Alternative Multiple Imputation Inference for Mean and Covariance Structure Modeling

Taehun Lee
University of Oklahoma, Norman

Li Cai
University of California, Los Angeles

Model-based multiple imputation has become an indispensable method in the educational and behavioral sciences. Mean and covariance structure models are often fitted to multiply imputed data sets. However, the presence of multiple random imputations complicates model fit testing, which is an important aspect of mean and covariance structure modeling. Extending the logic developed by Yuan and Bentler, Cai, and Cai and Lee, we propose an alternative method for conducting multiple imputation–based inference for mean and covariance structure modeling. In addition to computational simplicity, our method naturally leads to an asymptotically chi-square model fit test statistic. Using simulations, we show that our new method is well calibrated, and we illustrate it with analyses of three real data sets. A SAS macro implementing this method is also provided.

Introduction

Model-based multiple imputation is widely accepted as one of the most flexible methods for handling missing data in a variety of applied research settings (Allison, 2001; Little & Rubin, 1987; Schafer, 1997). The formulas that combine point estimates and standard errors across multiple imputations have become familiar sights to educational and behavior scientists. The concept of replacing unobserved values by random draws from their proper posterior predictive distributions has not only fundamentally changed how methodologists and statisticians deal with missing observations in data collection and study design but also exerted far-reaching influence on statistical modeling and computation involving latent variables. After all, latent variables and missing data are synonymous.

As an inferential framework, multiple model-based imputation is extremely general. This generality stems from the usefulness of a *missing data formulation* popularized by the seminal paper on expectation-maximization (EM) algorithm by Dempster, Laird, and Rubin (1977). Unlike full likelihood–based methods

675

that often require significant amount of tailored development for each new specialized condition, with the availability of proper multiple imputations, the complete data modeling and estimation problems are often remarkably simple. This is most naturally illustrated by the multiple imputation methods for the treatment of survey nonresponse, but it is one of a number of contexts in which imputation is used as a critical device to obtain statistical adjustments that would otherwise be too complicated for routine application. We mention below two seemingly unrelated examples simply to highlight the fact that even though a bulk of our subsequent theoretical and empirical investigations will focus on missing observations in mean and covariance structure modeling, the approach we took can be equally palatable to other contexts in which imputations are required, even if no observation is apparently missing.

First, consider the *plausible value* methodology (see e.g., Mislevy, Beaton, Kaplan, & Sheehan, 1992), which is one of the backbones of the statistical framework employed in such large-scale educational assessment systems as the National Assessment of Educational Progress (NAEP) or Program for International Student Assessment (PISA). These assessments tend to focus on group-level (e.g., gender, ethnicity, country, etc.) inferences, but the individual students that make up the groups are usually administered few test items due to the complex study designs that are geared toward obtaining more efficient group-level inferences. This results in substantial uncertainty in the individual students' test scores. The simple aggregation of student scores yields statistically inconsistent group-level estimates. Glossing over a number of important details, at the end, the plausible values *are* multiple imputations based on a regression model that properly reflects the uncertainty in individual students' scores so that consistent group-level inferences can be drawn. Proper analysis of data sets containing plausible values requires the same tools as in multiple imputation for handling missing observations.

Next, consider Stuart and Rubin's (2008) matching method for estimating causal effects from observational data. In their approach, they proposed that one may construct a matched control group from multiple sources of control units when the original control group does not provide enough overlap with the treated group on observed covariates. Having more than one source of control units may introduce bias in estimated average treatment effect. To combat this potential bias, Stuart and Rubin relied on a regression-based multiple imputation procedure. Again, the analysis of the multiple sets of data requires the same basic statistical tools for combining multiple imputations.

Rather independently of the multiple imputation literature (though exceptions do exist, e.g., Rubin & Thayer, 1982), mean and covariance structure modeling (Browne & Arminger, 1995; Yuan & Bentler, 2007) has become one of the most widely used statistical techniques in social and behavioral sciences. However, when mean and covariance structure models must be fitted to multiply imputed data sets, a number of difficulties arise. First, the standard multiple imputation

inferential procedure of analyzing each imputation data set separately and combining the point estimates and standard errors at the very end is cumbersome at best. In practice, the researcher usually must fit a series of models to explore specification, to compare their fit, and to examine their relative substantive interpretability. In each step, a variety of statistical and heuristic indices are examined to guide the next move. If there are 20 imputation data sets, the researcher must carry 20 replications in each step of the search for model specification, a daunting task especially when some of the indices consist of entire matrices of numbers (e.g., residual correlations). Automated procedures (e.g., PROC MIANALYZE in SAS) for combining multiple imputation results may only alleviate some of the burden. Second, even setting the cumbersomeness aside, the standard procedure does not provide an overall model fit statistic, which forms the basis of model fit assessment in mean and covariance structure modeling. Current suggestions (e.g., Allison, 2001; Meng & Rubin, 1992) are either not accurate enough (see e.g., Allison, 2003) or require computations that are cumbersome and nonstandard, at least insofar as mean and covariance structure modeling is concerned. Moreover, the performances of those suggestions have not been evaluated in the context of mean and covariance structure modeling. Indeed, we view, as a serious drawback of the standard multiple imputation inferential procedure, the lack of a principled way for computing popular fit indices such as root mean square error of approximation (RMSEA; Browne & Cudeck, 1993), or incremental indices such as Tucker–Lewis index (TLI; Tucker & Lewis, 1973). Third, the current multiple imputation procedures are focused on obtaining corrected final parameter estimates and standard errors and do not encourage publishing enough intermediate results (e.g., mean vector and covariance matrix), so that the other researchers can replicate or meta-analyze the findings more easily. Cai and Lee (2009) also make this point.

In response to the technical difficulties, we propose a new two-stage procedure for conducting multiple imputation inference for mean and covariance structure modeling. We note that this research is not on how to build proper imputation models to produce multiple imputations. In the subsequent theoretical derivations, we also purposively make no distinction between the kinds of imputations involved, for example, missing observations, plausible values, adjustments to potential outcomes, and so on. We will treat the imputations as given and focus on statistical inference with the multiple imputations.

The guiding insight of our new approach is that at least for standard mean and covariance structure modeling, the combination of multiple imputations can occur in the beginning, before any structural models are even fitted. The underlying statistical theory is a direct extension of the results obtained by Yuan and Bentler (2000), Cai (2008), and Cai and Lee (2009) for the deterministic EM algorithm. The new procedure is computationally simple because once the imputations are combined, the structural equation modeler is back in the familiar territory of working directly with summary statistics such as the means and

covariances. We develop an asymptotically chi-square distributed overall model fit statistic based on Browne's (1984) residual-based test (Proposition 4) that can also be used naturally as a basis for additional fit indices.

Sometimes the use of multiple imputations is unavoidable (e.g., dealing with plausible values), but even when there is a choice, we argue that the flexibility afforded by multiple imputation can be decidedly advantageous. Full-information maximum likelihood (FIML; Anderson, 1957; Arbuckle, 1996) is an often recommended alternative estimation strategy for mean and covariance structure modeling when some *observations* are missing. Under Rubin's (1976) classification system of the types of missing data, FIML enjoys desirable large sample optimal properties when data are missing at random and the missing data mechanism is ignorable. Despite the asymptotic optimality, we show that our new procedure performs practically as well as the asymptotically optimal FIML estimator in finite sample sizes. In contrast to FIML, one of the key benefits of multiple imputation lies in the relative ease of including into the imputation model variables that are not part of the structural model, but are related to the missing data mechanism. Therefore, when data are not missing at random and the missing data mechanism is non-ignorable, which is arguably more realistic, our new procedure can easily reap the benefit of alternative multiple imputation systems that directly model the missing data mechanism without requiring significant change to how mean and covariance structure analysis is conducted in practice.

The remainder of the article is organized as follows. We will begin with some technical development to clarify the details of the new estimator and the assumptions that we are making. We will then use simulation studies to illustrate the performance of the new procedure relative to the FIML estimator and the standard multiple imputation inferential procedure in the context of missing responses. We apply the new procedure to the analysis of several well-known real data sets. We will conclude by noting the limitations of our new approach.

Notation and Existing Methods

A Mean and Covariance Structure Model

Throughout this article, we will assume that we are working with a multivariate normal data matrix $\mathbf{Y}$ with N independent rows and n variables. If the ith row of $\mathbf{Y}$ is $\mathbf{y}'_i$, then $\mathbf{y}_i$ follows a n-dimensional multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let $\boldsymbol{\omega} = [\boldsymbol{\mu}', \text{vech}(\boldsymbol{\Sigma})']'$, where the operator vech($\cdot$) stacks the $n(n+1)/2$ unique elements of a symmetric matrix. Clearly, the dimension of $\boldsymbol{\omega}$ is $d = n + n(n+1)/2$. For these moments, consider a structural model:

$$\boldsymbol{\omega}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{\theta}) \\ \text{vech}[\boldsymbol{\Sigma}(\boldsymbol{\theta})] \end{pmatrix}, \tag{1}$$

where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is a q-dimensional vector of free parameters in a subset of $\mathbb{R}^q$.

