1986) and was extended to a wider class of structural equation models by Arbuckle (1996, 1997). Jamshidian and Bentler (1999) proposed computations based on an EM algorithm. The EM algorithm generates a sequence of parameter estimates by cycling iteratively between an expectation (E) step and a maximization (M) step. In the E-step, the conditional expectation of the complete data log-likelihood is computed given the observed data and the current parameter estimates. In the M-step, the expected log-likelihood obtained in the E-step is maximized to obtain updated parameter estimates under the structural model. The iteration process is stopped when its parameter estimates converge to some pre-established criterion. Unlike the other methods just described, ML methods are based on the assumption that the missingness mechanism is at least MAR, but not necessarily MCAR; on the other hand, it is assumed that the data are normally distributed. Still, initial indications are that ML methods may well outperform others discussed above even with nonnormal data (see Graham, Hofer, & MacKinnon, 1996).

Detailed descriptions and discussions of EM can be found in papers such as those by Dempster, Laird and Rubin (1977), Little and Rubin (1987, 1989/1990), Little and Schenker (1995), and Rubin

(1991), as well as a book by Schafer (1997). The EQS implementation of EM is that of Jamshidian and Bentler (1999).

Background

Before describing the method of our investigation, we summarize the published literature on Monte Carlo research involving EM and/or available-case (AC) methods for dealing with missing-data. This summary is largely extracted from the literature review in Gold & Bentler (2000).

Gleason and Staelin (1975) evaluated the available-case method and four imputation methods (mean; noniterated, nonstochastic multiple regression; iterated principal component; first principal component) for their abilities to reproduce correlation matrices following random deletion of data. They also varied sample size, number of variables, proportion of data deleted, and the square root of the mean squared off-diagonal correlation. Data were presumably (but not explicitly) multivariate normal. They found that reproduction of the correlation matrix was better for the available-case method than for mean imputation when correlations and the nominal number of cases were sufficiently large. However, the available-case method was outperformed by regression imputation and by iterated imputation using principal components when the root mean squared off-diagonal correlation exceeded about .12, at least when the data consisted of 100 cases, 15 variables, and 12% missing data.

Finkbeiner (1979) conducted a comparative simulation study involving estimation of parameters of a multiple factor model involving missing data. A case-based maximum-likelihood (ML) method was compared to five heuristic methods: complete cases; available cases; mean imputation; iterated, nonstochastic multiple regression imputation; and iterated principal component imputation with the number of components determined for each sample by scree test. Finkbeiner's simulated, multivariate normal data conformed to a six-variable, two-factor model. His data-deletion method cannot be briefly characterized, except to say that data in each simulated sample were deleted to conform to one of two distributions of frequencies with which certain combinations of variables were missing. Missingness for each simulated variable was independent of the *values* of all variables. Therefore, the observed data constituted a completely random sample of the simulated population from which they were drawn. The ML method generally produced the least biased and among the most efficient parameter estimates, with the available-

case method often performing nearly as well. By many measures, regression imputation also performed well, and mean regression also did surprisingly well by some measures. The principal component method performed relatively poorly. Thus, Finkbeiner's findings are somewhat inconsistent with those of and Gleason and Staelin (1975), although it is not obvious why.

Brown (1994) compared complete-case analysis, available-case analysis, mean imputation, hot-deck imputation, and a form of resemblance-based hot-deck imputation (RBHDI, termed "similar response pattern imputation" by Brown (1994) and initially described by Jöreskog and Sörbom (1993)). The resemblance-based hot-deck technique can be considered a variation on other hot-deck methods, in the sense that the imputed datum for one case is taken from another case in the same sample. The resemblance-based method differs from existing hot-deck methods in that the substituted value is determined not by random sampling, but by designating as the "donor" that case which most resembles the "recipient" on a set of standardized variables for which data are complete. *Resemblance* is defined in terms of a least-sum-of-squares criterion involving those variables. Brown's study also varied simulated sample size and proportion of data deleted at random. He used a four-factor, 10-variable structural model as the basis for generating his raw data matrices, which had a multivariate normal distribution. He evaluated the five missing data treatment methods in terms of model convergence, model fit, bias in parameter estimates, and size of the standard errors of those estimates.

Apparently no method was consistently superior to the others. *Model convergence* was a problem only for simulated $N = 100$, and least of all for the available-case method. (As might be expected, failure to converge also increased as the proportion of data missing increased.) At $N = 500$, where convergence was not a problem, the rate of *model rejection* deviated from the nominal 5% most for mean imputation, followed by the RBHDI, available-case, hot-deck, and complete-case methods. *Bias in parameter estimates* was generally smallest for RBHDI and greatest for the hot-deck, mean, and available-case methods. RBHDI also had the advantage for *standard errors of parameter estimates*: The average calculated standard deviation of the empirical sampling distribution of parameter estimates was almost always smaller for RBHDI than for the other four methods.

Graham, Hofer and MacKinnon (1996) estimated variances and covariances in a study involving simulated multivariate normal data. Of five variables, two were present for every subject; of the remaining

three, one was randomly chosen for deletion from each subject. The bias and efficiency of the estimated variances and covariances was then estimated for each of the following methods: multiple imputation; EM; raw-data ML; multiple-group structural equation modeling; mean imputation; EM-based single imputation; and available-case analysis. Estimates were essentially unbiased for all but two methods, i.e. single imputation (which underestimated the variances of the three variables with incomplete data) and mean imputation (which underestimated all variances and covariances involving the same three variables). As a measure of efficiency, the average standard error was .025 for each method except multiple imputation (.027) and available-case analysis (.034).

Graham et al. then replicated their simulation with two changes: (1) the simulated data were univariately nonnormal, with skew ranging from -.84 to 3.16 and kurtosis ranging from -.04 to 11.97; and (2) only the EM and available-case methods were used. Average residual of the estimates was .009 for both methods, and the root mean square residual was .024 for EM and .031 for available-case analysis. The average standard error was .059 for EM and .079 for available-case analysis. The deviations from normality thus produced some slight overestimation of the elements of the covariance matrix and loss of efficiency compared to the normal-data results. A third simulation with normally distributed MAR data revealed some loss of efficiency for all methods (but least of all for EM and raw ML) with increasing "at-random" missingness, and also showed increasing bias, except for EM and raw ML.

Marsh (1998) conducted a simulation study to evaluate the performance of the available-case method as a function of sample size and percentage of data missing completely at random. The method was applied to data generated from a three-factor population model, and performance was evaluated in terms of ML parameter estimates, fit indices, and the incidence of nonpositive definite matrices resulting when the available-case sample covariance matrices were analyzed with the true population model. Marsh's study was also an attempt to address a specific problem with available-case covariance matrices, namely that "the number of cases per variable (NPV) and the number of cases used in the construction of each covariance (NPC) varies so that there is a distribution of sample sizes rather than a single value" (p. 23). It is thus impossible to specify a correct single value for "the" size of the sample upon which the analysis is based. Accordingly, Marsh evaluated the suitability of the minimum, mean, and maximum NPV

and NPC as substitutes for the nominal sample size in SEM analyses with available-case covariance matrices.

