Published in final edited form as:
Sociol Methods Res. 2012 November ; 41(4): 598–629. doi:10.1177/0049124112460373.

ML versus MI for Missing Data with Violation of Distribution Conditions*

Ke-Hai Yuan,
University of Notre Dame

Fan Yang-Wallentin, and
Uppsala University, Sweden

Peter M. Bentler
University of California, Los Angeles

Abstract

Normal-distribution-based maximum likelihood (ML) and multiple imputation (MI) are the two major procedures for missing data analysis. This article compares the two procedures with respects to bias and efficiency of parameter estimates. It also compares formula-based standard errors (SEs) for each procedure against the corresponding empirical SEs. The results indicate that parameter estimates by MI tend to be less efficient than those by ML; and the estimates of variance-covariance parameters by MI are also more biased. In particular, when the population for the observed variables possesses heavy tails, estimates of variance-covariance parameters by MI may contain severe bias even at relative large sample sizes. Although performing a lot better, ML parameter estimates may also contain substantial bias at smaller sample sizes. The results also indicate that, when the underlying population is close to normally distributed, SEs based on the sandwich-type covariance matrix and those based on the observed information matrix are very comparable to empirical SEs with either ML or MI. When the underlying distribution has heavier tails, SEs based on the sandwich-type covariance matrix for ML estimates are more reliable than those based on the observed information matrix. Both empirical results and analysis show that neither SEs based on the observed information matrix nor those based on the sandwich-type covariance matrix can provide consistent SEs in MI. Thus, ML is preferable to MI in practice, although parameter estimates by MI might still be consistent.

Keywords

bias; standard error; consistency; observed information; sandwich-type variance; Monte Carlo

*The research was supported by NSF grant DMS04-37167, a grant from the National Natural Science Foundation of China (30870784), and Grants DA00017 and DA01070 from the National Institute on Drug Abuse.

Correspondence concerning this article should be addressed to Ke-Hai Yuan (kyuan@nd.edu).

Introduction

Incomplete or missing data exist in almost all areas of empirical research. They are especially common in longitudinal studies in social and behavioral sciences. Many statistical procedures have been developed for analyzing missing data. Two notable ones are maximum likelihood (ML) and multiple imputation (MI). Under the assumption of a correctly specified parametric model and that data are missing at random, both procedures generate consistent parameter estimates and consistent standard errors (e.g., Little & Rubin, 2002; Schafer, 1997). Recent developments indicate that the normal-distribution-based ML can still generate consistent parameter estimates and consistent standard errors (SE) even when the population distribution is unknown (Yuan, 2009). Although no analytical results exist for MI to generate consistent parameter estimates when the parametric model is misspecified, it has been stated in the literature that the normal-distribution-based MI generates reasonable parameter estimates and SEs with distribution violations (e.g., Schafer, 1997, p. 136; Schafer & Graham, 2002; Schafer & Olsen, 1998). The purpose of this paper is to compare the robustness of the two major missing data methods. Using Monte Carlo simulation, we will study the biases in parameter estimates by ML and MI. We will also compare the formula-based SEs by ML and MI against their respective empirical SEs. Information on the relative efficiency of the two classes of estimators will also be obtained by comparing their empirical SEs.

Missing data can occur for various reasons. The process by which data become incomplete was called the missing data mechanism by Rubin (1976). Missing completely at random (MCAR) is a process in whichmissingness of data is independent of both the observed and the missing values; missing at random (MAR) is a process in which missingness is independent of the missing values given the observed data. When missingness depends on the missing values themselves given the observed data, the process is missing not at random (MNAR). When all missing values are MCAR, most ad hoc procedures still generate consistent parameter estimates. When ignoring the process for data that are MAR, including MCAR, ML and MI can generate consistent and efficient parameter estimates under a correctly specified parametric model. Thus, missing data with MCAR and MAR mechanisms are sometimes referred to as ignorable non-responses. When missing values are MNAR, one has to correctly model the missing data mechanism in order to get consistent parameter estimates in general. In this paper, we mainly study the normal-distribution-based ML and MI when missing values are MAR.