A typical example of a mean and covariance structure model is the extended factor analytic simultaneous equation model employed in the LISREL framework (Jöreskog & Sörbom, 2001). In LISREL, one may represent the measurement model for the ith case as

$$\mathbf{y}_i = \boldsymbol{\tau_\theta} + \boldsymbol{\Lambda_\theta}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i, \tag{2}$$

where $\boldsymbol{\eta}_i$ is a $p \times 1$ vector of latent common factors that are orthogonal to the n unique factors in $\boldsymbol{\varepsilon}_i$. Let the unique factors have zero means and covariance matrix $\boldsymbol{\Phi_\theta}$. The parameter matrices $\boldsymbol{\tau_\theta}$ and $\boldsymbol{\Lambda_\theta}$ contain the measurement intercepts and factor loadings, respectively. We use the subscript $\boldsymbol{\theta}$ to explicitly denote the dependence of the parameter matrices on the vector of free parameters. The structural equations are defined as

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha_\theta} + \mathbf{B_\theta}\boldsymbol{\eta}_i + \boldsymbol{\zeta}_i, \tag{3}$$

where $\boldsymbol{\zeta}_i$ is a $p \times 1$ vector of equation disturbance terms that have zero means and covariance matrix $\boldsymbol{\Psi_\theta}$, and is orthogonal to $\boldsymbol{\varepsilon}_i$ by implication. A rearrangement of Equation 3 leads to $\boldsymbol{\eta}_i = \mathbf{A_\theta}[\boldsymbol{\alpha_\theta} + \boldsymbol{\zeta}_i]$, where $\mathbf{A_\theta} = [\mathbf{I}_p - \mathbf{B_\theta}]^{-1}$. One has to assume that the difference between the identity matrix and the regression coefficient matrix $\mathbf{B}$ is invertible. When substituted into Equation 2, the reduced form equation is

$$\mathbf{y}_i = \boldsymbol{\tau_\theta} + \boldsymbol{\Lambda_\theta}\mathbf{A_\theta}[\boldsymbol{\alpha_\theta} + \boldsymbol{\zeta}_i] + \boldsymbol{\varepsilon}_i. \tag{4}$$

Taking expectations, we see that the model implies a linear mean structure

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = E(\mathbf{y}_i) = \boldsymbol{\tau_\theta} + \boldsymbol{\Lambda_\theta}\mathbf{A_\theta}\boldsymbol{\alpha_\theta}. \tag{5}$$

The implied covariance structure model is

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathrm{var}(\mathbf{y}_i) = \boldsymbol{\Lambda_\theta}\mathbf{A_\theta}\boldsymbol{\Psi_\theta}\mathbf{A}'_\theta\boldsymbol{\Lambda}'_\theta + \boldsymbol{\Phi_\theta}. \tag{6}$$

The joint mean and covariance structure model is obviously

$$\boldsymbol{\omega}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{\theta}) \\ \mathrm{vech}[\boldsymbol{\Sigma}(\boldsymbol{\theta})] \end{pmatrix} = \begin{pmatrix} \boldsymbol{\tau_\theta} + \boldsymbol{\Lambda_\theta}\mathbf{A_\theta}\boldsymbol{\alpha_\theta} \\ \mathrm{vech}[\boldsymbol{\Lambda_\theta}\mathbf{A_\theta}\boldsymbol{\psi_\theta}\mathbf{A}'_\theta\boldsymbol{\Lambda}'_\theta + \boldsymbol{\Phi_\theta}] \end{pmatrix}. \tag{7}$$

The exact form of the structural equation model is not essential here. Equation 7 is merely one of the many instantiations of the general mean and covariance structure model given in Equation 1. Other formulations (e.g., the Bentler-Weeks model) can be used. For our purposes, the model in Equation 7 is sufficient.

Estimation and Inference Under Standard Conditions

If there are no missing data in $\mathbf{Y}$ or if variables in $\mathbf{Y}$ do not involve complexities such as more than one set of imputations, we can use standard software programs

such as PROC CALIS in SAS to estimate $\boldsymbol{\theta}$ from the sample summary statistics. Let the sample estimates of the mean vector and the covariance matrix be

$$\hat{\boldsymbol{\mu}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{y}_i, \tag{8}$$

and

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{y}_i - \hat{\boldsymbol{\mu}})(\mathbf{y}_i - \hat{\boldsymbol{\mu}})', \tag{9}$$

respectively. We define $\hat{\boldsymbol{\omega}} = \left[\hat{\boldsymbol{\mu}}', \text{vech}\left(\hat{\boldsymbol{\Sigma}}\right)' \right]'$.

In practice, estimation is typically accomplished by invoking the multivariate normality assumption of $\mathbf{y}_i$ and minimizing the following maximum likelihood discrepancy function (see e.g., Yuan & Bentler, 2007):

$$F(\boldsymbol{\theta}; \hat{\boldsymbol{\omega}}) = \text{tr}\{\hat{\boldsymbol{\Sigma}}[\boldsymbol{\Sigma}(\boldsymbol{\theta})]^{-1}\} - \log|\hat{\boldsymbol{\Sigma}}[\boldsymbol{\Sigma}(\boldsymbol{\theta})]^{-1}| - n + [\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}(\boldsymbol{\theta})]'[\boldsymbol{\Sigma}(\boldsymbol{\theta})]^{-1}[\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}(\boldsymbol{\theta})]. \tag{10}$$

The minimizer of $F(\boldsymbol{\theta}; \hat{\boldsymbol{\omega}})$ is the maximum likelihood estimator of the structural parameters, and we denote it as

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} F(\boldsymbol{\theta}; \hat{\boldsymbol{\omega}}). \tag{11}$$

Under appropriate conditions, it is consistent, asymptotically normal, and asymptotically efficient. Furthermore, the statistic

$$T_F = (N-1)F(\hat{\boldsymbol{\theta}}; \hat{\boldsymbol{\omega}}), \tag{12}$$

is distributed in large samples as a central chi-square variable with $d - q$ degrees of freedom, if the model is specified correctly. The second derivative matrix of the discrepancy function can be used to construct a large sample covariance matrix of the estimates.

Conventional Multiple Imputation Estimation and Inference

If missing values are present, or if some or all variables in $\mathbf{Y}$ have more than one copy of imputations, estimation of $\boldsymbol{\theta}$ can be carried out in the following manner. First, either the researcher would produce ($M > 1$) multiply imputed data sets, denoted as $\mathbf{Y}_1, \ldots, \mathbf{Y}_M$, or the M data sets have already been pre-imputed, for example, in the case of plausible values in surveys. The mean and covariance structure model is then fitted to each imputation. The point estimates and standard errors are aggregated at the end.

In brief, consider the mth imputation (or plausible value) $\mathbf{Y}_m$, $m = 1, \ldots, M$. Based on this complete data set, we can easily compute the sample mean vector and sample covariance matrix using Equations 8 and 9. Denote the sample mean

vector and covariance matrix based on the *m*th imputation as $\hat{\boldsymbol{\mu}}_m$ and $\hat{\boldsymbol{\Sigma}}_m$, respectively. Then for each imputation, we can define a maximum likelihood discrepancy function for our mean and covariance structure model $F(\boldsymbol{\theta}; \hat{\boldsymbol{\omega}}_m)$, where $\hat{\boldsymbol{\omega}}_m = [\hat{\boldsymbol{\mu}}_{m'}', \text{vech}(\hat{\boldsymbol{\Sigma}}_m)']'$. Minimization of $F(\boldsymbol{\theta}; \hat{\boldsymbol{\omega}}_m)$ leads to the parameter estimate based on the *m*th imputation. Let us denote it as $\hat{\boldsymbol{\theta}}_m$. At the same time, we also obtain a covariance matrix of $\hat{\boldsymbol{\theta}}_m$ as a by-product. Let us denote it as $U_m = \text{var}(\hat{\boldsymbol{\theta}}_m)$.

Standard multiple imputation formulas can be used to combine the point estimates and variability information (Schafer, 1997). The final point estimate is the average of the multiple imputation estimates: $\bar{\theta} = M^{-1} \sum_{m=1}^{M} \hat{\boldsymbol{\theta}}_m$, and the multiple imputation variance approximation is $(\bar{\theta}) = \bar{U} + (1 + M^{-1})V$, where $\bar{U} = M^{-1} \sum_{m=1}^{M} U_m$ is the average of the within-imputation covariance matrices and $V = (M-1)^{-1} \sum_{m=1}^{M} (\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta})'$ is the between-imputation component. This provides the basis for conducting Wald tests involving the structural parameters in $\boldsymbol{\theta}$ as well as construction of confidence intervals.

For combining likelihood ratio goodness-of-fit test statistics, we may use the two statistics D_2 and D_3 proposed by Li, Meng, Raghunathan, and Rubin (1991) and Meng and Rubin (1992), respectively. In essence, D_2 is based on $M^{-1} \sum_{m=1}^{M} F(\hat{\boldsymbol{\theta}}_m; \hat{\boldsymbol{\omega}}_m)$, the average of the minimum fit-function value in each of the M imputed data sets, while D_3 requires additionally $M^{-1} \sum_{m=1}^{M} F(\bar{\theta}; \hat{\boldsymbol{\omega}}_m)$, the average of the fit-function values evaluated at $\bar{\theta}$, and $M^{-1} \sum_{m=1}^{M} F(\tilde{\boldsymbol{\omega}}; \hat{\boldsymbol{\omega}}_m)$, where $\tilde{\boldsymbol{\omega}} = M^{-1} \sum_{m=1}^{M} \hat{\boldsymbol{\omega}}_m$. Approximations based on the F-distribution are constructed. The approximate degrees of freedom can vary as a function of the number of imputations, the model's degrees of freedom, and the fraction of missing information.

