7-1-2001

The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models

Craig K. Enders
School of Education, University of Miami, Craig.Enders@asu.edu

Deborah L. Bandalos
University of Nebraska - Lincoln

The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models

Craig K. Enders

School of Education
University of Miami

Deborah L. Bandalos

Educational Psychology Department
University of Nebraska Lincoln

A Monte Carlo simulation examined the performance of 4 missing data methods in structural equation models: full information maximum likelihood (FIML), listwise deletion, pairwise deletion, and similar response pattern imputation. The effects of 3 independent variables were examined (factor loading magnitude, sample size, and missing data rate) on 4 outcome measures: convergence failures, parameter estimate bias, parameter estimate efficiency, and model goodness of fit. Results indicated that FIML estimation was superior across all conditions of the design. Under ignorable missing data conditions (missing completely at random and missing at random), FIML estimates were unbiased and more efficient than the other methods. In addition, FIML yielded the lowest proportion of convergence failures and provided near-optimal Type 1 error rates across both simulations.

Missing data is a common problem for researchers who utilize structural equation modeling (SEM) techniques. Applied researchers have traditionally relied on ad hoc procedures that have no theoretical justification; listwise deletion and pairwise deletion appear to be most popular (Marsh, 1998; Roth, 1994). Theory-based maxi-

---

mum likelihood (ML) approaches for treating missing data have been known in the technical literature for some time and have recently begun to appear in statistical packages. However, the extent to which the theoretical benefits associated with these methods may be realized in practice is unclear. The goal of this study is to examine the performance of one such method, full information maximum likelihood (FIML) estimation.

Before reviewing the extant missing data literature, it is useful to discuss the mechanisms that lead to missing data, as assumptions about the missing data mechanism differ among various methods. Rubin (1976) provided a classification system for missing data mechanisms and argued that missing data can be ignored (i.e., unbiased estimates can be obtained) under two conditions: missing completely at random (MCAR) and missing at random (MAR). According to Rubin, missing data are MCAR when the missing values on a variable Y are independent of other observed variables as well as the values of Y itself. When this occurs, the observed values of Y are simply a random subsample of the hypothetically complete data. However, MCAR is a stringent assumption that may not be tenable in practice (Muthén, Kaplan, & Hollis, 1987). The MAR assumption provides a second, less restrictive condition under which missing data can be ignored. Under the MAR condition, the probability that an observation is missing on variable Y can depend on another observed variable but not on the values of Y itself. The MAR assumption is less restrictive in the sense that the observed values need not be a simple random sample of the hypothetically complete data set. For example, suppose that two exams, X and Y, are administered to a group of examinees. After administering exam X, it is found that low-scoring respondents have a tendency to drop out or refuse to take exam Y. Thus, the propensity to complete exam Y depends only on scores from exam X and is unrelated to performance on Y for any group of individuals with identical scores on exam X.

MISSING DATA TECHNIQUES

The following discussion is not an exhaustive list of the available missing data techniques and focuses only on those methods examined in this study. Perhaps the most notable exclusion is mean imputation. Regression and SEM studies have equivocally demonstrated that mean imputation results in biased parameter estimates under both MCAR and MAR (Brown, 1994; Wothke, 2000), so the method was not examined in this study. Wothke summarized this situation, stating that mean imputation yielded "very precise estimates of exactly the wrong parameter."

Listwise Deletion

Listwise deletion utilizes only those cases that are complete on all variables; observations with any missing values are discarded. Although the method has desirable

qualities such as ease of implementation and comparability of univariate statistics, it can waste a great deal of complete data, resulting in inefficient parameter estimates. It is important to note that listwise deletion requires MCAR data and can yield biased parameter estimates when this condition does not hold. To illustrate, consider the situation described previously involving the administration of two exams, X and Y. If the missing values on exam Y follow an MCAR pattern, the listwise deletion of cases would yield a random sample of the hypothetically complete data set, and only efficiency would be compromised. However, if missing values on exam Y are dependent on exam X scores, the listwise deletion of cases would truncate the marginal distributions of both X and Y and would likely result in biased parameter estimates.

Past research on listwise deletion is largely consistent with theoretical expectations. For example, listwise deletion has been shown to yield unbiased parameter estimates under MCAR (Arbuckle, 1996; Brown, 1994; Wothke, 2000) and biased estimates under MAR (Arbuckle, 1996; Muthén et al., 1987; Wothke). Also, listwise deletion has been shown to yield less efficient parameter estimates than other methods (Arbuckle, 1996; Wothke).

Pairwise Deletion

Pairwise deletion attempts to utilize all available data by discarding cases on a variable by variable basis; each element within a covariance matrix is calculated separately, and only those cases with missing values on a particular bivariate pair are discarded. Research on pairwise deletion has yielded equivocal findings. For example, Brown (1994) found biased parameter estimates under MCAR, whereas more recent work by Marsh (1998) reported unbiased estimates. In addition, Brown reported that pairwise deletion rejection rates were below the nominal 5% value, whereas Kaplan (1995) and Marsh found chi-square values to be positively biased. Brown's findings are somewhat puzzling in light of statistical theory. As pointed out by Bollen (1989) and Kaplan, Wishart distribution assumptions are violated to some extent when using pairwise deletion, and, as a result, the goodness-of-fit test statistic may not follow the appropriate central chi-square distribution.

Like listwise deletion, pairwise deletion requires the MCAR assumption and should yield biased parameter estimates under MAR; empirical research has suggested that this is the case (Arbuckle, 1996; Muthén et al., 1987; Wothke, 2000). The mechanism that leads to bias under MAR is identical to that described previously for listwise deletion. However, because the deletion of cases is performed separately for each variable pair, one might expect the resulting bias to be restricted to only those parameter estimates that involve missing observations, whereas bias due to listwise deletion might be propagated throughout the model, particularly when variables are highly correlated.

Although pairwise deletion and listwise deletion both exhibited bias under MAR, Muthén et al. (1987) noted differential performance between the two methods in confirmatory factor analysis (CFA) and full structural models. Pairwise deletion outperformed listwise deletion in terms of CFA model bias, whereas the reverse was true for the full structural model. Finally, the regression literature suggests that pairwise deletion may provide optimal performance when the level of association among variables is low (Haitovsky, 1968; Kim & Curry, 1977; Little, 1992). However, this has not been systematically explored in the SEM literature.

Similar Response Pattern Imputation

Jöreskog and Sörbom (1993b) recently introduced similar response pattern imputation in the PRELIS 2 computer program. Briefly, the method attempts to impute values from another case with similar observed values and does so using a minimization criterion on a set of matching variables. If no observation exists that has complete data on the set of matching variables, imputation does not take place for a case i. However, if such a case is found, the missing value on variable X is imputed with the observed value from the case with the similar response pattern. The reader is encouraged to consult Jöreskog and Sörbom for the computational details. Similar response pattern imputation has no theoretical rationale, and very little is known about the method and the conditions under which it might be successfully applied. To date, Brown (1994) is the only study to investigate the performance of the method. Under MCAR, Brown's results were generally favorable toward similar response pattern imputation; little bias was observed across the structural parameters of the model. However, Type 1 error rates were inflated somewhat.

ML Estimation

Theory-based ML approaches for dealing with missing data have been known in the technical literature for some time (Anderson, 1957; Dempster, Laird, & Rubin, 1977; Finkbeiner, 1979; Hartley & Hocking, 1971; Wilks, 1932). Because the methods require the less restrictive MAR assumption, unbiased parameter estimates should result under both MCAR and MAR. In addition, ML methods should yield more efficient estimates than listwise and pairwise deletion under MCAR. Despite the theoretical benefits, ML methods have not been widely utilized, perhaps due to the lack of statistical software integrating the approach (Wothke, 2000). Both Allison (1987) and Muthén et al. (1987) demonstrated how to implement ML estimation using the LISREL software package (Jöreskog & Sörbom, 1983), but the method proved impractical for most applied settings.