Out of 4000 covariance matrices generated in Marsh's study, only 27 were nonpositive definite, and all of these were in the condition that combined the smallest nominal sample size (N = 200) with the largest proportion of data missing (50%). Unlike Brown (1994, described above), Marsh found that parameter estimates (i.e. factor loadings, factor correlations, and measured-variable uniqueness) based on the available-case sample covariance matrices were unbiased regardless of nominal sample size and proportion of data missing. Effects of nominal sample size, proportion of data missing, and their interaction on the standard deviations of the parameter estimates were "largely eliminated" by a simple adjustment in which each observed standard deviation was multiplied by (mean NPC/1000)^{½}, where the mean NPC is calculated over the 45 covariances per measured-variable covariance matrix. This good news regarding parameter estimates and the incidence of nonpositive definite matrices notwithstanding, fit indices (chi-square, relative noncentrality index, nonnormed fit index, and root mean square error of approximation) were all adversely affected by decreasing nominal sample size, by increasing proportion of data missing, and by their interaction. This bias was not fully rectified by any a priori adjustment in which the minimum, mean or maximum NPC or NPV was specified for the sample size.

Gold and Bentler (2000) compared the accuracy of variance and covariance estimates generated for a four-factor, ten-variable path model by four missing-data treatments: structured-model EM, saturated-model EM, resemblance-based hot-deck imputation, and iterated stochastic regression imputation. The simulated data were distributed (in separate conditions of the study) either normally or univariately nonnormally, all with an MCAR missing-data mechanism. Each simulated sample consisted of either 100 or 500 cases, and the percentage of data missing varied from 4 to 16 percent. In almost every condition of the study, the EM methods were found superior to the other two methods.

Wothke (2000) compared the abilities of four methods (full information ML, mean imputation, pairwise deletion, and listwise deletion) to reproduce parameter values in a growth curve model involving four measurements per subject. Missing data probabilities were zero at time 1, 10% at time 2, 20% at time 3, and 30% at time 4. Data were multivariate normal and used 400 samples of n = 500. In the first simulation, involving a MCAR mechanism, the six parameters were estimated essentially without bias by

all methods except mean imputation; efficiency of estimates was best for full information ML and worst for listwise deletion. The second simulation differed from the first by using an MAR mechanism, with realized missing data rates of zero at time 1, 16 to 20% at time 2, 25 to 33% at time 3, and 30 to 40% at time 4. In this simulation, only full information ML estimated the parameters essentially without bias.

Enders and Bandalos (2001) compared four missing-data methods (full information maximum likelihood, listwise deletion, pairwise deletion, and similar response pattern imputation) with respect to convergence failure, bias and efficiency of parameter estimates, and model fit, with all conditions of the study using normal data. They found that full information ML was equivalent or superior to the other three methods on all four outcome measures, regardless of factor loading magnitude, sample size, missing data rate, and missing-data mechanism (MCAR or MAR).

Enders (2001) compared full information ML estimation with listwise deletion, pairwise deletion, mean imputation, and similar response pattern imputation with respect to parameter estimate bias and efficiency, confidence interval coverage, and rate of model rejection, in models with nonnormal indicators and either an MCAR or an MAR missing-data mechanism. Under MCAR conditions, there was little or no bias, except with mean imputation; full information ML showed increasing efficiency advantages over the other methods with increasing rates of missing data; although there was little effect of missing-data method on confidence interval coverage, standard errors tended to be negatively biased; and model rejection rates were quite high for all methods, including full information ML. Under MAR conditions, full information ML showed the least parameter estimate bias; parameter estimate accuracy (i.e. the combination of efficiency with lack of bias) depended interactively on missing-data method, degree of nonnormality, and sample size, with some ad hoc methods performing better than full information ML under conditions of extreme nonnormality and small sample size; there was, again, little effect of missing-data method on confidence interval coverage; and model rejection rates were again quite high for all methods.

Enders (2001) conducted an additional study that addressed the distortion in significance tests of both model fit and individual parameter estimates that is found with full information ML. Specifically, he evaluated the use, in the missing-data context, of robust standard errors and rescaled chi-square test statistics (see Yuan & Bentler, 2000), as compared with the use of bootstrap methods. Enders found that

robust standard errors tended to be somewhat positively biased, and that the rescaled chi-square test statistics yielded model rejection rates close to the nominal 5% rate.

Thus these studies, mostly involving MCAR data (but with four involving MAR data), seem to indicate that EM and other ML (as well as multiple-imputation) methods produced results with the least bias and greatest efficiency; the available-case method also produced unbiased results (at least in the MCAR case), but usually performed suboptimally with respect to efficiency and fit indices.

ADF and ML approaches

As indicated above, the main goal of this article is to evaluate a modification to available-case analysis that might justify the rehabilitation of the latter as a method of dealing with incomplete data and make it especially useful with nonnormal data. The modification is based on a paper by Van Praag, Dijkstra, and Van Velzen (1985).

Van Praag et al. (1985) suggested an asymptotically distribution-free available-case (ADFAC) approach that addresses at least some of the disadvantages of ordinary available-case analysis in a linear regression context. Their method involves the calculation of an omega ($\mathbf{\Omega}$) matrix "with typical element $\Omega_{ij,mn} = \mu_{ijmn} - \sigma_{ij}\sigma_{mn}$, where $\mu_{ijmn} \equiv E(z_i - \mu_i)(z_j - \mu_j)(z_m - \mu_m)(z_n - \mu_n)$" and $\sigma_{ij} \equiv E(z_i - \mu_i)(z_j - \mu_j)$ (Van Praag et al., 1985, p. 27); in practice, each expectation is replaced by the corresponding sample mean, thus yielding the estimate $\tilde{\Omega}_{ij,mn}$. When data are missing, each $\tilde{\Omega}_{ij,mn}$ must be calculated on an available-case basis and then corrected by the factor $p_{ijmn}/(p_{ij}p_{mn})$, where each p_{ijmn}, p_{ij}, or p_{mn} value reflects the proportion of the sample cases in which the variables indexed are simultaneously present. This yields the corrected available-case estimated omega matrix $\breve{\Omega}$. This matrix is identical to the inverse of the weight matrix used in ADF parameter estimation with complete data. In principle, the method yields a consistent estimated covariance matrix, and the use of the nominal sample size yields correct values of test statistics; in practice, nonpositive definite covariance matrices appear to be a small-sample problem (Van Praag et al., 1985).