Conventional FIML Estimator

As mentioned earlier, full-information maximum likelihood (Arbuckle, 1996) is a popular estimation method for mean and covariance structure models under ignorable missing data. In normal theory FIML estimation, an individual case's contribution to the log likelihood can be defined as

$$\log L(\boldsymbol{\theta}|\mathbf{y}_i) \propto -\frac{1}{2}\log|\mathbf{D}_i\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{D}_i'| - \frac{1}{2}[\mathbf{D}_i\mathbf{y}_i - \mathbf{D}_i\boldsymbol{\mu}(\boldsymbol{\theta})]'[\mathbf{D}_i\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{D}_i']^{-1}[\mathbf{D}_i\mathbf{y}_i - \mathbf{D}_i\boldsymbol{\mu}(\boldsymbol{\theta})],$$

(13)

where $\mathbf{D}_i$ is a selection matrix that depends on the missing data pattern in $\mathbf{y}_i$. In general, $\mathbf{D}_i$ consists only of zeros and ones. It has n columns, and as many rows as there are observed variables $\mathbf{y}_i$. It can be obtained by removing rows of an $n \times n$ identity matrix when the corresponding rows in $\mathbf{y}_i$ are missing. For instance, consider three observed variables, y_{i1}, y_{i2}, and y_{i3}. If y_{i2} is missing, then the selection matrix is

$$\mathbf{D}_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

This is obtained by removing the second row of a 3×3 identity matrix. Premultiplication of $\mathbf{y}_i$ by $\mathbf{D}_i$ results in the subset of $\mathbf{y}_i$ that is observed, and $\mathbf{D}_i\boldsymbol{\mu}(\boldsymbol{\theta})$ effectively selects the part of the mean vector for which observation i can contribute information. Similarly, $\mathbf{D}_i\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{D}_i'$ is the subset of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ that corresponds to the observed portion of $\mathbf{y}_i$. When $\mathbf{y}_i$ contains no missing values, $\mathbf{D}_i$ is an identity matrix and Equation 13 becomes the usual multivariate normal log likelihood function. The estimator is called full-information because it utilizes all available observations in the data matrix $\mathbf{Y}$. In the next step, the individual log likelihoods are summed up to obtain the log likelihood of the full sample

$$\log L(\boldsymbol{\theta}|\mathbf{Y}) \propto \sum_{i=1}^{N} \log L(\boldsymbol{\theta}|\mathbf{y}_i). \tag{14}$$

The FIML estimator directly maximizes the log-likelihood function in Equation 14 to obtain the structural model parameter estimates. Let this maximum be denoted $\breve{\boldsymbol{\theta}}$. Minus one times the inverse of the second derivative matrix of $\log L(\boldsymbol{\theta}|\mathbf{Y})$, when evaluated at $\breve{\boldsymbol{\theta}}$ provides a large sample covariance matrix of the parameter estimates. FIML estimation is implemented in major structural equation modeling software packages.

The FIML estimator also naturally leads to a chi-square distributed fit statistic based on likelihood ratio comparisons. The above FIML log likelihood can be used to obtain an unstructured/saturated maximum likelihood estimate of the mean vector and covariance matrix under missing data. Let the saturated estimate of the mean vector be denoted $\breve{\boldsymbol{\mu}}$ and the saturated covariance matrix be $\breve{\boldsymbol{\Sigma}}$. The following statistic

$$T_{\text{FIML}} = -2[\log L(\breve{\boldsymbol{\theta}}|\mathbf{Y}) - \log L(\breve{\boldsymbol{\mu}}, \breve{\boldsymbol{\Sigma}}|\mathbf{Y})] \tag{15}$$

is asymptotically distributed as a central chi-square variable with $d - q$ degrees of freedom when the structural model is correctly specified.

The New Two-Stage Estimator

The conventional multiple imputation procedure outlined previously is ideal for such modeling frameworks as linear regression analysis where the focus is on estimating and testing parameters. For reasons discussed earlier in the introduction section, it can be ill-suited to the practice of mean and covariance structure modeling. We shall call our new estimator the Multiple Imputation Two Stage (MI2S) estimator. As opposed to the standard practice of conducting separate structural analysis before combining the inferences, in MI2S, we combine the multiple imputations first and then estimate the structural parameters. By

adapting a theorem (Proposition 4) proposed by Browne (1984), an asymptotically chi-square goodness-of-fit test statistic can be constructed.

Stage One: Combining Multiple Imputations

We assume the availability of M properly imputed data sets, $\mathbf{Y}_1, \ldots, \mathbf{Y}_M$. By proper, we mean that the imputations are based on a well-constructed imputation model that includes key variables properly accounting for the missing data mechanism. In addition, if the Markov chain Monte Carlo (MCMC) method is required to produce the imputations, we assume that necessary numerical precautions have been exercised so that the imputations are drawn from an MCMC sampler after a sufficiently long burn-in period and that appropriate subsampling is taken so that the imputations are approximately uncorrelated. In other contexts, for example, plausible values, the imputations are already given, and we assume that they are proper. For the MI2S estimator in this article, we make an additional assumption that the imputations are produced by a multivariate normal imputation model (see e.g., Schafer, 1997). We note that this assumption is made when one uses the popular imputation software program SAS PROC MI.

Consider imputation m. From the complete data set $\mathbf{Y}_m$, we can easily compute estimates of the mean vector $\hat{\boldsymbol{\mu}}_m$ and the covariance matrix $\hat{\boldsymbol{\Sigma}}_m$ as per Equations 8 and 9. Writing the estimate of $\boldsymbol{\omega}$ from imputation m as $\hat{\boldsymbol{\omega}}_m = [\hat{\boldsymbol{\mu}}'_{m'}, \text{vech}(\hat{\boldsymbol{\Sigma}}_m)']'$, the combined estimate of $\boldsymbol{\omega}$ across the M imputations is the average

$$\tilde{\boldsymbol{\omega}} = \frac{1}{M} \sum_{m=1}^{M} \hat{\boldsymbol{\omega}}_m. \tag{16}$$

The above equation is nothing more than a direct application of the standard multiple imputation combination rule to the unstructured/saturated first- and second-order moments of the complete data.

In essence, Equation 16 is a stochastic counterpart of the deterministic EM algorithm for handling missing data in the multivariate normal model (Dempster et al., 1977). EM produces maximum likelihood estimates of the unstructured/saturated first and second order moments based on the observed data. Here, the estimation of $\boldsymbol{\omega}$ eschews EM in favor of multiple random imputations. We believe that this is a more flexible method than the EM algorithm. For example, when multiple sets of plausible values are present, Equation 16 easily produces the combined estimates of means and covariances, whereas the EM algorithm would be entirely useless because the data sets containing the plausible values do not have missingness.

Under multivariate normality of the complete data model, we can obtain the following estimate of the complete data Fisher information matrix based on imputation m

$$\mathfrak{I}(\hat{\boldsymbol{\omega}}_m) = \begin{pmatrix} \hat{\boldsymbol{\Sigma}}_m^{-1} & 0 \\ 0 & \frac{1}{2}\mathbf{K}_n'(\hat{\boldsymbol{\Sigma}}_m^{-1} \otimes \hat{\boldsymbol{\Sigma}}_m^{-1})\mathbf{K}_n \end{pmatrix}, \tag{17}$$

where $\mathbf{K}_n$ is the $n^2 \times n(n+1)/2$ duplication matrix as defined in Schott (1997). The covariance matrix of $\hat{\boldsymbol{\omega}}_m$ is equal to $N^{-1}[\mathfrak{I}(\hat{\boldsymbol{\omega}}_m)]^{-1}$. Again applying standard multiple imputation variance formula, we have an estimate of the covariance matrix of $\tilde{\boldsymbol{\omega}}$

$$\text{var}(\tilde{\boldsymbol{\omega}}) = N^{-1}\tilde{\Xi} = \left[\frac{1}{M}\sum_{m=1}^{M}\frac{1}{N}[\mathfrak{I}(\hat{\boldsymbol{\omega}}_m)]^{-1}\right] + \left[\frac{1+M^{-1}}{M-1}\sum_{m=1}^{M}(\hat{\boldsymbol{\omega}}_m - \tilde{\boldsymbol{\omega}})(\hat{\boldsymbol{\omega}}_m - \tilde{\boldsymbol{\omega}})'\right]. \tag{18}$$

The form of $N^{-1}\tilde{\Xi}$ has a natural interpretation. If there are no missing data, or equivalently stated, if there is no between-imputation variability due to missing information, $\tilde{\Xi}$ solely depends on the multivariate normal theory Fisher information matrix—the first component in square brackets. If there is missing information, the between-imputation component captures the added uncertainty due to missing data—the second set of square brackets. Note that Equations 17 and 18 are based on the multivariate normality assumption. Therefore, $\tilde{\Xi}$ is a model-based estimate of the covariance matrix of $\tilde{\boldsymbol{\omega}}$. This point will subsequently be important because having a model-based covariance matrix for $\tilde{\boldsymbol{\omega}}$ requires substantially smaller sample size to achieve stable estimation and inference for the structural parameters of interest.