ML approaches for dealing with missing data have recently become available in a number of SEM programs; this study investigated the FIML approach found in the AMOS program (Arbuckle, 1995). It should be noted that the expectation-maximization (EM) algorithm, another ML algorithm, is also available in current software packages (e.g., SPSS); see Little and Rubin (1987) and Schafer (1997) for a thorough discussion of this approach. Briefly, the FIML approach computes a casewise likelihood function using only those variables that are observed for case i. Assuming multivariate normality, the casewise likelihood of the observed data is obtained by maximizing the function

$$\log L_i = K_i - \frac{1}{2}\log|\Sigma_i| - \frac{1}{2}(x_i - \mu_i)' \Sigma_i^{-1}(x_i - \mu_i) \tag{1}$$

where K_i is a constant that depends on the number of complete data points for case i, x_i is the observed data for case i, and μ_i and Σ_i contain the parameter estimates of the mean vector and covariance matrix, respectively, for the variables that are complete for case i. The casewise likelihood functions are accumulated across the entire sample and maximized as follows.

$$\log L(\mu, \Sigma) = \sum_{i=1}^{N} \log L_i \tag{2}$$

The reader is encouraged to consult Arbuckle (1996) for further computational details.

From Equation 1, it should be clear that all available data are utilized during parameter estimation; a case i contributes to the estimation of all parameters for which there are complete data. However, it may not be obvious that the inclusion of data from partially complete cases contributes to the estimation of parameters that involve the missing portion of the data as well. To illustrate, consider the bivariate example described previously, where exam Y scores were missing for individuals who scored poorly on exam X. Clearly, the inclusion of partially complete data (i.e., data from the exam Y attributers) in the previous likelihood equation contributes to the estimation of X parameters, but it also contributes to the estimation of Y parameters via the correlation between X and Y. That is, probable values for the missing Y data points are implied by the observed X values, and the inclusion of the partially complete data increases the precision and accuracy of Y parameter estimates. Although it is important to note that the FIML algorithm does not impute missing values, this borrowing of information from the observed portion of the data is conceptually analogous to replacing missing Y data points with the conditional expectation of Y given X.

The inclusion of partially recorded cases into the likelihood function has important implications for parameter estimation and provides the basis for the theoreti-

cal advantages of likelihood-based inference. Under MCAR, the partially observed cases serve to increase the efficiency of ML parameter estimates relative to listwise deletion; the random nature of missing data implies that accuracy (i.e., bias) will be unaffected. However, under MAR, the partially observed cases provide important information about the underlying marginal distributions of the incomplete variables and hence may reduce the bias that would result from the listwise deletion of cases.

Although technical literature has existed on ML methods for some time, little is known about the performance of these methods relative to popular ad hoc methods. Four studies are particularly relevant to this investigation. Arbuckle (1996) examined the performance of FIML using a simple CFA model, whereas Wothke (in press) did so using a latent growth curve model. It is important to note that these studies were somewhat limited in their scope; both appeared in book chapters and were likely intended as demonstrations of the FIML method rather than comprehensive Monte Carlo studies. Graham, Hofer, and MacKinnon (1996) examined the use of ML algorithms (EM, FIML, and the multiple-group approach) under planned missingness conditions, but did so only for covariance matrix elements and regression weights, not SEM. Finally, a study by Muthén et al. (1987) examined the related multiple-group ML approach and did so using both a CFA and full structural model.

The results of these studies were generally consistent with theoretical expectations. For example, ML parameter estimates were unbiased under both MCAR and MAR (Arbuckle, 1996; Graham et al., 1996; Muthén et al, 1987; Wothke, 2000), whereas listwise and pairwise estimates were biased under MAR; listwise deletion was not included in the Graham et al. (1996). In addition, ML parameter estimates were more efficient than those of listwise and pairwise deletion (Arbuckle, 1996; Graham et al., 1996; Wothke). Although the efficiency of FIML was far superior to listwise deletion, conflicting results were observed relative to pairwise deletion; Arbuckle (1996) reported dramatic efficiency gains relative to pairwise deletion, whereas Graham et al. and Wothke reported minimal gains. Furthermore, the performance of FIML varied substantially across the six factor loadings in Arbuckle's (1996) model. Although not examined by previous research, Arbuckle suggested that the relative efficiency gains realized by using FIML might, in part, depend on the level of association among the observed variables. Also, Arbuckle's results might have been affected by the use of real, rather than simulated, data; he correctly noted that the effects of nonnormality and model misfit might have influenced results. Finally, previous studies have not examined the impact of FIML on convergence failures and model fit. However, Kaplan (1995) examined the chi-square test for the related multiple-group approach and found that the mean, variance, and rejection rates of the empirical chi-square distribution closely matched those of the appropriate central chi-square distribution.

PURPOSE

The purpose of this study is to investigate the performance of FIML relative to other widely available missing data techniques for use with structural equation models. Specifically, FIML was compared to both listwise and pairwise deletion, as well as Jöreskog and Sörbom's (1993b) similar response pattern imputation method. A Monte Carlo study was designed to address four research questions: how do the four missing data methods differ with respect to

1. Proportion of convergence failures.
2. Parameter estimate bias.
3. Parameter estimate efficiency.
4. Model goodness of fit.

METHOD

Design

The simulation study utilized two 9-variable structural equation models (Figures 1 and 2). In the first, a three-factor CFA model was specified with three indicators per latent variable, each indicator loading on only a single factor. The correlations among the three latent variables were set at .40, and factor loadings (λ) were uniformly set at one of three values (.40, .60, and .80). Uniqueness terms were equal to one minus λ^2. As seen in Figure 2, this CFA model was also parameterized as an equivalent full structural model with three latent variables. The choice to examine two different parameterizations of the same model was influenced primarily by Muthén et al. (1987), who found that listwise and pairwise deletion yielded different amounts of bias under MAR for CFA and full structural models. The study design consisted of three between-subjects factors and included four levels of sample size (100, 250, 500, and 750), five levels of missing data (2%, 5%, 10%, 15%, and 25%), and three levels of factor loading magnitude (.40, .60, and .80).