The Van Praag et al method can be extended to a structural equation modeling context as follows. We assume that $\sigma = \sigma(\theta)$ and that, as $n \to \infty$, $\tilde{s} \to \sigma$ and $\sqrt{n}(\tilde{s} - \sigma) \to N(0, \Omega)$, where σ is the vectorized triangular population covariance matrix Σ and $\tilde{s}$ is the similarly vectorized pairwise estimate of σ. Using

$\breve{\Omega}$ as a quadruplewise-present estimate of Ω, we find $\tilde{\theta}$ by minimizing $(\tilde{s}-\sigma(\theta))'\breve{\Omega}^{-1}(\tilde{s}-\sigma(\theta))$. As a result, based on standard arguments (e.g. Browne, 1984), $n(\tilde{s}-\sigma(\tilde{\theta}))'\breve{\Omega}^{-1}(\tilde{s}-\sigma(\tilde{\theta}))$ approaches a chi-square distribution with p^*-q degrees of freedom, where p is the number of measured variables, $p^*=\frac{p(p+1)}{2}$, and q is the number of free parameters. Furthermore, $\sqrt{n}(\tilde{\theta}-\theta)\rightarrow N(0,(\Delta\Omega^{-1}\Delta')^{-1})$, where $\Delta=\frac{\partial\sigma(\theta)}{\partial\theta}$; therefore the estimated standard error of θ is given by the square root of the diagonal elements of $\frac{(\tilde{\Delta}\breve{\Omega}^{-1}\tilde{\Delta}')^{-1}}{n}$ where $\tilde{\Delta}$ is an estimate of Δ. Thus the ADFAC method provides a means for obtaining parameter estimates, estimated standard errors, and statistics to evaluate model fit.

A second goal of this article is to evaluate a version of EM in which adjustments have been made to enhance its robustness to departures from normality. A disadvantage of ML methods, mentioned above, is that they are based on assumptions of normally distributed data. Although ML parameter estimates are largely robust to moderate violations of normality, ML methods are not robust with respect to standard error estimates and overall model fit chi-squares. These deficiencies have been addressed in complete data with some success by the development of formulas for robust standard error estimates and scaling constants for the chi-square statistic (e.g. Satorra and Bentler, 1994). These formulas have been extended to the case of incomplete data (Arminger and Sobel, 1990; Yuan and Bentler, 2000) and are implemented in EQS 6.0.

According to Arminger and Sobel (1990), the robust standard error for a parameter estimate $\hat{\theta}$ is given by $\frac{\Omega_{\hat{\theta}}}{n}$, where $\Omega_{\hat{\theta}}=A_\theta^{-1}B_\theta A_\theta^{-1}$, $A_\theta=-\lim_{n\rightarrow\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 l_i(\theta_0)}{\partial\theta_0\partial\theta_0'}$, $B_\theta=\lim_{n\rightarrow\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial l_i(\theta_0)}{\partial\theta_0}\frac{\partial l_i(\theta_0)}{\partial\theta_0'}$, $l_i(\theta_0)=\frac{p_i}{2}\log(2\pi)-\frac{1}{2}\{\log|\Sigma_i(\theta_0)|+(X_i-\mu_i(\theta_0))'\Sigma_i^{-1}(\theta_0)(X_i-\mu_i(\theta_0))\}$, p_i is the dimension of case X_i, $\Sigma_i(\theta_0)=\mathrm{cov}(X_i)$ is a submatrix of the population covariance matrix Σ, and $\mu_i(\theta_0)=\mathrm{E}(X_i)$ is a subvector of the population mean vector μ. In practice, A_θ may be estimated by the sample information matrix, and B_θ as the sample covariance matrix of the score vector.

According to Yuan and Bentler (2000), the scaled chi-square statistic is given by

$$T_2^* = \frac{(p + p^* - q)T_2}{\mathrm{tr}(\hat{\Omega}_{\hat{\beta}}\hat{V}_2)}, \text{where } T_2 = 2[l(\hat{\beta}) - l(\hat{\theta})],\ l(\hat{\beta}) = \sum_{i=1}^{n} l_i(\hat{\beta}),\ \ l(\hat{\theta}) = \sum_{i=1}^{n} l_i(\hat{\theta}),$$

$$l_i(\hat{\beta}) = \frac{p_i}{2}\log(2\pi) - \frac{1}{2}\{\log\left|\Sigma_i(\hat{\beta})\right| + (X_i - \mu_i(\hat{\beta}))'\Sigma_i^{-1}(\hat{\beta})(X_i - \mu_i(\hat{\beta}))\},$$

$$l_i(\hat{\theta}) = \frac{p_i}{2}\log(2\pi) - \frac{1}{2}\{\log\left|\Sigma_i(\hat{\theta})\right| + (X_i - \mu_i(\hat{\theta}))'\Sigma_i^{-1}(\hat{\theta})(X_i - \mu_i(\hat{\theta}))\}, \text{ where } \Sigma_i(\hat{\beta}) = S_n \text{ and } \mu_i(\hat{\beta}) = \bar{X}_n \text{ are}$$

treated as a saturated-model EM-estimated ML sample covariance matrix and mean vector, and

$\Sigma_i(\hat{\theta})$ and $\mu_i(\hat{\theta})$ are similar EM/ML estimates computed under the structured model. Further,

$$\hat{\Omega}_{\hat{\beta}} = A_\beta^{-1}B_\beta A_\beta^{-1},\ A_\beta = -\frac{1}{n}\sum\nolimits_{i=1}^{n}\frac{\partial^2 l_i(\hat{\beta})}{\partial\hat{\beta}\partial\hat{\beta}'},\ B_\beta = \frac{1}{n}\sum\nolimits_{i=1}^{n}\frac{\partial l_i(\hat{\beta})}{\partial\hat{\beta}}\frac{\partial l_i(\hat{\beta})}{\partial\hat{\beta}'}; \text{ and where } \hat{V}_2 \text{ is an estimate}$$

$$\text{of } V_2 = H_2 - H_2\dot{\beta}(\dot{\beta}'H_2\dot{\beta})^{-1}\dot{\beta}'H_2,\ H_2 = \begin{pmatrix} \frac{1}{2}\lim\limits_{n\to\infty}\frac{1}{n}\sum\limits_{i=1}^{n}\kappa_i'(\Sigma_i^{-1}\otimes\Sigma_i^{-1})\kappa_i & 0 \\ 0 & \lim\limits_{n\to\infty}\frac{1}{n}\sum\limits_{i=1}^{n}\tau_i'\Sigma_i^{-1}\tau_i \end{pmatrix},\ \kappa_i = \frac{\partial\mathrm{vec}(\Sigma_i)}{\partial\sigma'},$$

$\Sigma_i = \mathrm{cov}(X_i)$, $\sigma = \mathrm{vech}(\Sigma)$, vec(•) is an operator that transforms a matrix into a vector by stacking the columns of the matrix, $\tau_i = \frac{\partial\mu_i}{\partial\mu'}$, $\mu_i = E(X_i)$, and $\dot{\beta} = \frac{\partial\beta(\theta_0)}{\partial\theta_0'}$. In practice, A_β and B_β may be estimated by the sample information matrix and the sample covariance matrix of the score vector.