A basic tenet of the theory behind multiple imputation is that if the imputations are proper, and if M is large, $\tilde{\boldsymbol{\omega}}$ provides a consistent estimate of $\boldsymbol{\omega}$ (Rubin, 1996). Furthermore, when M is large, the limiting distribution of $\sqrt{N}(\tilde{\boldsymbol{\omega}} - \boldsymbol{\omega})$ is normal, with zero means, and an asymptotic covariance matrix that can be consistently estimated by $\tilde{\Xi}$. Thus, similar (*in spirit*) to Cai and Lee's (2009) two-stage estimator wherein the first stage is based on the EM algorithm, we can use $\tilde{\boldsymbol{\omega}}$ directly in discrepancy function based estimation of the structural parameters in $\boldsymbol{\theta}$.

Stage Two: Estimation and Inference

More formally, for the mean and covariance structure model $\boldsymbol{\omega}\boldsymbol{\theta}$, we define the MI2S estimator of $\boldsymbol{\theta}$ as

$$\tilde{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} F(\boldsymbol{\theta}; \tilde{\boldsymbol{\omega}}), \tag{19}$$

where $F(\boldsymbol{\theta}; \tilde{\boldsymbol{\omega}})$ is the maximum likelihood discrepancy function as defined in Equation 10. Because $\tilde{\boldsymbol{\omega}}$ is a consistent estimate of $\boldsymbol{\omega}$ as M tends to infinity, and the estimating equations defined by the maximum likelihood discrepancy function is unbiased, the resulting two-stage estimator $\tilde{\boldsymbol{\theta}}$ should be consistent and asymptotically normal, which can be shown using Yuan and Jennrich's (1998) asymptotic distribution theory for generalized estimating equations.

A comparison of Equations 11 and 19 clearly reveals the fact that the multiple imputation estimate of means and covariances $\tilde{\boldsymbol{\omega}}$ serves essentially the same role as sample means and sample covariances (i.e., $\hat{\boldsymbol{\omega}}$) in standard complete data estimation. There is, however, one hidden problem. One would naturally hope that $N - 1$ times the minimum discrepancy function value

$$T_M = (N - 1)F(\tilde{\boldsymbol{\theta}}; \tilde{\boldsymbol{\omega}}) \qquad (20)$$

is similarly distributed as a chi-square variable, but as Yuan and Bentler (2000) noted, in general this is not true. Due to the presence of multiple imputations, the asymptotic covariance matrix of $\tilde{\boldsymbol{\omega}}$ is not of a standard form that is assumed in the maximum likelihood discrepancy function. Equation 18 makes this point amply clear. While the first part of $\tilde{\Xi}$ (the within-imputation variance) comes from standard multivariate normal theory, the second part (the between-imputation variance) depends on the variability of the imputations, which is a function of the fraction of missing information. Thus, in general, the discrepancy function $F(\boldsymbol{\theta}; \tilde{\boldsymbol{\omega}})$ is not correctly specified (in the sense of Browne, 1984). According to the theory developed in Yuan and Bentler (2000), when the discrepancy function is not correctly specified, though the estimator itself may still be consistent and asymptotically normal, the test statistic T_M is at best distributed as a mixture of single degree of freedom chi-square variates.

A related problem is that the standard errors based on the second derivatives of $F(\tilde{\boldsymbol{\theta}}; \tilde{\boldsymbol{\omega}})$ are also incorrect. They will in general be too small. An intuitive explanation is as follows. The standard errors are related to sample size. The larger N is, the smaller the standard errors are. However, when there is missing information, the sample size associated with $\tilde{\boldsymbol{\omega}}$ is no longer unequivocally equal to N. The means and covariances analyzed in the second stage of MI2S are based on less information than the number of cases N would suggest.

To remedy the situation, we apply Browne's (1984) Proposition 4 to the context of multiple imputation inference. Proposition 4 contains a residual-based test statistic that is asymptotically chi-square for any consistent and asymptotically normal estimator of $\boldsymbol{\theta}$. Let $\mathbf{J}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\omega}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ be the $d \times q$ Jacobian of the mean and covariance structure model. Under Browne's (1984) regularity conditions, $\mathbf{J}(\boldsymbol{\theta})$ should be of full column rank, so there exists a $d \times (d - q)$ matrix $\mathbf{J}^c(\boldsymbol{\theta})$ that is an orthogonal complement of $\mathbf{J}(\boldsymbol{\theta})$, such that $[\mathbf{J}^c(\boldsymbol{\theta})]'\mathbf{J}(\boldsymbol{\theta}) = 0$.

Under MI2S estimation, let $\boldsymbol{\omega}(\tilde{\boldsymbol{\theta}})$ be the model-implied moments. The residual moments are simply $\mathbf{e} = \tilde{\boldsymbol{\omega}} - \boldsymbol{\omega}(\tilde{\boldsymbol{\theta}})$. Furthermore, recall that $\tilde{\Xi}$ as defined in Equation 18 is a $d \times d$ symmetric matrix that consistently estimates the limiting covariance matrix of $\sqrt{N}\tilde{\boldsymbol{\omega}}$. Then, given $H_0 : \boldsymbol{\omega} - \boldsymbol{\omega}(\boldsymbol{\theta}) = \mathbf{0}$ for some $\boldsymbol{\theta}_0 \in \Theta$ versus $H_1 : \boldsymbol{\omega} - \boldsymbol{\omega}(\boldsymbol{\theta}) \neq \mathbf{0}$ for any $\boldsymbol{\theta}$, the following goodness of fit (GOF) statistic

$$T_{\text{BM}} = N\mathbf{e}'\boldsymbol{\Gamma}\mathbf{e} \qquad (21)$$

is asymptotically distributed as a central chi-square variable with $d - q$ degrees of freedom under the null hypothesis, where $\mathbf{\Gamma} = [\mathbf{J}^c(\tilde{\boldsymbol{\theta}})] \big\{ [\mathbf{J}^c(\tilde{\boldsymbol{\theta}})]' \tilde{\Xi} [\mathbf{J}^c(\tilde{\boldsymbol{\theta}})] \big\}^{-1}$ $[\mathbf{J}^c(\tilde{\boldsymbol{\theta}})]'$. We emphasize that though T_{BM} can be viewed as an application of Browne's (1984) Proposition 4, it is not the same as the asymptotically distribution free (ADF) test statistic to which the article is more famously associated. In fact, T_{BM} uses a weight matrix that is model-based, as opposed to the sample-based weight matrix in Browne's (1984) ADF statistic. As we will demonstrate using simulations, the statistic works well even when N is as small as 100.

Following a result in Browne and Arminger (1995), the limiting covariance matrix of the MI2S estimator $\sqrt{N}(\tilde{\boldsymbol{\theta}})$ can be consistently estimated as $[\mathbf{J}(\tilde{\boldsymbol{\theta}})' \tilde{\Xi}^{-1} \mathbf{J}(\tilde{\boldsymbol{\theta}})]^{-1}$, whose diagonal elements are the squared standard errors for the structural parameter estimates. Finally, note that because of the chi-squaredness of T_{BM} and the fact that $\mathbf{e}' \mathbf{\Gamma} \mathbf{e}$ does not explicitly depend on N, we can compute the familiar RMSEA statistic (Browne & Cudeck, 1993), using $\mathbf{e}' \mathbf{\Gamma} \mathbf{e}$ as a discrepancy function value. Other fit indices, for example, the TLI, can be computed similarly, given the availability of combined summary statistics in $\tilde{\boldsymbol{\omega}}$.

Simulation Studies

The simulation studies involve a series of comparisons involving the FIML estimator, the standard multiple imputation inferential procedure, with the newly developed MI2S estimator in the context of missing observations. For MI2S, the multiple imputations are generated under the multivariate normal model with the MCMC method. Specifically, we implement a data augmented Gibbs sampler described by Schafer (1997). The starting values of the Gibbs sampler are obtained by running an EM algorithm that produces the maximum likelihood estimates of the unstructured means and covariances. The burn-in period of the Gibbs sampler is set to 1,000 iterations. Throughout this section, $M = 20$ imputations are taken with a thinning interval of 200 iterations between each imputation. These numerical settings are quite conservative. For structural model estimation, we use the default sequential quadratic programming solver in GAUSS (Aptech Systems, Inc., 2003) to optimize the log likelihood or to minimize the discrepancy function. The code for data generation, EM estimation, MCMC imputation, FIML estimation, and discrepancy function–based model fitting are entirely programmed in GAUSS.