Data Generation

A population covariance matrix was generated from the three-factor CFA model described previously. Two hundred fifty normally distributed raw data matrices ($n \times 9$) were generated within each of the 60 between-subjects design cells using the RANNOR function in the Statistical Analysis Systems IML procedure. These uncorrelated vectors of random normal variates were linearly transformed to the desired covariance structure using Cholesky factorization.
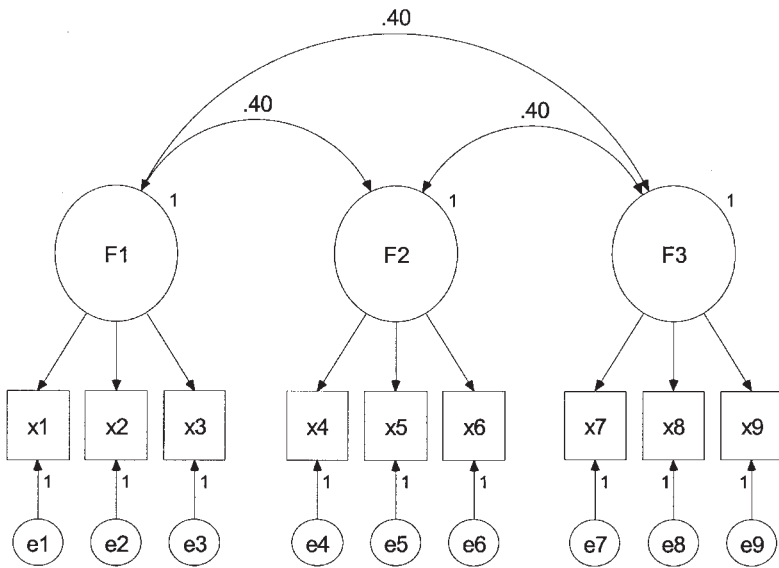
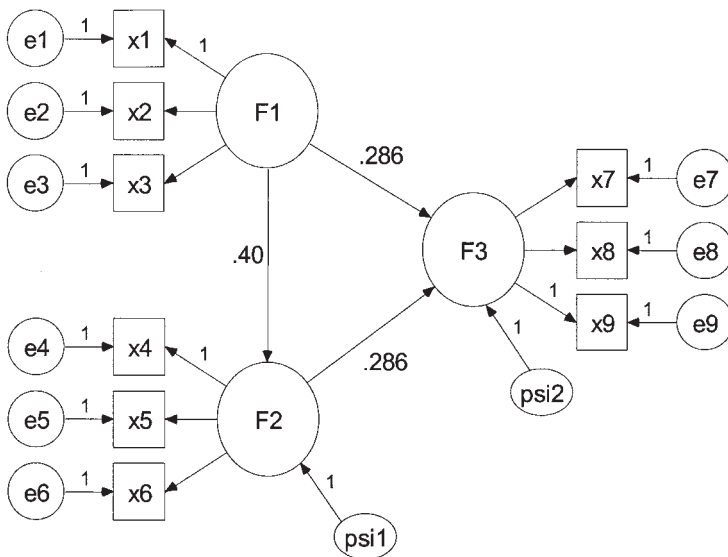FIGURE 1   Confirmatory factor analysis model used for simulation.

FIGURE 2   Full structural model used for simulation.

Following data generation, missing values were created under the MCAR and MAR mechanisms. MCAR data were simulated by randomly deleting a specified percentage of observations from each of the raw data matrices. To accomplish this, the $(n \times 9)$ raw data matrices were paired with corresponding matrices of uniform random numbers, and these random numbers were used to select observations for deletion according to the five missing data rates described previously. The deletion process was performed in such a way that all nine observed variables had exactly the same proportion of missing observations. In contrast, missing values under MAR were restricted to three observed variables: $X7$, $X8$, and $X9$. Missingness on these three variables was dependent on the values of the remaining six observed variables, but not on the values of $X7$, $X8$, and $X9$ themselves. Using a method similar to that of Muthén et al. (1987), a selection variable was calculated as a weighted composite of the six completely observed variables such that the indicators of the first latent variable were twice as influential as those of the second latent variable.

$$s = 2/3 \ (X1 + X2 + X3) + 1/3 \ (X4 + X5 + X6) \qquad (3)$$

Under the MAR mechanism, $X7$, $X8$, and $X9$ values were deleted for observations whose percentile rank on the selection variables was less than or equal to the desired percentage of missing data. This procedure could simulate the situation in which participants are administered three batteries of tests, and missing values on the third battery resulted for those examinees who performed poorly on the first two batteries. As noted previously, the resulting data were fit to both a CFA and full structural model; the same data were used for both parameterizations.

Following the deletion of observations, the PRELIS 2 computer program (Jöreskog & Sörbom, 1993b) was used to implement similar response pattern imputation. Whereas raw data was required to implement the FIML method, sample covariance matrices were generated for the other three methods using Statistical Analysis Systems. It should be noted that the pairwise deletion sample size used in this study was the minimum n of the covariance terms within each matrix. This is consistent with Brown (1994). The resulting sample covariance matrices and raw data matrices (FIML) were used as input into the AMOS 4.0 computer program, and ML estimates were obtained for all model parameters. This process was automated using a Visual Basic program that repeatedly called the AmosEngine programming interface.

Analysis

Four dependent variables were examined:

1. Proportion of convergence failures.

2. Parameter estimate bias.
3. Parameter estimate efficiency.
4. Model goodness of fit.

Nonconvergence was assessed by counting the number of models that failed to converge within a number of iterations equal to three times the number of estimated model parameters (Jöreskog & Sörbom, 1993a). Allowing for comparability with previous research, bias was expressed as a percentage of sample bias relative to the true parameter value. This was calculated as

$$\% BIAS = \left[\frac{\hat{\theta}_{ij} - \theta_i}{\theta_i} \right] * 100 \tag{4}$$

where θ_i is the true population parameter for the ith element of θ and $\hat{\theta}_{ij}$ is the corresponding parameter estimate taken from the jth iteration. The mean percentage of bias was subsequently calculated across the 250 replicates within each design cell.

Efficiency was measured using the variance of the empirical sampling distribution for each parameter estimate taken across the 250 replications in each cell of the design. Following Arbuckle (1996) and Wothke (2000), a measure of relative efficiency was calculated by forming a ratio of the sampling variances for each model parameter. This is simply

$$RE = \frac{\sigma^2_{AH}}{\sigma^2_{FIML}} \tag{5}$$

where σ^2_{AH} is the sampling variance of a specified parameter from one of the three ad hoc methods and σ^2_{FIML} is the corresponding sampling variance of the FIML estimate.

Finally, the effects of missing data method on model fit were assessed by calculating the percentage of model rejections ($p < .05$) for each cell of the design.

RESULTS

Only models that converged within 90 iterations (i.e., three times the number of estimated parameters) and had admissible parameter estimates were considered for the analyses. In screening the parameter estimates from the admissible solutions, it was noted that a number of iterations produced solutions that would almost certainly be discarded in practice (e.g., a factor loading of 58.56). For this reason, an additional screening mechanism was implemented in an attempt to identify unreasonable solutions; any solution that included one or more parameter estimates exceeding a ±400% bias threshold were discarded. Although the choice of such a

threshold was clearly arbitrary, it was felt that parameter estimates that exceeded the ±400% threshold would be identified as problematic in most applied situations. In most cases this threshold value represented a departure of at least 10 standard deviation units from the true parameter value. An additional 1% to 2% of the replicates were screened out based on this threshold value, and roughly the same percentage of replicates were deleted from each of the estimation methods.

Simulation 1: MCAR Data

Convergence failures.   Collapsed across all cells in the design, the FIML estimation method resulted in the highest percentage of admissible solutions (90.3%) followed by pairwise deletion (89.6%), similar response pattern imputation (83.5%), and listwise deletion (70.7%). The percentage of convergence failures by loading magnitude, sample size, missing data rate, and estimation method is shown in Table 1; due to space limitations, only selected design cells are displayed. Not surprisingly, convergence failures increased as the percentage of missing data increased. However, factor loading magnitude and sample size moderated this effect. The majority of convergence problems occurred in cells that crossed the low factor loading ($\lambda = .40$) and low sample size ($n = 100$) conditions, and the frequency of inadmissible solutions decreased as factor loading magnitude and sample size increased. At the high factor loading condition ($\lambda = .80$), the missing data rate primarily impacted listwise deletion convergence, and sample size had little effect; most cells had no convergence failures. Because the vast majority of design cells had no convergence failures, additional tables for the $\lambda = .80$ condition are not presented.

Parameter estimate bias.   As would be expected under MCAR, the vast majority of cells yielded negligible levels of bias (less than ±1% to 2% of the true population parameter, on average). This was true across parameters of both the CFA and full structural models. A small number of bias values exceeded 10%, and these were primarily found in cells that crossed the low factor loading ($\lambda = .40$) and low sample size ($n = 100$) conditions; this was particularly true for full structural model factor loadings. It should be noted that a similar pattern of results was observed when running the simulations on the complete data matrices. The population covariance matrix at the low loading condition contained several covariance elements that were quite small in magnitude and was thus subject to high sampling variability. As such, it would appear that the bias in these cells was not entirely due to the estimation methods, but was at least in part due to the instability of the covariance structure in these cells.