Thus EM can be applied to nonnormal missing data in a way that adds estimated standard errors and model fit statistics to the ML parameter estimates.

This article compares the ADFAC method to the adjusted EM method in a Monte Carlo simulation study that varies parameter estimation/missing-data treatment method, simulated sample size, and proportion of data missing, and investigates their performance with nonnormal data that are incomplete by virtue of MCAR and MAR mechanisms. Performance will be assessed in terms of bias of parameter estimates and standard error estimates, efficiency of parameter estimates, and overall model fit.

METHOD

Terminology

Some potential for confusion arises from the fact that a Monte Carlo study is an *actual* investigation of the behavior of statistics resulting from a large number of *simulated* investigations. Both the simulated investigations and the actual investigation of which they are a part involve samples, data, variables and (in a sense) subjects. Therefore, to reduce the likelihood of terminological confusion, we distinguish between (1) the samples, data, variables and subjects of the simulated investigations and (2) what we call the *metasamples*, *metadata*, *metavariables* and *metasubjects* of the actual (Monte Carlo) investigation that is the focus of this paper. For similar reasons, we describe the design of the actual investigation separately from the design of the simulated investigations.

Design of Simulated Investigations

Each simulated investigation involved 16 manifest variables (all continuous, with a zero mean vector) and either 500 or 5000 subjects. Data were randomly generated by EQS 6.0, subject to a path model consisting of four latent variables, each with four manifest-variable indicators. The model is depicted in **Figure 1**. In addition, the simulation process ensured that the data were univariately nonnormal. The specific variance, skew, and kurtosis values assigned to each variable are shown in **Table 1**.)

Design of the Monte Carlo Study

Independent Metavariables

The Monte Carlo study involved four independent metavariables that will combine to produce 2 x 2 x 2 x 2 = 16 cells in a balanced design. The independent metavariables were:

(1) *Nominal size of simulated raw data matrix (Sample Size)*: $N = 500$ or $N = 5000$.

(2) *Proportion of data deleted/missing (Missingness)*: .15 or .30.

(3) *Missing-data mechanism (Mechanism)*: MCAR or MAR.

(4) *Missing data treatment and parameter estimation method (Treatment)*: (a) asymptotically distribution-free available-case (ADFAC) analysis; or (b) structured-model EM with non-normality corrections.

Metasample Size, Data Generation, Metadata Generation, and Statistical Analysis of Metadata

Metasample Size

The study used 100 samples (raw data matrices, metasubjects) per combination of levels of the three independent metavariables (i.e. each "meta-n" = 100), thus yielding a total of 1600 raw data matrices.

Data Generation

The 1600 data matrices were constructed in the following three-step process:

(1) *Initial matrix creation*. As was indicated previously, the random number generator in EQS 6.0 produced all initial data, subject to the model in Figure 1. All EQS-created data consisted of complete raw-data matrices (with data deletion accomplished by a subsequent procedure, as described below). Two metasamples (i.e. 200 unique, complete raw data matrices) were thus produced: one metasample for each sample size (N = 500 vs. N = 5000). Accordingly, sample size constitutes a between-metasamples metavariable (analogous to a "between-groups variable"). Following this step, there were 200 raw data matrices.

(2) *Replication and data deletion*. Each metasubject (raw data matrix) was used in all combinations of missing data percentage (15 vs. 30 percent) with missing-data treatment (ADFAC vs. structured-model EM). To accomplish this within-metasubject manipulation, each data matrix from step one was replicated four times, and each replicate matrix was processed through one of two data deletion programs written in the Microsoft Windows version of MatLab 5.3. Accordingly, each input matrix yielded five output matrices (the original plus four replicates) differing only in proportion of data missing and missingness mechanism. One (MAR) data deletion program deleted all variables except V1 and V5 with a probability of .343 for any case for which V1 + V5 is in the upper half of its distribution (thus yielding approximately 15% MAR missingness) or with a probability of .686 for any case for which V1 + V5 is in the upper half of its distribution (thus yielding approximately 30% MAR missingness). The other (MCAR) program was written so as to ensure (for each matrix) random selection without replacement, subject to the requirement that the proportion of data missing was equal to that for the corresponding matrix in the MAR condition. Therefore, (a) every data point within a matrix had an equal and independent

probability of deletion, and (b) random deletion continued until exactly the specified proportion of data points was deleted from the matrix. The randomness of the process was not be restricted in any other way, and all deleted data resulting from this program were therefore MCAR. Following this step, there were 800 raw data matrices with missing data (plus the 200 intact matrices created in step one).

(3) *Replication and missing-data treatment*. Each incomplete data matrix from the previous step was then replicated to produce two identical matrices. Each of the resulting two identical matrices was then processed through one of the two missing data treatments described previously. Following this step, there were 1600 raw data matrices, each yielding parameter estimates, standard error estimates, and a model chi-square value (plus the original 200 intact matrices from step one). This concluded the data generation process.

Metadata Generation:

Calculations of Dependent Metavariables

Each of the 1600 incomplete raw data matrices yielded 37 parameter estimates, 37 standard error estimates, and a model chi-square value. The parameter estimates for each raw data matrix consist of 17 regression coefficients (12 factor loadings and five path coefficients) and 20 variances (16 error variances, three disturbance variances, and one factor variance).

Bias of parameter estimates. We assessed bias in parameter estimates by subtracting from them the corresponding known parameter values, and then dividing the difference by the known parameter values, thus yielding a measure here called a "relative difference". Then the root mean square relative difference (RMSD) across the 17 regression coefficients and the root mean square relative difference across the 20 variances were calculated for each raw data matrix.

Bias of standard error estimates. We assessed bias in standard error estimates by subtracting the empirical standard deviation of each estimate from its corresponding standard error estimate, and then dividing the difference by the empirical standard deviation, thus yielding another relative difference measure. As with the parameter estimates, root mean square relative differences were calculated for each data matrix, for each type of parameter.

Efficiency of parameter estimates. Within each cell of the design, the mean empirical standard deviation of each type of parameter estimate (i.e. variances and regression coefficients) across all 100 data matrices was calculated for subsequent comparison with the corresponding values from other cells of the design.