Type I Error Rates

We conduct this study to show that the proposed MI2S fit test statistic T_{BM} is indeed chi-square distributed under the null hypothesis. The data generating process is a confirmatory factor analysis (CFA) model with nine manifest variables and three correlated factors. The generating factor loading matrix is

$$\mathbf{\Lambda}' = \begin{pmatrix} \underline{1.0} & 0.7 & 0.9 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \underline{1.0} & 0.9 & 0.9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \underline{1.0} & 1.0 & 1.0 \end{pmatrix},$$

where the underlined values are considered as fixed for the scale identification of the latent factors. And the factor covariance matrix is

$$\mathbf{\Psi} = \begin{pmatrix} 0.6 & & \\ 0.4 & 0.9 & \\ 0.3 & 0.2 & 0.5 \end{pmatrix}.$$

The unique factor variances are given by $\mathbf{\Phi} = \operatorname{diag}(.7, .9, .5, .3, .4, .4, .6, .4, .5)$. For this model, $q = 21$ free parameters make up $\mathbf{\theta}$. They are the six nonzero free loadings, the six factor variances and covariances, and the nine unique variances. Single-group factor analysis models do not have an explicit mean structure, and the generating $\mathbf{\mu}$ is taken as a null vector. The covariance structure model of Equation 6 also simplifies to $\mathbf{\Sigma}(\mathbf{\theta}) = \mathbf{\Lambda}\mathbf{\Psi}\mathbf{\Lambda}' + \mathbf{\Phi}$. The degrees of freedom is equal to 24.

Complete multivariate normal data having the generating factor analysis covariance structure are first simulated. To simulate missingness, the complete data sets are then subject to two kinds of missing data mechanisms: missing completely at random (MCAR) and missing at random (MAR). For the MCAR condition, for each row in the simulated data matrix, a fair dice is rolled to decide whether there should be any missing values. Next, for a case that is chosen to contain missing observations, the values for the last three indicators are set to missing. This leaves about 16% of all observations missing. This pattern may occur in practice when a subset of individuals do not complete the questionnaire by study design (planned missingness), for example, due to the high cost of measuring all variables for all individuals.

For the MAR condition, the probability of missingness of the last three manifest variables is set to be linearly related to the mean of the first six manifest variables, say, Z. Specifically, we divided the distribution of Z into quartiles and set the missingness probabilities for the four regions to (.50, .20, .075, .025). Under this MAR condition, cases with lower values of Z have higher probabilities of missingness on the last three variables, leaving about 20% of all cases contain missing observations. As a benchmark, a condition with no missing data (NOMIS) is also included.

The three missing data conditions (NOMIS, MCAR, and MAR) are crossed with three sample sizes: $N = 100, 300, 500$, resulting in nine simulation conditions. In each condition, 1,000 replications are attempted. For each replication, both the MI2S estimator and the FIML estimator are used to fit the CFA model. Non-converged replications are discarded.

Because the fitted model is correctly specified, our hypothesis is that T_{BM} should be distributed as a central chi-square variable with 24 degrees of freedom

regardless of the amount or kind of missing data. On the other hand, the naive fit statistic T_M (Equation 20) should not be chi-square distributed under MAR or MCAR. Under MCAR and MAR, T_{FIML} is also asymptotically chi-square distributed with the same degrees of freedom as T_{BM}. Under NOMIS, both T_{FIML} and T_M reduce to T_F (Equation 12), and all statistics are distributed as central chi-square variable with 24 degrees of freedom.

Table 1 presents a summary of the results of this simulation study. First, under NOMIS, both T_{FIML} and T_{BM} are clearly behaving as they should. The means of the empirical distribution of these two test statistics are very close to 24, which is the theoretical expected value of a central chi-square variable with 24 degrees of freedom. The variances are also close to the theoretical value of 48. The empirical Type I error rates at the .01, .05, and .10 levels closely follow the nominal alpha level. While T_{FIML} tends to be more liberal at a smaller N, T_{BM} is more conservative. The performance of both T_{FIML} and T_{BM} appears to improve as sample size increases. Next, for the MCAR condition, the naive test statistic T_M is obviously not chi-square distributed. Its means and variances are too large, and the empirical rejection rates are far above the nominal level. On the other hand, both T_{FIML} and T_{BM} maintain adequate control over Type I error rates, and their mean–variance relations much better approximate that of a central chi-square variable with 24 degrees of freedom. The type I error rates of D_2 and D_3 are very close to the nominal alpha level. We observe essentially the same phenomena under the MAR condition. Our conclusion is that as far as Type I error rates are concerned, the MI2S statistic T_{BM} performs at least as well as T_{FIML}, D_2, and D_3 across the conditions examined. For relatively small sample sizes, T_{BM} may even lead to a slightly better calibrated test than T_{FIML}, D_2, or D_3 when the data are MAR.

Power to Detect Model Misspecification

In this simulation, we deliberately mis-specify the fitted model so that we may investigate the power of the MI2S statistic. The generating model is essentially the same as the model used in the previous simulation study. The only difference is that the generating model contains an extra parameter—a cross-loading for the ninth indicator on the first factor λ_{91} equaling .7. The fitted model is still a three-factor CFA model, but it omits the cross-loading as a free parameter. A relatively mild degree of misspecification is chosen so that the Monte Carlo simulation can show more nuances of the tests, especially at large N. The reference distribution of T_{FIML} and F_{BM} is still central chi-square with 24 degrees of freedom. However, because of the misspecification, the empirical distribution of the test statistics will not be central chi-square distributed.

The missing data conditions remain the same as in the previous simulation study (NOMIS, MCAR, and MAR). There are three sample size conditions: 100, 300, and 500. In each of the nine conditions, 1,000 replications are

TABLE 1
Type I Error Rates

N	Converged	Statistic	M	Var	Min	Max	Significance Level 0.01	0.05	0.10
NOMIS									
100	1000	T_{FIML}	24.781	55.356	9.561	52.877	.018	.078	.136
		T_{BM}	22.975	41.906	9.969	43.998	.002	.030	.066
300	1000	T_{FIML}	24.650	50.120	6.810	52.452	.004	.046	.124
		T_{BM}	23.973	47.337	6.814	52.472	.006	.038	.094
500	1000	T_{FIML}	23.958	51.927	8.304	59.810	.008	.050	.088
		T_{BM}	23.607	49.473	8.553	60.703	.008	.044	.078
MCAR									
100	1000	T_{FIML}	25.409	50.248	8.299	48.459	.007	.072	.145
		T_M	29.457	71.189	10.725	58.731	.070	.214	.317
		T_{BM}	23.044	40.911	7.811	45.177	.003	.028	.064
		D_2					.007	.069	.156
		D_3					.013	.058	.125
300	1000	T_{FIML}	24.100	49.869	8.103	48.459	.013	.066	.111
		T_M	28.266	72.197	8.670	60.163	.064	.157	.240
		T_{BM}	23.694	49.164	7.437	53.040	.013	.055	.095
		D_2					.014	.063	.109
		D_3					.016	.059	.110
500	1000	T_{FIML}	24.208	47.453	9.157	55.986	.012	.045	.102
		T_M	28.324	65.597	11.558	62.001	.047	.154	.249
		T_{BM}	24.245	46.042	8.661	56.345	.009	.051	.095
		D_2					.008	.050	.102
		D_3					.013	.042	.094
MAR									
100	1000	T_{FIML}	25.434	55.990	9.565	54.813	.019	.087	.158
		T_M	29.509	78.706	9.887	64.771	.089	.207	.300
		T_{BM}	23.137	46.512	8.943	50.960	.007	.045	.097
		D_2					.026	.096	.172
		D_3					.021	.091	.141
300	1000	T_{FIML}	24.557	46.944	8.864	49.969	.008	.057	.110
		T_M	28.831	67.573	10.450	59.723	.059	.174	.276
		T_{BM}	24.161	45.589	8.568	52.914	.006	.053	.098
		D_2					.006	.059	.110
		D_3					.005	.059	.105
500	1000	T_{FIML}	23.997	43.973	8.681	49.240	.008	.042	.096
		T_M	28.235	61.816	10.305	55.509	.051	.151	.229
		T_{BM}	24.055	44.915	8.357	54.484	.005	.051	.097
		D_2					.004	.052	.096
		D_3					.009	.048	.093

Note: The entries in the "Converged" column refer to the number of converged replications in each condition.

attempted. For each replication, we compare the rejection rates of T_{BM} with T_{FIML} at the .01, .05, and .10 nominal alpha levels.

Table 2 presents a summary of the simulation results. The trend is quite clear. The test based on T_{BM} is slightly less powerful than T_{FIML} and D_2 and D_3, when N is small. However, considering the mild degree of misspecification, T_{BM} shows acceptable levels of power, and the difference in power diminishes as N increases to 300 and 500.

Bias and Variability

To take a closer look at the MI2S estimator, we examine the estimated relative bias (RB) and root mean square deviation (RMSD) of the estimates from true generating parameter values. We continue to use the same CFA data generating model as in the first simulation study. The fitted model is correctly specified, so in effect, we are examining parameter recovery. For a generic parameter θ, we define estimated relative bias as

$$\text{RB} = \frac{1}{L} \sum_{l=1}^{L} \frac{\hat{\theta}_l - \theta}{\theta},$$

where θ is the true value, L is the number of Monte Carlo replications (1,000 in this case), $\hat{\theta}_l$ is the parameter estimate in replication l. We define RMSD as

$$\text{RMSD} = \sqrt{L^{-1} \sum_{l=1}^{L} (\hat{\theta}_l - \theta)^2}.$$

Table 3 presents the true parameter values, RB, and RMSD for key parameters (factor loadings and correlations) under MCAR when N is 100. Full results for other conditions are available upon request, but in summary, FIML, MI2S, and the standard multiple imputation method are all relatively unbiased in the conditions examined. The maximum absolute value of RB never exceeded 6%. Though MI2S is slightly more variable when N is small, presumably because of the additional random imputation variability, its RMSDs are different from those of the FIML and the standard multiple imputation estimators only in the third decimal place when N is large.