TABLE 1
Percentage of MCAR Inadmissible Solutions by Factor Loading, Sample Size, and Missing
Data Rate

Condition			Estimation Method			
Factor Loading	Sample Size	Missing Data (%)	FIML	LD	PD	SRPI
0.40	100	2	73.6	81.6	74.4	72.0
0.40	100	5	73.2	88.4	75.6	80.4
0.40	100	10	78.4	97.6	79.6	85.2
0.40	100	15	81.2	99.2	85.2	90.8
0.40	100	25	89.2	100.0	90.4	99.2
0.40	250	2	17.2	23.2	18.0	18.4
0.40	250	5	16.4	34.4	17.2	20.4
0.40	250	10	26.0	71.2	26.0	44.4
0.40	250	15	28.4	83.2	30.4	62.4
0.40	250	25	47.6	99.6	53.2	94.4
0.60	100	2	4.0	8.8	4.4	4.4
0.60	100	5	2.8	16.4	4.0	5.2
0.60	100	10	4.8	50.8	7.6	16.0
0.60	100	15	9.2	86.4	12.0	26.4
0.60	100	25	16.0	99.6	23.6	79.2
0.60	250	2	0.0	0.0	0.0	0.0
0.60	250	5	0.0	0.0	0.0	0.0
0.60	250	10	0.0	2.0	0.0	0.0
0.60	250	15	0.0	17.2	0.0	0.4
0.60	250	25	0.4	86.8	0.4	21.2

Note.   MCAR = missing completely at random; FIML = full information maximum likelihood; LD = listwise deletion; PD = pairwise deletion; SRPI = similar response pattern imputation.

Efficiency.   As seen previously in Equation 5, the relative sampling variance was used as a basis for efficiency comparisons among estimation methods. Because the sampling variances are inversely related to sample size, the relative efficiency statistic can provide an indication of the increase in sample size required to reach the level of efficiency provided by FIML. For example, suppose that the relative efficiency of a listwise deletion factor loading was calculated at 1.76. This value suggests that the sample size for listwise deletion would have to be increased by approximately 76% to reach the level of efficiency afforded by FIML. It should be noted that the previous interpretation is limited to a specific parameter, sample size, and missing data rate.

Relative efficiency values were fairly consistent across parameters of the CFA and full structural models. As such, results from two representative parameters are presented here. Table 2 shows the relative efficiency for a selected factor loading ($X8$) and factor correlation ($F1 \leftrightarrow F3$) from the CFA model, whereas Table 3 shows relative efficiency values for the $X8$ factor loading and $F1 \rightarrow F3$ structural

TABLE 2
Relative Efficiency of Confirmatory Factor Analysis Parameter Estimates by Factor Loading, Sample Size, and Missing Data Rate
(MCAR)

Factor Loading	Sample Size	Missing Data (%)	Variance Ratio (X8 Loading)[a]			Variance Ratio (F1–F3 Correlation)[a]		
			LD/FIML	PD/FIML	SR/FIML	LD/FIML	PD/FIML	SR/FIML
0.40	500	2	1.15	0.99	0.98	0.96	0.97	1.03
0.40	500	5	1.84	1.03	1.23	1.27	1.03	1.07
0.40	500	10	2.05	1.02	1.23	1.52	0.96	1.29
0.40	500	15	2.11	1.02	1.46	2.05	0.98	1.34
0.40	500	25	7.82	1.00	2.44	2.37	1.02	1.81
0.60	500	2	1.11	1.00	1.01	1.17	1.01	1.03
0.60	500	5	1.63	1.03	1.10	1.43	1.02	1.12
0.60	500	10	2.69	1.04	1.30	1.95	1.01	1.34
0.60	500	15	3.34	1.01	1.67	2.31	1.01	1.57
0.60	500	25	8.13	1.15	2.88	4.42	1.07	2.16
0.80	500	2	1.14	1.02	1.05	1.16	1.00	1.03
0.80	500	5	1.47	1.04	1.15	1.47	1.02	1.14
0.80	500	10	2.23	1.03	1.26	2.64	1.02	1.36
0.80	500	15	3.88	1.16	1.57	3.23	1.08	1.67
0.80	500	25	10.18	1.18	2.48	5.99	1.16	2.99

Note. MCAR = missing completely at random; LD = listwise deletion; FIML = full information maximum likelihood; PD = pairwise deletion; SR = similar response.

[a]Values greater than 1.0 indicate that FIML estimates were more efficient.

TABLE 3
Relative Efficiency of SEM Parameter Estimates by Factor Loading, Sample Size, and Missing Data Rate (MCAR)

Factor Loading	Sample Size	Missing Data (%)	Variance Ratio (X8 Loading)[a]			Variance Ratio (F1–F3 Structural Path)[a]		
			LD/FIML	PD/FIML	SR/FIML	LD/FIML	PD/FIML	SR/FIML
0.40	500	2	1.19	1.06	1.00	1.26	1.00	1.03
0.40	500	5	1.55	1.05	0.91	2.49	1.01	1.25
0.40	500	10	1.61	1.11	1.14	2.04	1.00	1.34
0.40	500	15	1.86	1.03	1.36	1.95	0.99	1.41
0.40	500	25	3.08	0.96	1.94	2.69	1.17	3.16
0.60	500	2	1.21	0.99	1.01	1.18	1.01	1.07
0.60	500	5	1.60	1.02	1.09	1.59	1.04	1.12
0.60	500	10	1.62	1.02	1.02	2.61	1.04	1.27
0.60	500	15	1.94	1.04	1.49	4.17	0.97	1.50
0.60	500	25	4.58	1.11	2.07	12.86	1.15	4.34
0.80	500	2	1.12	1.02	1.05	1.06	1.05	1.07
0.80	500	5	1.46	1.11	1.09	1.35	1.04	1.19
0.80	500	10	2.34	1.11	1.29	2.13	1.11	1.17
0.80	500	15	2.68	1.14	1.61	3.43	1.21	1.41
0.80	500	25	4.92	1.09	2.47	10.24	1.24	2.65

Note. MCAR = missing completely at random; LD = listwise deletion; FIML = full information maximum likelohood; PD = pairwise deletion; SR = similar response.

[a]Values greater than 1.0 indicate that FIML estimates were more efficient.

path from the full structural model. Both tables are organized by loading magnitude, missing data rate, and estimation method. It should be noted that larger sample sizes resulted in reduced sampling variability for all estimation methods, but the resulting relative efficiency values were not impacted by sample size. As such, only the results for the $N = 500$ condition are displayed in the tables.

As seen in the tables, FIML consistently yielded small gains in efficiency relative to pairwise deletion across most conditions, typically about 5%. These differences are most evident for the parameters of the full structural model; FIML efficiency gains were slightly less for the CFA model. Although it is difficult to identify a trend, the efficiency of FIML relative to pairwise deletion is most obvious at the highest factor loading condition where high rates of missing data are present. In this case, 10% to 20% efficiency gains were observed. In contrast, large efficiency gains were observed for FIML relative to listwise deletion and similar response imputation. For both estimation methods, the relative superiority of FIML increased as the percentage of missing data increased. This pattern held across sample size and factor loading conditions and did not appear to interact with either of these factors. Clearly, the effects of missing data on efficiency were most pronounced for listwise deletion; at the 25% missing data rate the listwise deletion sample size would have to be increased by 400% to 900% to reach the level of precision provided by FIML. However, some caution is warranted in interpreting the results from the 25% missing data condition, as several of the listwise deletion efficiency estimates were based on smaller numbers of replicates due to convergence failures in these cells.