Finally, the *chi-square* values were evaluated by determining, for each combination of experimental conditions, the rate of rejection of the null hypothesis under a nominal alpha = .05. Since the fitted model is the true model, this rate should approximate .05 for the intact raw data matrices. The extent to which the rejection rate for the incomplete raw data matrices deviates from .05 is a measure of the quality of the missing-data treatment under the specified conditions. In addition, the mean and variance of the chi-square values was calculated for each cell of the design and was compared to their expected values.

Design Summary and Metadata Analysis

The 1600 matrices resulting from the above-described data generation process constitute a balanced 2 (sample size) x 2 (percentage of data missing) x 2 (missing-data mechanism) x 2 (missing-data treatment) mixed design, in which the first independent metavariable were manipulated between metasamples and the last three were manipulated within metasubjects. The resulting metadata were statistically treated by analysis of variance. For purposes of clarity, the ANOVA of this four-factor design was actually analyzed in two three-factor parts, i.e. one part for the MCAR metadata and another part for the MAR metadata.

RESULTS

Analytic Approach

Analysis of variance was used to evaluate the metadata. To simplify presentation and interpretation of the results, only the missing-data treatment main effect and interactions involving Treatment are reported, since only these effects are relevant to the purposes of the study. Prior to analysis, the metadata were screened to assure adequate compliance with analytic requirements, with particular attention to missing and outlying metadata.

MCAR Metadata

Bias (RMSDs) in Parameter Estimates. Table 2 gives the means and standard deviations of RMSDs in parameter estimates for the MCAR metadata. As is apparent at a glance, the RMSDs are smaller for ML than for ADFAC, and this effect is particularly strong for the smaller sample size and for the larger proportion of data missing. All relevant effects (Treatment, Treatment x Sample Size, Treatment x Missingness, and Treatment x Missingness x Sample Size) were statistically significant at or beyond the .004 level, and all had partial η^2 of at least .1, except for the three-way interaction (for which partial η^2 = .088 for variances and .043 for regression coefficients). Due to the significant interaction effects, simple main effects of Treatment were analyzed with *t* tests. At every combination of Missingness and Sample Size, the simple main effect of Treatment was found significant beyond the .001 level. Thus, ML was uniformly superior to ADFAC with respect to bias in parameter estimates.

Bias (RMSDs) in Standard Error Estimates. Table 3 gives the means and standard deviations of RMSDs in standard error estimates for the MCAR metadata. Contrary to the situation with parameter estimates, the RMSDs in standard error estimates are not uniformly smaller for ML than for ADFAC. In fact, they are uniformly smaller for ADFAC than for ML when Sample Size is 5000. In addition, the RMSD for the standard error of variances is smaller for ADFAC than for ML when Sample Size is 500 and Missingness is .30. All relevant effects were statistically significant at or beyond the .0005 level, except the Treatment main effect for regression coefficients, which was nonsignificant at the .2 level. All relevant significant effects had partial η^2 of at least .480. Due to the significant interaction effects, simple main effects of Treatment were analyzed with t tests. At every combination of Missingness and Sample Size except the combination of Missingness = .30 with Sample Size = 500, the simple main effect of Treatment was found significant beyond the .0005 level. (For the combination of Missingness = .30 with Sample Size = 500, the simple main effect of Treatment on the RMSD for the standard error of variances was not significant at the .34 level.) Thus, ML was usually superior to ADFAC with respect to bias in the estimated standard errors when Sample Size = 500, but ADFAC was uniformly superior to ML in the estimated standard errors when Sample Size = 5000.

Efficiency of Parameter Estimates. The mean empirical standard deviations of parameter estimates for MCAR metadata are shown in Table 4. For both variances and regression coefficients, and

for every combination of Sample Size and Missingness, the mean empirical standard deviation is smaller for ML than for the corresponding value in ADFAC. Thus for nonnormal MCAR data, parameter estimation is more efficient for ML than for ADFAC.

Chi-Square Values. The rate of rejection of the null hypothesis under a nominal alpha = .05, and the means and variances of the chi-square values, are shown for the MCAR metadata in Table 5. Recall that the expected value of a chi-square statistic equals its degrees of freedom, and the variance is twice its degrees of freedom. For the present study, df = 99.

As Table 4 shows, the rejection rate is near .05 in all ML conditions of the study, as well as in the ADFAC condition for which Sample Size = 5000 and Missingness = .15; for all other ADFAC conditions, there is a large departure from the nominal alpha, especially when Sample Size = 500. According to sign tests, the differences in rejection rate between ML conditions and their corresponding ADFAC conditions were all significant at or beyond $\underline{p} = .0005$, except where Sample Size = 5000 and Missingness = .15 (for which the rejection rates were identical) and where Sample Size = 5000 and Missingness = .30 (for which $\underline{p} = .18$).

As Table 5 also shows, the means of the chi-square values come close to the expected value of 99, except for the ADFAC conditions with Sample Size = 500, where they are rather inflated. The variances of the chi-square values tend to be somewhat inflated beyond their expected values of 198, particularly in the ADFAC conditions with Sample Size = 500. The one exception to this pattern of inflated variances is in the ML condition for which Sample Size = 500 and Missingness = .30. In that condition, the variance is actually somewhat deflated below its expected value.

MAR Metadata

Bias (RMSDs) in Parameter Estimates. Table 6 gives the means and standard deviations of RMSDs in parameter estimates for the MAR metadata. The RMSDs are uniformly smaller for ML than for ADFAC, especially for the smaller sample size. According to the analyses of variance, all relevant effects were statistically significant beyond alpha = .0005 with partial η^2 of at least .138, except the three-way interaction, which was not significant for either variances ($\underline{p} = .71$) or regression coefficients ($\underline{p} = .64$). Due to the significant two-way interaction effects, the simple main effects of Treatment were analyzed with *t* tests. At every combination of Missingness and Sample Size, the simple main effect of Treatment was

found significant beyond the .001 level. Thus, ML was uniformly superior to ADFAC with respect to bias in parameter estimates.

Bias (RMSDs) in Standard Error Estimates. Table 7 gives the means and standard deviations of RMSDs in standard error estimates for the MAR metadata. For both variances and regression coefficients, ML shows less bias (smaller RMSDs) than ADFAC when Sample Size = 500, regardless of Missingness, but ADFAC shows less bias than ML when Sample Size = 5000. According to the ANOVAs, all relevant effects were statistically significant at the .0005 level, with partial η^2 of at least .151, with two exceptions. For both variances and regression coefficients, the Treatment x Missingness interaction failed to reach significance even at the .3 level. Due to the significant three-way interaction effects, the simple main effects of Treatment were analyzed with t tests. At every combination of Missingness and Sample Size, the simple main effect of Treatment was found significant beyond the .0005 level. Thus, bias in the standard error estimates was less for ADFAC than for ML when Sample Size = 5000, but was less for ML than for ADFAC when Sample Size = 500.