Applications to Real Data

In this section, we illustrate the performance of the MI2S estimator with applications to three real data sets. We wrote a SAS macro fully implementing the MI2S estimator and used it to obtain the results reported here. The macro and data sets are freely available from http://lcai.bol.ucla.edu/. Its usage is documented in Appendix A. The first two applications highlight the fact that the MI2S estimator can provide an asymptotically chi-square distributed fit test statistic,

TABLE 2
Power to Detect Model Misspecification

N	Converged	Statistic	M	Var	Min	Max	Significance Level		
							0.01	0.05	0.10
NOMIS									
100	992	T_{FIML}	32.738	165.096	11.711	225.052	.132	.294	.420
		T_{BM}	28.456	52.715	11.323	54.125	.036	.126	.238
300	1000	T_{FIML}	47.108	145.751	21.846	86.621	.608	.800	.872
		T_{BM}	41.329	89.796	20.091	74.779	.418	.686	.802
500	1000	T_{FIML}	62.331	193.208	28.528	111.770	.936	.986	.996
		T_{BM}	53.420	108.027	28.012	86.632	.842	.960	.986
MCAR									
100	1000	T_{FIML}	31.262	78.324	11.113	60.166	.111	.264	.369
		T_{M}	36.509	112.548	13.636	73.336	.249	.443	.590
		T_{BM}	26.774	48.786	11.082	51.608	.025	.086	.168
		D_{2}					.085	.231	.347
		D_{3}					.118	.261	.375
300	1000	T_{FIML}	42.267	130.214	14.534	85.742	.439	.688	.794
		T_{M}	50.682	193.870	15.669	106.162	.690	.857	.918
		T_{BM}	37.523	91.982	13.273	79.797	.250	.501	.657
		D_{2}					.360	.596	.718
		D_{3}					.469	.700	.810
500	1000	T_{FIML}	53.551	164.487	18.168	134.664	.789	.923	.951
		T_{M}	65.483	253.279	21.067	166.018	.938	.978	.986
		T_{BM}	47.709	111.301	17.901	108.624	.658	.860	.929
		D_{2}					.711	.868	.928
		D_{3}					.831	.939	.965
MAR									
100	1000	T_{FIML}	31.456	80.279	10.786	64.347	.099	.253	.384
		T_{M}	36.751	118.558	10.992	74.062	.253	.492	.604
		T_{BM}	27.115	53.851	8.508	60.846	.028	.115	.196
		D_{2}					.086	.227	.337
		D_{3}					.132	.288	.397
300	1000	T_{FIML}	42.298	142.001	17.718	81.384	.445	.667	.766
		T_{M}	51.275	208.942	20.686	94.843	.696	.850	.911
		T_{BM}	37.856	90.996	16.693	72.636	.282	.531	.672
		D_{2}					.378	.618	.711
		D_{3}					.497	.706	.794
500	1000	T_{FIML}	52.935	174.141	23.530	92.953	.761	.912	.955
		T_{M}	66.107	285.771	24.479	128.708	.936	.986	.993
		T_{BM}	47.633	112.088	18.441	81.747	.640	.854	.928
		D_{2}					.683	.865	.935
		D_{3}					.810	.936	.972

Note: The entries in the "Converged" column refer to the number of converged replications in each condition.

TABLE 3
Bias and Variability of Parameter Estimates Under MCAR (N = 100)

	True Value	FIML		MI2S		MI	
		RB (%)	RMSD	RB (%)	RMSD	RB (%)	RMSD
λ_{21}	0.7	3.35	0.204	3.41	0.208	3.49	0.206
λ_{31}	0.9	3.40	0.238	3.68	0.248	3.62	0.240
λ_{52}	0.9	0.30	0.107	0.27	0.108	0.28	0.107
λ_{62}	0.9	0.61	0.107	0.60	0.108	0.60	0.108
λ_{83}	1.0	3.26	0.258	4.45	0.301	4.65	0.274
λ_{93}	1.0	2.63	0.249	4.06	0.298	4.35	0.332
ψ_{11}	0.6	0.85	0.197	0.68	0.198	0.79	0.197
ψ_{21}	0.4	−1.11	0.111	−1.45	0.112	−1.46	0.112
ψ_{31}	0.3	−3.07	0.106	−2.32	0.113	−2.47	0.107
ψ_{22}	0.9	−0.68	0.173	−0.65	0.174	−0.66	0.173
ψ_{32}	0.2	0.13	0.097	0.66	0.105	0.39	0.098
ψ_{33}	0.5	2.62	0.179	5.32	0.208	5.64	0.188

missing information adjusted standard error estimates, as well as additional model fit indices such as RMSEA (with confidence interval) and TLI. The third application shows that the MI2S estimator can conveniently handle mean and covariance structure modeling for data sets containing multiple sets of plausible values. When we have to create multiple imputations using MCMC sampling, we generally use a burn-in period of 1,000 iterations and subsampling intervals of 200.

CFA

We first apply the MI2S procedure to the well-known Open-Book Closed-Book data set in Mardia, Kent, and Bibby (1979). The original data set contains test scores on five subject areas obtained from 88 examinees, that is, $n = 5$ and $N = 88$. The tests of the first two subjects are open-book, and the rest are closed-book exams.

The model specification is as follows. The factor loading matrix is

$$\Lambda' = \begin{pmatrix} \lambda_{11} & \lambda_{21} & 0 & 0 & 0 \\ 0 & 0 & \lambda_{32} & \lambda_{42} & \lambda_{52} \end{pmatrix},$$

reflecting the testing mode (open-book vs. closed-book), and the factors are correlated

$$\Psi = \begin{pmatrix} 1.0 & \\ \psi_{21} & 1.0 \end{pmatrix}.$$

The unique factor covariance matrix is given by $\Phi = \mathrm{diag}(\phi_{11}, \ldots, \phi_{55})$. The degrees of freedom is equal to 4. For this complete data set, it is known that a

two-factor CFA model fits very well, yielding a chi-square of 2.07 under maximum likelihood estimation.

To illustrate the newly proposed MI2S estimator, we artificially created MAR data by applying the same procedure used in Cai and Lee (2009), wherein the scores on the last three variables were set to missing if the sum of the first two was less than 80, resulting in 28 cases with missing values. Next, based on 20 imputations ($M = 20$), the combined estimate of means, variances, and covariances, $\tilde{\omega}$, was obtained by Equation 16. The two-factor CFA model was then fitted using the MI2S estimator.

The empirical results mirror the simulation study. More specifically, with 4 degrees of freedom, the MI2S test statistic T_{BM} is $6.67, p = .15$, whereas the naive statistic T_{M} is $10.89, p = .03$. Having analyzed the complete data set, we know that the model should fit well. In this case, the difference between T_{BM} and T_{M} is large enough to lead to a possibly erroneous decision on the fit of the hypothesized CFA model. RMSEA and TLI for this model are, based on our MI2S estimator, equal to .086 and .808, respectively. A 90% confidence interval of RMSEA is [.000, .199].

For the purpose of comparison, we fitted the model using the FIML estimator and obtained $T_{\text{FIML}} = 6.64, p = .16$, a result in close agreement with the MI2S estimator. We also computed D_2 and D_3 statistics and the associated p values for this model. The p value associated with D_2 is $F(4, 91.35) = 1.75, p = .15$ and the p value associated with D_3 is $F(4, 302.82) = 1.42, p = .23$.

We also calculated adjusted standard error estimates for the structural parameter estimates. The results are presented in Table 4. The entries in the "Adjusted" column show the standard error estimates produced by MI2S estimator and the entries in the "Unadjusted" column show the naive standard error estimates obtained by treating the combined estimate of the means and the covariance matrix as if it comes from complete data. Notice that the unadjusted standard error estimates are in general smaller than the MI2S standard error estimates. It is an artifact caused by neglecting the fraction of missing information in the combined estimates. In contrast, the MI2S standard errors are properly inflated, accounting for the missing information. For the purpose of comparisons, the standard error estimates from the standard MI estimator and FIML estimator are presented in the columns of "Standard" and "FIML," respectively. It can be seen that these standard error estimates are not only properly inflated adjusting for the missing information but also very close to those from the MI2S estimator.