Model fit.   Using only the admissible solutions for each estimation method, the percentage of model rejections ($p < .05$) based on the chi-square statistic was calculated for each cell of the design. Because data were generated from a population model with perfect fit, a 5% value represents the nominal rejection rate. Overall, the FIML method yielded near-optimal rejection rates (6.1%), as did listwise deletion (6.8%). Similar response pattern imputation and pairwise deletion performed considerably worse, with overall rejection rates of 14.8% and 15.2%, respectively. Table 4 displays the percentage of model rejections by loading magnitude and missing data rate, collapsed across sample size conditions. As seen in the table, the missing data rate had no effect on FIML rejection rates. Values were quite stable and ranged between approximately 3% and 8%. Listwise deletion rejection rates were also fairly stable but tended to increase slightly as the percentage of missing data increased. In contrast, the effects of the missing data rate were quite evident for pairwise deletion and similar response pattern imputation; Type 1 error rates increased as the percentage of missing data increased. It is also clear that loading magnitude mediates the effects of the missing data rate for pairwise deletion; rejection rates increased as factor loading magnitude increased. Pairwise deletion

TABLE 4
MCAR (Missing Completely at Random) Model Rejection Rates by Factor Loading and
Missing Data Rate

Condition		Estimation Method			
Factor Loading	Missing Data (%)	FIML	LD	PD	SRPI
0.40	2	4.8	4.1	4.7	5.7
0.40	5	4.1	3.6	4.4	10.1
0.40	10	3.3	4.0	4.0	12.3
0.40	15	4.7	2.3	5.0	19.2
0.40	25	4.4	0.0	6.1	15.4
0.60	2	5.6	5.7	5.6	7.0
0.60	5	8.3	6.6	9.2	13.3
0.60	10	6.2	6.6	9.7	15.4
0.60	15	6.8	6.1	12.1	20.4
0.60	25	6.2	8.7	16.3	21.8
0.80	2	6.6	8.4	9.0	8.5
0.80	5	6.7	5.9	12.7	12.8
0.80	10	6.9	7.1	23.3	18.1
0.80	15	6.9	9.1	35.9	18.8
0.80	25	7.2	15.6	56.1	24.2

Note. FILM = full information maximum likelihood; LD = listwise deletion; PD = pairwise deletion; SRPI = similar response pattern imputation.

provided near-optimal rejection rates at the low factor loading condition ($\lambda = .40$), whereas rejection rates exceeded 50% at the high factor loading condition.

Simulation 2: MAR Data

Convergence failures. Consistent with previous results, FIML produced the highest percentage of admissible solutions overall (91.6%), followed by pairwise deletion (91.1%), similar response pattern imputation (89%), and listwise deletion (78.1%). In general, convergence failures for the MAR simulations were isolated to cells within the low factor loading condition ($\lambda = .40$). Minimal convergence problems occurred at the medium factor loading condition under small sample sizes ($n = 100$), and virtually all other cells in the design had 100% convergence. Table 5 shows the percentage of convergence failures by sample size, missing data rate, and estimation method; only values for the low loading condition are presented. Not surprisingly, convergence problems increased as the percentage of missing data increased. However, the effects of the missing data on convergence failures diminished as sample size increased. When sample size was 500 or larger, only listwise deletion resulted in substantial convergence problems.

TABLE 5
Percentage of MAR Inadmissible Solutions by Sample Size and Missing Data Rate

Condition			Estimation Method			
Factor Loading	Sample Size	Missing Data (%)	FIML	LD	PD	SRPI
0.40	100	2	73.2	80.8	72.8	76.4
0.40	100	5	68.8	87.6	70.4	74.0
0.40	100	10	70.4	91.6	73.2	76.0
0.40	100	15	75.2	94.0	78.0	79.2
0.40	100	25	82.8	96.8	85.2	89.6
0.40	250	2	18.4	28.4	17.6	18.0
0.40	250	5	15.2	44.4	16.4	20.4
0.40	250	10	19.2	71.2	22.4	32.0
0.40	250	15	19.2	84.0	24.0	33.2
0.40	250	25	31.2	96.0	34.0	54.8
0.40	500	2	1.2	3.2	0.8	1.2
0.40	500	5	1.2	16.0	1.2	2.4
0.40	500	10	0.8	37.2	0.8	3.2
0.40	500	15	0.8	70.4	1.6	5.2
0.40	500	25	3.2	90.8	3.6	22.8
0.40	750	2	0.0	0.0	0.0	0.0
0.40	750	5	0.0	0.0	0.0	0.0
0.40	750	10	0.0	7.6	0.0	0.0
0.40	750	15	0.0	34.0	0.0	2.8
0.40	750	25	0.8	89.2	1.2	40.8

Note.   MAR = missing at random; FIML = full information maximum likelihood; LD = listwise deletion; PD = pairwise deletion; SRPI = similar response patter imputation.

Parameter estimate bias.   Table 6 shows the mean percentage of CFA factor loading bias by factor loading magnitude, missing data rate, and estimation method. Effect size estimates for the sample size factor were generally quite small, so values in the table are collapsed over this factor. Furthermore, bias results were highly consistent across loadings of a given latent factor, so results for a single representative loading from each factor are presented. As seen in the table, CFA model factor loadings exhibited little or no bias under MAR for three of the estimation methods: FIML, pairwise deletion, and similar response pattern imputation. However, listwise deletion resulted in biased parameter estimates for the $F1$ and $F2$ loadings; the $F3$ loadings were generally unbiased. Not surprisingly, the amount of bias increased as the missing data rate increased. Muthén et al. (1987) suggested that bias values less than 10% to 15% may not be serious in most SEM contexts, and it is clear from the table that listwise deletion yielded levels of bias that far exceeded this heuristic value.

Unlike the CFA model, factor loading estimates from the full structural model were generally unbiased across the four estimation methods. Table 7 shows the

mean percentage of full structural model factor loading bias for a single loading from each of the latent factors. Tabled values are again collapsed across sample size conditions and are presented by loading magnitude, missing data rate, and estimation method. As explained earlier, a small number of larger bias values were observed in the low factor loading ($\lambda = .40$) and low sample size ($n = 100$) cells; in collapsing over the sample size factor this contributed somewhat to the slightly larger bias values in the $\lambda = .40$ cells in Table 7. This was also observed when running simulations on the complete data matrices and is likely due to the instability of the covariance structure in these cells. Nevertheless, the larger values in these cells generally did not reach problematic levels (Muthén et al., 1987). It is also important to point out that listwise deletion yielded the largest bias values in these cells.

Table 8 displays the mean percentage of bias for the CFA factor correlation estimates by loading magnitude, missing data rate, and estimation method. As before, results are collapsed across the sample size factor. As would be expected, FIML yielded unbiased parameter estimates of the factor correlations in the CFA model. In contrast, listwise deletion yielded biased estimates for all three correlation estimates, whereas pairwise deletion and similar response pattern imputation yielded biased estimates of the correlations that involved $F3$ ($F1 \leftrightarrow F3$ and $F2 \leftrightarrow F3$), the factor with missing values on its indicator variables. As seen in the table, listwise deletion resulted in the highest level of overall bias across the three parameters. Of the three biased methods, similar response pattern imputation yielded the least amount of overall bias. For those parameters that exhibited bias, the amount of bias increased as the percentage of missing data increased. It should be noted that CFA uniqueness terms were generally unbiased for all estimation methods.

Results for the structural regression weights from the full structural model followed the same pattern as the CFA model factor correlation estimates. The mean percentage of structural path bias by loading magnitude, missing data rate, and estimation method are presented in Table 9; consistent with the previous three tables, results are collapsed across sample size conditions. As seen in the table, FIML estimates were unbiased, listwise deletion yielded biased estimates of all structural parameters, and pairwise deletion and similar response pattern imputation yielded biased estimates of those paths that involved $F3$ ($F1 \rightarrow F3$ and $F2 \rightarrow F3$), the factor with missing values on its indicator variables. As before, listwise deletion yielded the highest levels of bias, followed by pairwise deletion and similar response pattern imputation, respectively.