Efficiency of Parameter Estimates. The mean empirical standard deviations of parameter estimates for MAR metadata are shown in Table 8. For both variances and regression coefficients, and for every combination of Sample Size and Missingness but two, the mean empirical standard deviation is smaller for ML than for the corresponding value in ADFAC. The two exceptions involved both the variances and the regression coefficients for Sample Size = 5000 and Missingness = .15, where the mean empirical standard deviations of parameter estimates for ADFAC were equal to those of ML. Thus for nonnormal MAR data, parameter estimation is at least as efficient for ML as for ADFAC.

Chi-Square Values. The rate of rejection of the null hypothesis under a nominal alpha = .05, and the means and variances of the chi-square values, are shown for the MAR metadata in Table 9.

As Table 9 shows, the ML rejection rate tends to fall below the nominal alpha when Sample Size = 500, and it inflates the rejection rate when Sample Size = 5000 and Missingness = .30. The ADFAC rejection rate, by comparison, is inflated, usually severely. According to sign tests, the differences in rejection rate between ML conditions and their corresponding ADFAC conditions were all significant at or beyond p = .003.

As Table 9 also shows, the means of the ML chi-square values come reasonably close to the expected value of 99, although tending to fall somewhat below that for Sample Size = 500. The means of the ADFAC chi-square values all exceed 99, usually by a large amount. The variances of both ML and ADFAC chi-square values tend to be inflated, except for ML with Sample Size = 500, where they are somewhat deflated. The inflated variances are especially severe for ADFAC.

DISCUSSION

The main purpose of this study was to evaluate a modification to available-case analysis that might make it particularly useful in dealing with nonnormal incomplete data, and to compare this modification to an implementation of expectation-maximization that gives standard-error estimates and model fit statistics as well as maximum-likelihood parameter estimates. The basic findings were as follows.

For the MCAR metadata, EM produced parameter estimates that were uniformly less biased than those produced by ADFAC. EM also produced standard-error estimates that were usually less biased than ADFAC when Sample Size = 500, but were more biased than ADFAC when Sample Size = 5000. The empirical standard deviations of parameter estimates were uniformly smaller for EM than for ADFAC, and the chi-square rejection rate was usually closer to the nominal rate (alpha = .05) with EM than with ADFAC.

For the MAR metadata, EM again produced parameter estimates that were uniformly less biased than those produced by ADFAC. Again, EM produced standard-error estimates that were less biased than ADFAC when Sample Size = 500, but were more biased than ADFAC when Sample Size = 5000. The empirical standard deviations of parameter estimates were usually smaller for EM than for ADFAC. Finally, the chi-square rejection rate was always closer to the nominal rate with EM than with ADFAC.

Thus the main take-home message of these results is that the EM implementation of ML tends to outperform ADFAC in almost all respects, regardless of missing-data mechanism. Estimates are less biased, and have less sampling variability. The chi-square test statistic is more trustworthy. The only important respect in which ADFAC outperforms ML involves bias in standard-error estimates with large samples. This implies that the formulas used to estimate standard errors need improvement. In the case of

the adjusted EM procedure, these formulas are given by a sandwich-type triple product matrix $\Omega_{\hat{\theta}} = A_{\theta}^{-1} B_{\theta} A_{\theta}^{-1}$ (Yuan & Bentler, 2000). It has recently been noted that "the sandwich estimate is often far more variable than the usual parametric variance estimate" (Kauermann & Carroll, 2001, p. 1387). Since a parametric variance estimate is not available in this situation, the only possibility is to improve the estimator. Possibly this could be done by using the Hessian ("observed information matrix") as an estimate of A_{θ} instead of the information matrix, as was done here. Alternatively, an adjustment akin to those proposed by Kauermann and Carroll may be derivable for the structural modeling context. Finally, a resampling estimator might be considered instead. In the case of the ADFAC methodology, the asymptotically minimum variance formula is implemented, and it performs well when N = 5000. However, as with standard ADF methodology, at N = 500 the formula is quite inaccurate and there is a need for an improved estimator.

One surprising finding involved the influence of sample size and missingness on the chi-square test statistic when ADFAC was used on MAR data. In the complete-data case, it has been found (Curran, West & Finch, 1996; Hu, Bentler & Kano, 1992) that ADF parameter estimation yields close-to-nominal chi-square rejection rates with very large sample sizes (e.g. $N = 5000$). In the current study, such a result was not found with MAR data when Missingness = .30. To the contrary, the chi-square mean actually was greater when Sample Size = 5000 than when Sample Size = 500. To determine whether there was something anomalous about the Sample Size = 5000 / Missingness = .30 cell, we ran two additional conditions, i.e. Missingness = .20 and Missingness = .25. The resulting pattern was that both rejection rate and chi-square mean increased with increasing missingness. This finding underscores the importance of the MCAR assumption for the ADFAC method. Evidently, ADFAC becomes essentially useless with MAR data as missingness increases. In fact, of all conditions of the study, it seems that ADFAC is potentially useful only with very large datasets having relatively little data missing due to an MCAR mechanism.

Thus we come to our recommendation, which is essentially that, in any practical missing-data situation, the ADF available-case method not be used. Instead, ML methods seem to be the methods of choice (e.g. Enders & Bandalos, 2001; Graham, Hofer & MacKinnon, 1996; Gold & Bentler, 2000). When data are nonnormal, ML methods are still preferable, although they should be used with robust standard

errors and rescaled chi-square statistics (e.g. Enders, 2001 and the present study), or with the Bollen-Stine and naïve bootstraps (Enders, 2001). Of course, this recommendation comes from a study that investigated only two sample sizes, only two missing-data proportions, specific values of univariate skew and kurtosis, and one population model. There is, as usual, no guarantee that similar results would be obtained under different conditions. Also, the study reported here did not evaluate the performance of missing-data methods under conditions in which data are missing due to an NMAR mechanism. Still, it would be quite surprising if ADFAC, which was developed under the assumption of an MCAR mechanism, performed any better with an NMAR mechanism than it has with the MCAR and MAR mechanisms used in the present study.

REFERENCES

Arbuckle, J.L. (1996). Full information estimation in the presence of incomplete data. In G.A. Marcoulides & R.E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 243-277). Mahwah, NJ: Erlbaum.

Arbuckle, J.L. (1997). *Amos users' guide version 3.6*. Chicago, IL: SmallWaters Corporation.

Arminger, G., & Sobel, M.E. (1990). Pseudo-maximum likelihood estimation of mean and covariance structures with missing data. *Journal of the American Statistical Association, 85*(409), 195-203.

Bentler, P.M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software, Inc.