Conditional Latent Curve Analysis

In the second application, the data from a symposium at the 1997 meeting of the Society for Research on Child Development (Curran, 1997) were analyzed. The data set contains four repeated measures of $N = 405$ participants from the National Longitudinal Survey of Labor Marketing Experience in Youth on their aggressive

TABLE 4

Standard Errors of Structural Parameter Estimates for the Open-Book Closed-Book Data

		SE			
	Estimate	Adjusted	Unadjusted	Standard	FIML
λ_{11}	11.84	1.88	1.76	1.93	1.88
λ_{21}	10.63	1.43	1.29	1.49	1.42
λ_{32}	10.81	1.21	0.90	1.23	1.32
λ_{42}	14.36	1.94	1.48	2.04	2.16
λ_{52}	14.82	2.48	1.70	2.59	2.41
ψ_{21}	0.91	0.08	0.06	0.08	0.09
ϕ_{11}	162.20	31.70	29.07	34.54	33.13
ϕ_{22}	57.95	19.76	15.27	22.63	20.12
ϕ_{33}	8.51	6.58	4.87	6.34	6.26
ϕ_{44}	77.93	16.36	14.74	16.61	17.33
ϕ_{55}	131.38	28.79	22.39	27.78	24.86

Note: The entries in the "Adjusted" column show missing information adjusted standard error estimates produced by MI2S estimator, whereas the entries in the "Unadjusted" column show the incorrect standard error estimates when the combined estimate $\tilde{\omega}$ (see Equation 16) is treated as if it comes from complete data. The entries in the "Standard" and "FIML" columns show the standard error estimates produced by the standard MI estimator and the FIML estimator, respectively.

behavior. A number of time-invariant covariates, including gender and mother's age, are also available. The data set exhibits substantial attrition of respondents. Roughly half of the cases have missing observations in the aggressive behavior variable on one or more measurement occasions beyond baseline. To handle the missingness, we created $M = 20$ imputations and utilized the MI2S estimator.

The goal is to fit a linear latent curve model (Bollen & Curran, 2006) to the repeated measurements of aggressive behavior. We included gender and mother's age as time-invariant covariates, making $n = 6$. We freely estimated the covariance matrix of the intercept and slope factors. The time-specific residual variances were allowed to be heteroskedastic. This model has 9 degrees of freedom.

The MI2S test statistic and the FIML test statistic come out to be very similar, that is, $T_{BM} = 16.23, p = .06$ and $T_{FIML} = 16.56, p = .06$, respectively. In contrast, the naive statistic T_M is $21.07, p = .01$. Once again, the difference between T_M and T_{BM} is large enough to lead to qualitatively different conclusions about model fit. The p value associated with D_2 is $F(9, 199.21) = 1.68, p = .10$ and the p value associated with D_3 is $F(9, 1926.62) = 1.58, p = .12$. The RMSEA and TLI based on MI2S estimator are equal to .044 and .893, respectively, with a 90% confidence interval of RMSEA being [.000, .078]. Though not reported here due to space constraints, unadjusted standard error estimates for the structural parameter estimates are all erroneously smaller than the adjusted standard error estimates produced by the MI2S estimator.

Analysis Involving Plausible Values: PISA

As the third application, the MI2S estimator is applied in the mean and covariance structure modeling of a data set containing multiple sets of plausible values: PISA. Conducted triennially by the Organization for Economic Co-operation and Development (OECD), PISA is a system of international assessments that measures 15-year-olds' literacy in reading, mathematics, and science. The PISA data also include numerous items on student characteristics, student family background, and student perceptions, just to name a few. More information about PISA can be found at http://www.pisa.oecd.org/.

A distinctive characteristic of the PISA data set is that student performance on literacy scales and subscales are reported as $M = 5$ sets of plausible values for each of the scales or subscales. In other words, instead of reporting a single scale score, PISA uses multiple imputation to represent the unknown values of students' literacy. As such, the MI2S estimator can be effectively employed. Indeed, it can be even simpler than the case of missing data because we do not have to generate the imputations ourselves.

In this particular application, we use data from the assessment conducted in 2006, focusing on the United States school sample. PISA 2006 uses a two-stage sampling design in which schools are first sampled and then students are sampled in the participating schools. Therefore, it is required to use sampling weights or to model the clustered data structure explicitly (or maybe both) for sound statistical analyses.[1] However, for a simple illustration of the MI2S estimator under the presence of multiple sets of plausible values, we ignored the sampling weights.

First, purely for the sake of illustration, we developed a hypothetical structural equation model in which a student's science literacy (SCI) is regressed on his or her mathematics literacy (MATH) and the value of science (VoSCI) that he or she holds. Figure 1 presents a conceptual path diagram of our structural equation model. MATH is an observed variable consisting of five sets of plausible values. SCI is a latent variable, measured by three subscale variables: explaining phenomena scientifically (EPS), using scientific evidence (USE), and identifying scientific issues (ISI). EPS is a scaling indicator with a fixed factor loading of 1.0. Each of the science subscale variables contains five sets of plausible values. VoSCI is a latent variable, measured by two manifest variables: the general value of science (GSCI) and perceptions of the personal value of science (PSCI). GSCI is the scaling indicator for VoSCI. Neither GSCI nor PSCI contains plausible values. We freely estimated the covariance of a student's mathematics literacy (MATH) and his/her value of science (VosCI). To aid interpretation, we standardized all observed variables. The analysis sample includes cases with complete data for all variables ($N = 5531$).

Next, we combined the $M = 5$ data sets containing plausible values and produced a single estimate of means and covariance matrix using Equation 16. We then applied the MI2S estimator to assess the plausibleness of the postulated
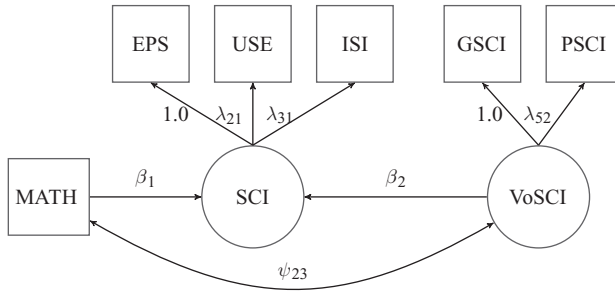
FIGURE 1. *A conceptual path diagram for the hypothetical structural equation model for PISA data. Only the key structural parameters are shown.*

TABLE 5.
Structural Parameter Estimates and Standard Errors for the PISA Data

	Estimate	*SE*
λ_{21}	1.01	.005
λ_{31}	0.99	.006
λ_{52}	0.77	.031
β_1	0.07	.009
β_2	0.81	.008
ψ_{23}	0.29	.014

Note: The standard errors are missing information adjusted standard error estimates produced by MI2S estimator.

model. The overall model fit statistic, T_{BM}, came out to be 114.87 on 7 degrees of freedom, $p < .001$. We can also compute RMSEA and TLI for this model, which are equal to .053 and .947, respectively, with a 90% confidence interval of RMSEA being [.045, .062]. Furthermore, we obtained the structural parameter estimates and the associated standard error estimates (see Table 2). We also computed D_2 and D_3 statistics and the associated p values. The p value associated with D_2 is $F(7, 104.95) = 15.80, p < .001$ and the p value associated with D_3 is $F(7, 553.98) = 14.70, p < .001$. Taken together, the results indicate that the MI2S estimator is equally applicable to the case of plausible values.

Discussions

In this article, we proposed a new two-stage estimator and inferential tools (MI2S) for mean and covariance structure modeling under the presence of multiple imputations. While standard multiple imputation theory dictates that one should fit a mean and covariance structure model M times (once to each

imputation) and combine the results at the very end, the MI2S estimator permits us to combine the multiple imputations at the beginning of data analysis. It produces a single estimate of the means and covariance matrix by averaging the unstructured means and covariance matrices across multiple imputations. Adjustment for the between-imputation variation is also carried out at the beginning, yielding a proper asymptotic covariance matrix of the estimated means and covariances. Using Browne's (1984) results, this asymptotic covariance matrix can be used to obtain an asymptotically chi-square distributed model fit statistic, T_{BM}, and the standard error estimates associated with the structural parameter estimates.

Our simulation studies and empirical data analysis show that the MI2S estimator performs as well as the existing FIML estimator under MCAR and MAR conditions not only in terms of Type I error rates and statistical power of overall model fit statistics but also in terms of bias and variability of parameter estimates.

The MI2S estimator is potentially more flexible than FIML. For example, handling plausible values is natural in the MI2S estimator, whereas FIML estimator will not be of much help because there is technically no missingness in the data sets. In addition, inclusion of missing-data-relevant covariates is very straightforward because MI2S is within a multiple imputation framework. In contrast, though it is possible to include missing data relevant covariates for the FIML estimator (Collins, Schafer, & Kam, 2001; Graham, 2003), significantly more efforts are required if the missing data mechanism needs to be explicitly modeled. Moreover, as the number of missing-data-relevant covariates increases, model estimation and identification becomes increasingly problematic for FIML (Savalei & Bentler, 2009, p. 494).

Compared with the standard multiple imputation procedures, the MI2S estimator is computationally more efficient (i.e., a single model-fitting vs. multiple model-fittings) and produces a chi-square distributed test statistic T_{BM} that can form the basis of mean and covariance structure model fit evaluations. Using T_{BM}, it becomes straightforward to conduct tests of close fit (Browne & Cudeck, 1993) and obtain various fit indices such as RMSEA and TLI. On the other hand, using the standard multiple imputation procedure, it is not clear how to combine multiple sets of various fit indexes obtained from the multiple model-fittings because imputation generates a slightly different saturated model.