Efficiency.   Under the MAR mechanism it can be argued that efficiency results were not as substantively interesting as the MCAR simulation due to the large amount of bias observed in the parameter estimates. Although highly efficient estimators might be desirable to some, even in the presence of some estimation bias, the magnitude of bias observed in the ad hoc methods is probably unacceptable to

TABLE 6
Mean Percentage of CFA Model Factor Loading Bias by Factor Loading and Missing Data Rate (MAR)

| Factor Loading | Missing Data (%) | Estimation Method | | | | | | | | | | | |
| | | FIML | | | LD | | | PD | | | SRPI | | |
		X2 Lo	X5 Lo	X8 Lo	X2 Lo	X5 Lo	X8 Lo	X2 Lo	X5 Lo	X8 Lo	X2 Lo	X5 Lo	X8 Lo
0.40	2	1.8	1.6	1.3	−9.8	−2.4	0.6	1.8	1.8	0.5	2.3	1.9	0.6
0.40	5	0.5	1.4	0.7	−20.3	−6.5	−1.1	1.0	1.2	−0.5	0.8	1.1	−0.2
0.40	10	1.2	0.3	−0.5	−36.4	−10.4	−1.8	1.1	0.0	−1.6	1.5	0.2	−1.8
0.40	15	2.0	1.0	0.5	−47.0	−12.4	−0.6	1.6	1.3	−1.9	0.2	2.0	−2.3
0.40	25	1.0	0.2	1.4	−62.5	−10.5	−4.6	1.0	0.0	−0.8	0.2	0.2	2.1
0.60	2	0.2	−0.7	−0.9	−6.6	−3.6	−1.7	0.3	−0.7	−1.7	0.3	−0.6	−1.5
0.60	5	−0.1	−0.4	0.0	−13.2	−5.9	−1.6	−0.1	−0.4	−1.5	0.1	−0.4	−1.2
0.60	10	−0.1	−0.5	0.9	−21.7	−9.5	−1.7	0.0	−0.6	−1.5	−0.2	−0.6	−0.9
0.60	15	−0.2	0.1	−0.2	−27.3	−10.9	−2.8	−0.3	0.1	−3.0	−0.4	0.2	−2.6
0.60	25	−0.9	−0.3	0.5	−38.1	−15.4	−3.0	−1.0	−0.5	−3.2	−0.7	−0.2	−3.5
0.80	2	−0.1	−0.5	0.0	−5.4	−3.2	−1.0	−0.1	−0.5	−1.0	−0.1	−0.5	−0.5
0.80	5	−0.4	−0.7	−0.7	−10.4	−5.8	−2.5	−0.4	−0.7	−2.5	−0.4	−0.7	−1.7
0.80	10	−0.2	−0.1	−0.3	−16.1	−7.8	−3.2	−0.2	−0.1	−3.2	−0.2	−0.1	−2.6
0.80	15	−0.4	−0.2	0.2	−20.5	−10.0	−3.4	−0.4	−0.2	−3.4	−0.4	−0.2	−2.8
0.80	25	0.1	−0.6	−0.2	−27.2	−13.6	−5.0	0.0	−0.7	−5.0	0.1	−0.6	−5.0

Note.   CFA = confirmatory factor analysis; MAR = missing at random; FIML = full information maximum likelihood; LD = listwise deletion; PD = pairwise deletion; SRPI = similar response pattern imputation.

TABLE 7
Mean Percentage of SEM Model Factor Loading Bias by Factor Loading and Missing Data Rate (MAR)

| | | Estimation Method | | | | | | | | | | | |
| | | FIML | | | LD | | | PD | | | SRPI | | |
Factor Loading	Missing Data (%)	X2 Lo	X5 Lo	X8 Lo	X2 Lo	X5 Lo	X8 Lo	X2 Lo	X5 Lo	X8 Lo	X2 Lo	X5 Lo	X8 Lo
0.40	2	5.9	8.2	6.6	7.8	8.4	5.8	6.5	9.0	6.7	8.3	8.5	6.2
0.40	5	7.2	8.5	8.5	11.5	9.3	5.5	8.0	8.6	8.5	6.1	8.9	9.7
0.40	10	5.7	6.4	9.4	–2.3	11.6	10.5	4.9	7.4	8.9	6.0	8.6	11.0
0.40	15	7.2	9.0	7.1	–22.2	15.9	13.6	6.3	10.3	6.2	4.9	9.9	10.9
0.40	25	7.1	4.5	10.6	–61.8	10.5	5.6	7.9	5.3	11.0	7.1	5.1	21.3
0.60	2	2.4	2.2	1.6	3.5	2.2	1.6	2.5	2.2	1.6	2.5	2.2	1.8
0.60	5	2.1	2.5	2.1	4.0	3.1	2.2	2.4	2.6	2.1	2.3	2.6	1.7
0.60	10	2.4	0.8	4.6	4.8	2.4	4.2	2.6	0.9	4.3	2.4	0.7	7.0
0.60	15	1.5	1.5	3.2	5.5	3.0	3.2	1.7	1.7	3.3	1.0	1.3	5.5
0.60	25	1.8	1.2	4.2	6.8	2.1	4.8	2.0	1.1	4.2	1.7	1.4	6.9
0.80	2	0.4	0.5	0.4	0.4	0.3	0.4	0.4	0.5	0.4	0.4	0.5	0.5
0.80	5	0.3	0.6	0.2	0.4	0.8	0.2	0.3	0.6	0.2	0.3	0.6	0.3
0.80	10	1.0	0.5	0.1	1.5	0.6	0.1	1.1	0.5	0.1	1.1	0.5	–0.1
0.80	15	0.3	0.6	0.5	0.9	1.1	0.5	0.3	0.7	0.6	0.3	0.6	1.0
0.80	25	0.7	0.4	0.7	1.9	0.7	0.7	0.7	0.4	0.7	0.8	0.4	2.2

Note. MAR = missing at random; FIML = full information maximum likelihood; LD = listwise deletion; PD = pairwise deletion; SRPI = similar response pattern imputation.

TABLE 8
Mean Percentage of CFA Model Factor Correlation Bias by Factor Loading and Missing Data Rate (MAR)