Bernaards, C.A. (2000). *Nonresponse and factor analysis, with applications to rating scale data*. PhD dissertation. Utrecht: Universiteit Utrecht, Faculteit Sociale Wetenschappen.

Brown, R.L. (1994). Efficacy of the indirect approach for estimating structural equation models with missing data: A comparison of five methods. *Structural Equation Modeling, 1*(4), 287-316.

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

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

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological), 39*(1), 1-38.

Enders, C.K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods, 6*(4), 352-370.

Enders, C.K., & Bandalos, D.L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*(3), 430-457.

Finkbeiner, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika 44*(4), 409-420.

Gleason, T.C., & Staelin, R. (1975). A proposal for handling missing data. *Psychometrika, 40*(2), 229-252.

Gold, M.S., & Bentler, P.M. (2000). Treatments of missing data: A Monte Carlo comparison of RBHDI, iterative stochastic regression imputation, and expectation-maximization. *Structural Equation Modeling, 7*(3), 319-355.

Graham, J.W., Hofer, S.M., & MacKinnon, D.P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research, 31*(2), 197-218.

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

Jamshidian, M., & Bentler, P.M. (1999). ML estimation of mean and covariance structures with missing data using complete data routines. *Journal of Educational and Behavioral Statistics, 24*, 21-41.

Jöreskog, K.G., & Sörbom, D. (1993). *PRELIS 2 user's reference guide*. Chicago: Scientific Software International.

Kauermann, G., & Carroll, R.J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association, 96*, 1387-1396.

Lee, S.-Y. (1986). Estimation for structural equation models with missing data. *Psychometrika, 51*, 93-99.

Little, R.J.A., & Rubin, D.B. (1987). *Statistical analysis with missing data*. New York: John Wiley & Sons.

Little, R.J.A., & Rubin, D.B. (1989/1990). The analysis of social science data with missing values. *Sociological Methods and Research, 18*(2 & 3), 292-326.

Little, R.J.A., & Schenker, N. (1995). Missing data. In G. Arminger, C.C. Clogg, & M.E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 39-75). New York: Plenum Press.

Marsh, H.W. (1998). Pairwise deletion for missing data in structural equation models: nonpositive definite matrices, parameter estimates, goodness of fit, and adjusted sample sizes. *Structural Equation Modeling, 5*(1), 22-36.

Rubin, D.B. (1991). EM and beyond. *Psychometrika, 56*(2), 241-254.

Satorra, A., & Bentler, P.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. voy Eye & C.C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.

Schafer, J.L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

Van Praag, B.M.S., Dijkstra, T.K., and Van Velzen, J. (1985) Least-squares theory based on general distributional assumptions with an application to the incomplete observations problem. *Psychometrika, 50*, 25-36.

Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In T.D. Little, K.U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches and specific examples* (pp. 219-240, 269-281). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Yuan, K.-H., & Bentler, P.M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology 2000* (pp.165-200). Washington, D.C.: American Statistical Association.

Table 1

Variance, Skew, and Kurtosis Values Assigned to Each Simulated Variable

Simulated Variable	Variance	Skew	Kurtosis
F1	1.00	0.0	0
E1	.36	-0.5	0
E2	.51	1.5	4
E3	.64	-2.0	8
E4	.75	0.0	-1
E5	.36	-1.0	3
E6	.51	2.0	7
E7	.64	0.5	1
E8	.75	-1.5	5
E9	.36	0.0	2
E10	.51	1.0	6
E11	.64	-0.5	0
E12	.75	1.5	4
E13	.36	-2.0	8
E14	.51	0.0	-1
E15	.64	-1.0	3
E16	.75	2.0	7
D2	.75	0	0
D3	.92	0	0
D4	.91	0	0

Table 2

Means (and Standard Deviations) of RMSDs in Parameter Estimates: MCAR Metadata

	Adjusted Expectation-Maximization				Asymptotically Distribution-Free Available-Case			
	Variances		Regression Coefficients		Variances		Regression Coefficients	
Approximate Missing Data Proportion →	.15	.30	.15	.30	.15	.30	.15	.30
Sample Size								
N = 500	.143 (.029)	.161 (.030)	.141 (.040)	.155 (.043)	.219 (.028)	.269 (.042)	.176 (.042)	.214 (.052)
N = 5000	.045 (.009)	.054 (.009)	.042 (.012)	.047 (.013)	.056 (.010)	.070 (.012)	.046 (.013)	.057 (.014)

Table 3

Means (and Standard Deviations) of RMSDs in Standard Error Estimates: MCAR Metadata

	Adjusted Expectation-Maximization				Asymptotically Distribution-Free Available-Case			
	Variances		Regression Coefficients		Variances		Regression Coefficients	
Approximate Missing Data Proportion →	.15	.30	.15	.30	.15	.30	.15	.30
Sample Size								
$N = 500$	.292 (.049)	.352 (.043)	.206 (.027)	.286 (.029)	.322 (.028)	.346 (.028)	.319 (.034)	.374 (.044)
$N = 5000$	.174 (.024)	.272 (.018)	.142 (.013)	.250 (.014)	.126 (.020)	.109 (.017)	.089 (.012)	.094 (.018)

Table 4

Mean Empirical Standard Deviations of Parameter Estimates: MCAR Metadata

	Adjusted Expectation-Maximization				Asymptotically Distribution-Free Available-Case			
	Variances		Regression Coefficients		Variances		Regression Coefficients	
Approximate Missing Data Proportion →	.15	.30	.15	.30	.15	.30	.15	.30
Sample Size								
N = 500	.085	.094	.067	.076	.090	.114	.085	.113
N = 5000	.026	.031	.020	.024	.029	.037	.023	.029

Table 5

Rate of Rejection of the Null Hypothesis under Nominal Alpha = .05 and the Means and Variances of Chi-Square Statistics: MCAR Metadata

	Adjusted Expectation-Maximization		Asymptotically Distribution-Free Available-Case	
Approximate Missing Data Proportion →	.15	.30	.15	.30
Sample Size				
N = 500				
Rejection Rate	.03	.05	.64	.76
Chi-Square Mean	92.56	99.02	126.56	139.98
Chi-Square Variance	220.95	178.12	490.02	467.34
N = 5000				
Rejection Rate	.08	.07	.08	.13
Chi-Square Mean	98.06	100.16	102.74	104.17
Chi-Square Variance	274.62	250.91	236.59	246.98

Table 6

Means (and Standard Deviations) of RMSDs in Parameter Estimates: MAR Metadata

	Adjusted Expectation-Maximization				Asymptotically Distribution-Free Available-Case			
	Variances		Regression Coefficients		Variances		Regression Coefficients	
Approximate Missing Data Proportion →	.15	.30	.15	.30	.15	.30	.15	.30
Sample Size								
N = 500	.150 (.028)	.200 (.037)	.143 (.041)	.170 (.048)	.216 (.029)	.277 (.037)	.172 (.052)	.216 (.054)
N = 5000	.067 (.014)	.139 (.015)	.049 (.012)	.080 (.012)	.077 (.011)	.160 (.018)	.052 (.012)	.097 (.015)