Our research is not without limitations. First, the derivations are exclusively based on multivariate normal theory. The multivariate normality assumption implies that in higher order moments, interactions or nonlinearities among variables are not modeled in the imputation process. Therefore, if such complex associations are to be a crucial part of the analysis, biased consequences may be obtained due to the inconsistencies between the model used in the imputations and the model used in subsequent analyses of the imputed data sets. This remains an inherent issue for multiple imputation based estimators, MI2S included.

When the variables are clearly not normal (e.g., categorical variables or design variables), a different imputation approach can be employed based on log-linear

models or general location models (Schafer, 1997) or sequential generalized regression models (Raghunathan, Lepkowski, Van Hoewyk, & Solenberger, 2001). More research should be conducted in the future to explore the possible existence of non-normal versions of two-stage estimators that work under multiple imputation.

Second, our simulation studies only examined a small set of conditions. In particular, a closer examination of Table 1 reveals that, as sample size increases, Type I error rates of T_{FIML} tend to approach the nominal level from above, whereas the opposite trend holds for T_{BM}. Such tendencies may explain why T_{BM} can be a less powerful statistic than T_{FIML}, particularly when N is small. Before reaching a general conclusion, however, further research is warranted. Future simulation studies should also examine the robustness of MI2S under a wider variety of missing data conditions and types of structural models.

Third, we only implemented the proposed methods in SAS because of its widespread use in both academic and nonacademic settings. Though our macro is freely available, it is nevertheless restricted to a single software environment. We will explore the possibility of implementing the method in free statistical software such as R.

Appendix A

Usage of the SAS Macro *MI2S*

We use the conditional latent curve model as an example to describe the usage of the SAS macro. Suppose a SAS data set called *srcd* is present in the *work* library, and it contains the following six variables: *anti1–anti4*, *gen*, and *momage*. The first four *anti* variables are the repeated measures of aggressive behavior, and *gen* and *momage* are the gender and mother's age variables at baseline.

First, the SAS macro definitions should be included using the % *include* statement.

Next, multiple imputation of missing data should be performed. For instance, we use the following PROC MI statements:

```
proc mi data=srcd seed=10 nimpute=20 out=outmi20 noprint;
  mcmc   chain=single   nbiter=1000   niter=200   initial=em
prior=jeffreys;
  var anti1 anti2 anti3 anti4 gen momage;
run;
```

PROC MI will create a data set with 20 imputations. The imputed data sets are stored in the *outmi20*, with the SAS default index variable *_Imputation_* indicating the imputation numbers. This data set should be fed into the SAS macro as input data. Note that in the case of preexisting imputations (e.g., plausible values), the imputation step can be skipped. However, in that case, one must know the name of the index variable for the imputation numbers and sort the data by that index variable using PROC SORT.

Next, one must define the structural model to be fitted to the MI2S estimated means and covariances using one of the programming statements acceptable to PROC CALIS. The model definition *must* be a quoted SAS macro string and in this example we use the LINEQS-style statements:

```
%let mymodel=%str(
  lineqs
    anti1 = 0. * intercept + 1f1 + 0f2 + e1,
    anti2 = 0. * intercept + 1f1 + 1f2 + e2,
    anti3 = 0. * intercept + 1f1 + 2f2 + e3,
    anti4 = 0. * intercept + 1f1 + 3f2 + e4,
    f1 = al1 intercept + gamma1 gen + gamma2 momage + d1,
    f2 = al2 intercept + gamma3 gen + gamma4 momage + d2,
    gen = al3 intercept + d3,
    momage = al4 intercept + d4;
  std
    e1-e4 = th1 th2 th3 th4,
    d1-d4 = ph11 ph22 ph33 ph44;
  cov
    d1 d2 = ph12,
    d3 d4 = ph34;
);
```

Finally, the macro *MI2S* is invoked.

```
%MI2S(indata=outmi20, var=anti1 anti2 anti3 anti4 gen momage,
  nvar=6, nobs=405, nimp=20, impidx=_imputation_, calismodel=
&mymodel);
```

The *indata* and *var* options tell the macro the multiply imputed data set name and the names of the variables. The *nvar* and *nobs* options give the dimensions of the analysis in terms of the number of manifest variables and the number of observations. The *nimp* and *impidx* options tell the macro the number of imputations performed and the name of imputation index variable. These options should be specified in accordance with the number of imputations and index variable name specified previously. In this particular example, *nimp=20* and *impidx=_Imputation_*. Consider another example. When a data set consists of five sets of plausible values indexed by *_PV_*, then *nimp = 5* and *impidx = _PV_*.

Finally, the quoted macro string &mymodel is passed on to the macro to define the structural model. The macro produces the following output:

Goodness-of-fit Test for MI2S Estimator

T_BM	df	p value
16.2345	9.0000	.0621

Point Estimate of RMSEA
0.0445

The value of T_{MB}, the degrees of freedom, the p value, and RMSEA point estimate are printed. The macro also prints MI2S structural parameter estimates along with adjusted standard errors.

Declaration of Conflicting Interests

Funding

Note

1. For more information about the sampling design of PISA 2006 and its proper statistical handing, see OECD's *PISA 2006 Technical Report* (OECD, 2009b) and *PISA 2006 Data Analysis Manual* (OECD, 2009a).

References

Allison, P. D. (2001). *Missing data*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07–136. Thousand Oaks, CA: Sage.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, *112*, 545–557.

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

Aptech Systems, Inc. (2003). GAUSS (Version 6.08) [Computer software]. Maple Valley, WA: Author.

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. 697–715). Mahwah, NJ: Lawrence Erlbaum.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. New York, NY: John Wiley.

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

Browne, M. W., & Arminger, G. (1995). Specification and estimation of mean and covariance structure models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 185–249). New York, NY: Plenum.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Cai, L. (2008). SEM of another flavour: Two new applications of the supplemented EM algorithm. *British Journal of Mathematical and Statistical Psychology*, *61*, 309–329.

Cai, L., & Lee, T. (2009). Covariance structure model fit testing under missing data: An application of the supplemented EM algorithm. *Multivariate Behavioral Research*, *44*, 281–304.

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing-data procedures. *Psychological Methods*, *6*, 330–351.

Curran, P. J. (1997). *Comparing three modern approaches to longitudinal data analysis: An examination of a single developmental sample*. Symposium presented at the 1997 Biennial Meeting of the Society for Research in Child Development, Washington, DC.

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

Graham, J. W. (2003). Adding missing-data relevant variables to FIML-based structural equation models. *Structural Equation Modeling*, *10*, 80–100.

Jöreskog, K., & Sörbom, D. (2001). *LISREL user's guide*. Chicago, IL: SSI International.

Li, K.-H., Meng, X.-L., Raghunathan, T. E., & Rubin, D. B. (1991). Significance levels from repeated p-values with multiply-imputed data. *Statistica Sinica*, *1*, 65–92.

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

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. London, England: Academic Press.

Meng, X.-L., & Rubin, D. B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika*, *79*, 103–111.

Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, *29*, 133–161.

Organization for Economic Cooperation and Development. (2009a). *PISA 2006 data analysis manual*. Paris, France: Author.

Organization for Economic Cooperation and Development. (2009b). *PISA 2006 technical report*. Paris, France: Author.

Raghunathan, T., Lepkowski, J., Van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, *27*, 85–95.

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

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, *91*, 473–489.

Rubin, D. B., & Thayer, D. T. (1982). EM algorithms for ML factor analysis. *Psychometrika*, *47*, 69–76.

Savalei, V., & Bentler, P. M. (2009). A two-stage approach to missing data: Theory and application to auxiliary variables. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 477–497.

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

Schott, J. R. (1997). *Matrix analysis for statistics*. New York, NY: John Wiley.

Stuart, E. A., & Rubin, D. B. (2008). Matching with multiple control groups with adjustment for group differences. *Journal of Educational and Behavioral Statistics*, *33*, 279–306.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1–10.

Yuan, K.-H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, *30*, 165–200.

Yuan, K.-H., & Bentler, P. M. (2007). Structural equation modeling. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Psychometrics* (Vol. *26*, pp. 297–358). Amsterdam, Netherlands: North-Holland.

Yuan, K.-H., & Jennrich, R. I. (1998). Asymptotics of estimating equations under natural conditions. *Journal of the Multivariate Analysis*, *65*, 245–260.

Authors

TAEHUN LEE is an assistant professor in the Department of Psychology at the University of Oklahoma, Norman, OK 73019; e-mail: Taehun.Lee-1@ou.edu. His research interests concern methodological/statistical issues surrounding estimation, evaluation, and selection of latent variable models, which include factor analysis models, structural equation models, multilevel models and item response theory models. He is also interested in applications of statistical models for the study of individual differences and change over time.

LI CAI is an associate professor of education and psychology at UCLA, Box 951521, 315 GSEIS Building, Los Angeles, CA 90095; e-mail: lcai@ucla.edu. There he also serves as codirector of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). His methodological research agenda involves the development, integration, and evaluation of innovative latent variable models that have wide-ranging applications in educational, psychological, and health-related domains of study. A key component on this agenda is statistical computing, particularly as related to multidimensional item response theory (IRT) and multilevel modeling. He has also collaborated with substantive researchers at UCLA and elsewhere on projects examining measurement issues in mental health, substance abuse treatment, and patient-reported outcomes.