Factor Loading	Missing Data (%)	FIML			LD			PD			SRPI		
		F1–F2	F1–F3	F2–F3	F1–F2	F1–F3	F2–F3	F1–F2	F1–F3	F2–F3	F1–F2	F1–F3	F2–F3
0.40	2	−1.6	−4.2	−0.8	−22.5	−3.9	−3.1	−1.4	−13.1	−6.4	−1.2	−7.1	−2.7
0.40	5	0.2	−1.5	−0.9	−43.2	5.2	−3.7	0.2	−18.7	−11.7	0.8	−6.4	−5.7
0.40	10	0.3	0.2	−1.6	−95.8	7.3	−8.2	0.5	−26.3	−18.2	0.8	−10.6	−9.9
0.40	15	−1.4	−1.4	0.2	−135.8	1.9	−12.1	−1.0	−35.5	−20.5	−1.3	−17.9	−12.0
0.40	25	−0.5	−0.8	−1.8	−151.8	−67.2	−15.3	−1.4	−46.8	−28.4	−1.7	−26.5	−19.0
0.60	2	−2.1	0.6	−0.7	−16.2	−3.2	−4.4	−2.0	−7.7	−6.4	−2.0	−2.4	−3.1
0.60	5	−1.1	−0.7	−1.1	−29.1	−7.6	−8.3	−1.0	−16.2	−11.6	−1.2	−6.8	−5.0
0.60	10	−0.3	−0.7	−2.8	−50.1	−11.8	−13.8	−0.3	−25.4	−19.8	−0.5	−12.1	−10.4
0.60	15	−1.3	0.1	0.2	−70.3	−14.7	−15.1	−1.3	−31.8	−21.2	−1.5	−15.4	−11.2
0.60	25	−1.1	−0.7	−1.6	−108.8	−17.4	−21.0	−1.2	−43.8	−29.5	−1.5	−25.6	−19.7
0.80	2	0.8	−0.1	−0.6	−10.7	−5.3	−5.1	0.8	−7.2	−6.0	0.8	−2.5	−2.3
0.80	5	0.3	−0.3	−0.5	−22.8	−10.3	−9.1	0.3	−13.9	−10.8	0.3	−5.5	−4.4
0.80	10	0.9	−0.3	0.9	−37.2	−16.5	−12.4	0.9	−22.0	−15.0	0.9	−10.9	−7.1
0.80	15	−0.4	−0.1	−0.6	−52.5	−21.1	−18.0	−0.4	−28.0	−21.3	−0.4	−15.8	−12.0
0.80	25	0.0	0.2	0.2	−75.9	−29.0	−23.4	0.0	−38.1	−27.4	−0.1	−25.6	−20.7

Note. CFA = confirmatory factor analysis; MAR = missing at random; FIML = full information maximum likelihood; LD = listwise deletion; PD = pairwise deletion; SRPI = similar response pattern imputation.

TABLE 9
Mean Percentage of SEM Structural Path Bias by Factor Loading and Missing Data Rate (MAR)

| | | Estimation Method | | | | | | | | | | | | |
| | | FIML | | | LD | | | PD | | | SRPI | | |
Factor Loading	Missing Data (%)	F1–F2	F1–F3	F2–F3	F1–F2	F1–F3	F2–F3	F1–F2	F1–F3	F2–F3	F1–F2	F1–F3	F2–F3
0.40	2	3.8	−0.2	9.1	−9.9	23.5	16.7	4.4	−10.9	4.1	4.7	−2.8	5.7
0.40	5	6.5	7.9	7.7	−27.9	71.8	26.0	5.9	−18.6	0.9	6.2	−4.2	3.5
0.40	10	5.0	9.9	4.9	−83.6	112.1	56.3	4.1	−26.6	−6.3	3.3	−5.2	−3.2
0.40	15	3.1	6.7	7.9	−131.7	130.6	75.8	3.1	−42.0	−6.6	3.4	−18.3	−2.5
0.40	25	5.6	8.7	4.7	−159.4	72.6	90.8	4.7	−56.2	−10.9	5.1	−27.1	−5.6
0.60	2	−0.5	4.5	1.1	−10.2	12.1	3.9	−0.3	−6.7	−3.7	−0.4	0.1	−1.3
0.60	5	0.3	3.2	1.2	−20.1	20.1	7.7	0.6	−17.4	−7.6	0.1	−5.1	−1.6
0.60	10	2.6	1.1	−2.7	−38.9	32.5	9.7	2.8	−30.9	−16.6	2.5	−15.2	−10.6
0.60	15	0.9	2.3	1.7	−62.6	48.7	18.7	0.9	−39.7	−14.8	0.5	−18.7	−10.2
0.60	25	1.8	3.2	0.6	−112.8	93.1	33.2	2.1	−54.6	−20.1	1.4	−29.1	−17.5
0.80	2	1.1	0.6	−0.3	−7.9	2.7	0.6	1.1	−8.6	−5.5	1.1	−2.7	−2.1
0.80	5	0.3	0.6	−0.2	−18.4	5.1	1.8	0.4	−16.8	−10.1	0.3	−6.4	−4.0
0.80	10	2.1	0.2	2.0	−29.9	8.2	5.4	2.1	−28.0	−12.7	2.1	−13.6	−5.9
0.80	15	−0.1	1.8	0.1	−46.0	13.0	5.6	−0.1	−34.1	−19.1	−0.1	−18.9	−11.0
0.80	25	0.1	1.2	1.2	−71.0	20.0	9.3	0.1	−47.6	−23.5	0.0	−31.3	−20.9

Note. MAR = missing at random; FIML = full information maximum likelihood; LD = listwise deletion; PD = pairwise deletion; SRPI = similar response pattern imputation.

most applied researchers. Because FIML was the only estimator that yielded unbiased estimates under MAR, an extensive presentation of relative efficiency results is thus omitted. Consistent with MCAR results, the sampling variability of FIML parameter estimates was equal to or less than that of listwise deletion and similar response pattern imputation. However, the relative efficiency values were generally not as extreme as those observed under MCAR. For example, the mean relative efficiency values across the design cells for the three SEM structural parameters were 1.76 (SD = 1.15) and 1.27 (SD = .48) for listwise deletion and similar response pattern imputation, respectively. MCAR means across the same parameters were 3.42 (SD = 3.72) and 1.63 (SD = .91) for listwise deletion and similar response pattern imputation, respectively. Pairwise deletion results were quite similar to those observed under MCAR; the mean relative efficiency value across the MAR design cells for the three SEM structural parameters was .97 (SD = .11) compared to 1.04 (SD = .08) under MCAR.

Model fit.   Taken across all cells of the design, the FIML estimation method yielded near-optimal rejection rates (5.6%), as did listwise deletion (5.9%) and, to a lesser extent, pairwise deletion (6.8%). Similar response pattern imputation performed considerably worse, with rejection rates of 26.1%.

TABLE 10
Missing at Random (MAR) Model Rejection Rates by Factor Loading and
Missing Data Rate

Condition		Estimation Method			
Factor Loading	*Missing Data (%)*	*FIML*	*LD*	*PD*	*SRPI*
0.40	2	5.2	4.7	5.4	6.8
0.40	5	4.4	4.5	4.4	7.8
0.40	10	4.1	3.5	3.4	13.5
0.40	15	4.7	3.8	3.6	25.3
0.40	25	2.4	3.8	2.5	55.5
0.60	2	5.0	5.7	5.3	6.7
0.60	5	7.2	6.0	7.7	12.4
0.60	10	5.4	4.9	5.7	20.1
0.60	15	6.4	7.0	6.7	34.1
0.60	25	6.3	5.7	7.4	61.7
0.80	2	6.5	7.0	7.9	8.2
0.80	5	6.9	7.7	8.5	12.4
0.80	10	6.2	6.5	9.4	26.5
0.80	15	5.3	5.7	8.8	39.2
0.80	25	6.3	6.1	10.9	64.5

Note.   MAR = missing at random; FIML = full information maximum likelihood; LD = listwise deletion; PD = pairwise deletion; SRPI = similar response pattern imputation.

Table 10 displays the percentage of model rejections by loading magnitude, missing data rate, and estimation method; results are collapsed across the sample size factor. Consistent with MCAR results, FIML rejection rates were unaffected by the missing data rate and were quite close to the nominal 5% level. Listwise and pairwise deletion rejection rates were also fairly stable and were not dramatically different from the optimal rate. As seen in the table, pairwise deletion rejection rates increased slightly as factor loading magnitude increased. Finally, similar response pattern rejection rates were heavily impacted by the missing data rate; Type 1 errors increased dramatically as the percentage of missing data increased. At the highest missing data rate (25%), Type 1 error rates exceeded 60%.