Table 7

Means (and Standard Deviations) of RMSDs in Standard Error Estimates: MAR Metadata

	Adjusted Expectation-Maximization				Asymptotically Distribution-Free Available-Case			
	Variances		Regression Coefficients		Variances		Regression Coefficients	
Approximate Missing Data Proportion →	.15	.30	.15	.30	.15	.30	.15	.30
Sample Size								
$N = 500$	.277 (.064)	.330 (.073)	.168 (.025)	.239 (.028)	.364 (.024)	.440 (.025)	.341 (.029)	.424 (.035)
$N = 5000$	.171 (.036)	.222 (.028)	.102 (.011)	.195 (.010)	.130 (.021)	.164 (.019)	.074 (.010)	.156 (.015)

Table 8

Mean Empirical Standard Deviations of Parameter Estimates: MAR Metadata

	Adjusted Expectation-Maximization				Asymptotically Distribution-Free Available-Case			
	Variances		Regression Coefficients		Variances		Regression Coefficients	
Approximate Missing Data Proportion →	.15	.30	.15	.30	.15	.30	.15	.30
Sample Size								
N = 500	.084	.092	.063	.071	.085	.102	.076	.091
N = 5000	.026	.029	.020	.022	.026	.031	.020	.024

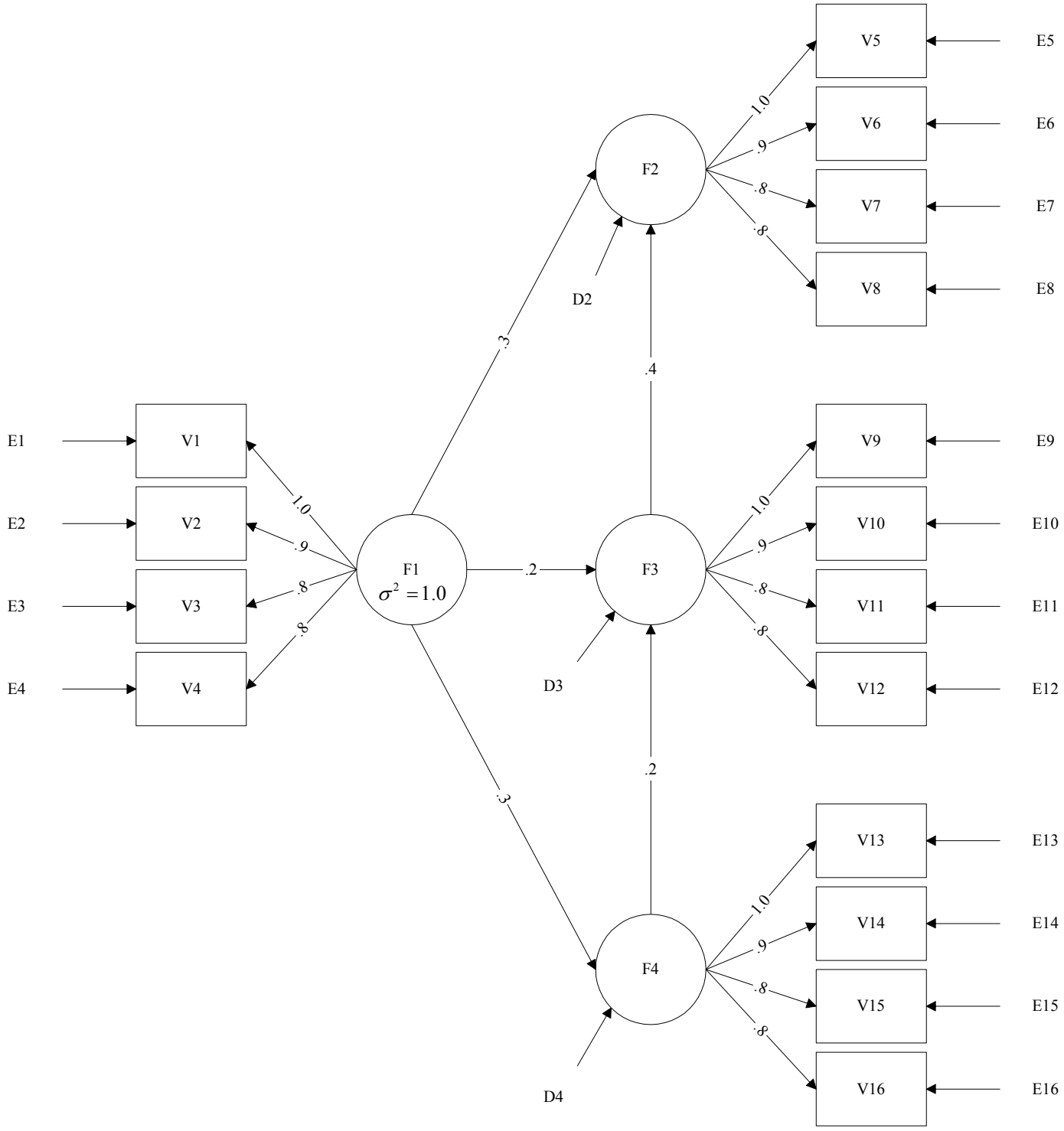
Table 9

Rate of Rejection of the Null Hypothesis under Nominal Alpha = .05 and the Means and Variances of Chi-Square Statistics: MAR Metadata

	Adjusted Expectation-Maximization		Asymptotically Distribution-Free Available-Case	
Approximate Missing Data Proportion →	.15	.30	.15	.30
Sample Size				
N = 500				
Rejection Rate	.02	.03	.70	.96
Chi-Square Mean	94.31	92.65	133.99	157.93
Chi-Square Variance	176.81	187.74	351.09	572.42
N = 5000				
Rejection Rate	.06	.11	.17	1.00
Chi-Square Mean	100.46	101.72	107.93	188.72
Chi-Square Variance	212.20	263.39	335.50	734.59

Figure Caption

Figure 1. Model used to generate the simulated data matrices for the Monte Carlo study.


E1
E2
E3
E4
V1
V2
V3
V4
1.0
.9
.8
.8
F1
$\sigma^2 = 1.0$
.3
.2
.3
F2
F3
F4
D2
D3
D4
.4
.2
V5
V6
V7
V8
V9
V10
V11
V12
V13
V14
V15
V16
E5
E6
E7
E8
E9
E10
E11
E12
E13
E14
E15
E16
1.0
.9
.8
.8
1.0
.9
.8
.8
1.0
.9
.8
.8