DISCUSSION

The convergence failures observed in this study are not surprising in light of previous research. Sample sizes of 100 or less have been shown to result in high rates of nonconvergence (Anderson & Gerbing, 1984; Boomsma, 1985), and this was the case in this study, even when there was no missing data. The fact that listwise deletion consistently resulted in convergence failures at high missing data rates was simply due to the reduction in the overall sample size. For example, the average listwise deletion sample size was less than 60 cases in cells that crossed the highest missing data rate (25%) and highest sample size ($n = 750$) conditions. Interestingly, Brown (1994) reported no convergence failures with sample sizes of 500, whereas high rates of nonconvergence occurred in this study under this sample size. These conflicting results can likely be explained by differences in the study conditions; Brown's loadings ranged between .80 and .90, and the highest missing data rate was 16%. No convergence problems were observed in this study under similar conditions ($\lambda = .80$ and 15% missing data).

Bias results for FIML, listwise, and pairwise deletion were consistent with theoretical expectations and past research (Arbuckle, 1996; Muthén et al., 1987; Wothke, 2000); the three methods were unbiased under MCAR, whereas only FIML was unbiased under the MAR mechanism. Consistent with Muthén et al. (1987), bias was generally confined to the structural parameters. As the authors pointed out, these parameters are generally the primary interest in latent variable models. The Muthén et al. study found that pairwise deletion estimates obtained from a CFA model were generally less biased than those of listwise deletion. The reverse was true for the equivalent full structural model; listwise deletion resulted in less bias than pairwise deletion. Results from this study did not appear to support this earlier finding, as listwise deletion resulted in greater amounts of bias than pairwise deletion in both CFA and SEM models. However,

it should be noted that differences in the MAR simulation procedure might have contributed to this discrepancy. Consistent with Brown (1994), the similar response pattern method yielded unbiased parameter estimates under MCAR. However, biased parameter estimates were observed under MAR, although the bias was generally less than that of listwise and pairwise deletion. This finding is a new contribution to the literature, as no previous studies have examined the behavior of this method under MAR.

Efficiency results were generally consistent with theoretical expectations. As would be expected under MCAR, FIML yielded more efficient estimates than either listwise or pairwise deletion. Although there were no theoretical expectations regarding the efficiency of similar response pattern imputation, it is probably not surprising that the method was less efficient than FIML. The performance of pairwise deletion was somewhat surprising in light of Arbuckle's (1996) study, which found FIML estimates to be twice as efficient as those of pairwise deletion in many cases. Our results suggested that 5% efficiency gains are common, and 10% to 20% gains might be possible in situations where factor loadings are strong ($\lambda = .80$). These results lend some support to Arbuckle's suggestion that larger efficiency gains might result when indicators have a large amount of shared variance, although this certainly does not account for the discrepancy between the results. In contrast, these efficiency results are quite consistent with Wothke (in press), who reported modest efficiency gains relative to pairwise deletion; the magnitude of these gains was quite similar to those presented here.

Unlike the three ad hoc estimation methods, FIML rejection rates were unaffected by the missing data rate across the two simulations. Based on these results, it would appear that FIML estimation results in near-optimal and stable model rejection rates. It should be noted that no previous studies have examined FIML model fit. Although MCAR rejection rates for pairwise deletion were quite different from those reported by Brown (1994), these findings are consistent with statistical theory and other past research (Kaplan, 1995; Marsh, 1998). Consistent with Kaplan's (1995) conclusions, our results suggested that the chi-square statistic is sensitive to violations of Wishart distribution assumptions that result from the use of pairwise deletion. Furthermore, the upward bias of the test statistic appeared to increase as the magnitude of the factor loadings increased. Rejection rates for listwise deletion and similar response pattern imputation were generally consistent with past research, albeit limited (Brown, 1994).

In summary, FIML estimation was superior across all conditions of the design. FIML estimates were unbiased and more efficient than the other methods across both the MCAR and MAR simulations. In addition, FIML yielded the lowest rate of convergence failures and provided near-optimal Type 1 error rates across both simulations. When data were MCAR, pairwise deletion generally yielded satisfactory performance. Parameter estimates were unbiased, and convergence failures

and efficiency were only slightly worse than FIML. However, inflated Type 1 error rates were observed, particularly when the magnitude of factor loadings was high. However, under MAR pairwise deletion, parameter estimates were substantially biased. Both listwise deletion and similar response pattern imputation performed substantially worse than FIML across most conditions of the study. The methods yielded high rates of nonconvergence and inefficient parameter estimates relative to FIML. These efficiency differences became more pronounced as the missing data rate increased. Under MAR, both methods yielded substantially biased parameter estimates. Although listwise deletion rejection rates were acceptable, those of similar response pattern imputation were substantially inflated.

In practice, applied researchers may not have knowledge about the missing data mechanism. In the absence of such knowledge, these results would suggest that FIML might be a superior method for dealing with missing data in structural equation models. Given the current availability and ease of implementing FIML in computer packages such as AMOS, it is recommended that applied researchers reduce their reliance on ad hoc methods for dealing with missing data. Although results may not be deleteriously impacted by the use of pairwise deletion when data are MCAR, the potential for biased estimates and, as a result, erroneous conclusions is clear. Based on the results of this simulation, the use of listwise deletion and similar response pattern should be viewed with caution.

Although a comprehensive set of factors was examined, it is not known whether these results will generalize to other situations commonly encountered in applied research. For example, the effects of nonnormal data on FIML estimation are unknown. Like complete-data ML methods, FIML assumes multivariate normality. The effects of nonnormality on ML estimation are well documented in the SEM literature, and it is reasonable to expect that nonnormality would also affect FIML estimates, although the extent of this impact is unknown. Second, as with most previous research, this study focused on ignorable patterns of missing data (MCAR and MAR). Although theory would predict that ML methods are biased under nonignorable patterns of missing data, the Muthén et al. (1987) study suggested that ML methods might be less biased than ad hoc methods. Finally, it is not known whether these results will generalize to other common analyses such as regression. Missing data literature in the area of regression is quite limited by today's standards, and applied researchers outside the field of SEM could widely benefit from a comprehensive study of FIML applied to regression models.

ACKNOWLEDGMENTS

REFERENCES

Allison, P. D. (1987). Estimation of linear models with incomplete data. In C. C. Clogg (Ed.), *Sociological methodology, 1987* (pp. 71–103). San Francisco: Jossey-Bass.

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika, 49,* 155–173.

Anderson, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association, 52,* 200–203.

Arbuckle, J. L. (1995). Amos user's guide [Computer software]. Chicago: SmallWaters.

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* (pp. 243–277). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika, 50,* 229–242.

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,* 287–316.

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, 39,* 1–38.

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

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,* 197–218.

Haitovsky, Y. (1968). Missing data in regression analysis. *Journal of the Royal Statistical Society, Series B, 30,* 67–82.

Hartley, H. O., & Hocking, R. R. (1971). The analysis of incomplete data. *Biometrics, 27,* 783–823.

Jöreskog, K. G., & Sörbom, D. (1983). LISREL VI: Analysis of linear structural relationships by the method of maximum likelihood [Computer software]. Mooresville, IN: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (1993a). LISREL 8 user's reference guide [Computer software]. Chicago: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (1993b). PRELIS 2 user's reference guide [Computer software]. Chicago: Scientific Software International.

Kaplan, D. (1995). The impact of BIB spiraling-induced missing data patterns on goodness-of-fit tests in factor analysis. *Journal of Educational and Behavioral Statistics, 20,* 69–82.

Kim, J., & Curry, J. (1977). The treatment of missing data in multivariate analyses. *Sociological Methods & Research, 6,* 215–240.

Little, R. J. A. (1992). Regression with missing X's: A review. *Journal of the American Statistical Association, 87,* 1227–1237.

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

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,* 22–36.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 52,* 431–462.

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology, 47,* 537–560.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63,* 581–592.

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

Wilks, S. S. (1932). Moments and distributions of estimates of population parameters from fragmentary samples. *Annals of Mathematical Statistics, 3,* 163–195.

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