ML with missing data has a long history. After Rubin (1976) justified ML with MAR data, ML procedures for missing data have been developed in almost every aspect of statistics (Little & Rubin, 2002; Molenberghs & Kenward, 2007; Schafer, 1997). MI was proposed by Rubin (1987), but its wide use is mainly due to various free and commercial programs (see e.g., Harel & Zhou, 2007; Horton & Kleinman, 2007). Nice nontechnical introductions to MI were given by Allison (2001) and Schafer and Olsen (1998). Nowdays, ML and MI are the recommended procedures in essentially all areas of data analysis with missing values (e.g., Allison, 2000, 2003; Buhi, Goodson & Neilands, 2008; Choi, Golder, Gilimore & Morrison, 2005; Croy & Novins, 2005; Jamshidian & Bentler, 1999; Kenward & Carpenter,

2007; King et al., 2001; Lee & Song, 2007; Olinsky, Chen & Harlow, 2003; Peng & Zhu, 2008; Peugh & Enders, 2004; Taylor & Zhou, 2009; Thomas, 2000).

Most developments for ML and MI with the MAR mechanism are based on correctly specified distributions. With complete data, we can use existing procedures to check the distributional properties of the sample before choosing a parametric model (e.g., D'Agostino, Belanger & D'Agostino, 1990). With missing data, especially when missing values are MAR, the observed data can be skewed and possess excess kurtosis even when the underlying population is normally distributed. Then most procedures for testing univariate or multivariate normality are not applicable (see e.g., Yuan, Lambert & Fouladi, 2004). Thus, we have to rely on the robust properties of ML or MI in data analysis with missing values. In the context of structural equation modeling (SEM) with distribution violations, Arminger and Sobel (1990) proposed to use a sandwich-type covariance matrix to estimate the SEs of the normal-distribution- based maximum likelihood estimates (MLE). Yuan (2009) and Yuan and Bentler (2010) recently showed that, even when the underlying population distribution is unknown, the normal-distribution-based MLEs are still consistent under the MAR mechanism, and the covariance matrix of the MLEs is consistently estimated by the sandwich-type covariance matrix proposed in Arminger and Sobel (1990). However, the performance of the SEs based on the sandwich-type covariance matrix has never been evaluated empirically with missing data. Enders (2001) evaluated biases in MLEs in the context of SEM when missing values are MAR, it is not clear why in his Table 3 the bias decreases as the proportion of missing values increases for a population with heavy tails. For the robustness of MI, Graham and Schafer (1999) performed a simulation study by treating a real data set as the population. They found that the absolute values of the biases are small while most of their population values of the regression parameters are also small. Actually, several biases of their estimates are greater than the population values of the regression parameters. It is not clear whether the small biases are due to the small values of the population parameters. The simulation reported in section 6.4 of Schafer (1997) is also based on a real data set. The study does not include a systematic evaluation of the effect of population skewness and kurtosis on parameter estimates by MI. Demirtas, Freels and Yucel (2008) recently conducted a more comprehensive simulation study on MI with two variables, one is complete and one contains missing values. They found that estimates of variance parameters by MI can suffer from serious bias when the proportion of missing data is large and the sample size is small, especially when the population is nonnormally distributed. None of the above literature compared MI against ML, and none systematically studied the performance of formula-based SEs of ML and MI either.

Because data sets in social sciences are seldom normally distributed (Micceri, 1989), it is important to know how ML and MI behave relative to each other under the condition of distribution violations. Actually, the results of Yuan (2009) and Yuan and Bentler (2010) for ML are all based on asymptotics. It is not clear whether MLEs are more biased than parameter estimates by MI at finite sample sizes. It is also not clear how the SEs based on the sandwich-type covariance matrix perform in practice. With real data, Schafer (1997, section 6.4) reported some results on the formula-based SEs of MI, where the normal-distribution-based information matrix is used to compute the covariance matrix of the parameter estimates with the (imputed) complete data. It is very likely that MI together with

a sandwich-type covariance matrix for complete data is more robust to distribution violations. This has been suggested by Schafer and Graham (2002, p. 170) in the context of SEM. However, it is not clear whether such a combination will generate consistent SEs or whether it will generate more accurate SEs than those based on the observed information matrix. There is also a need to compare this robust version of MI against the robust version of ML. Since both MI and ML are available in various statistical programs, with typical samples in social sciences coming from populations whose distributions are unknown, answers to the above questions will give the needed information for applied researchers to choose a more appropriate missing data procedure.

We will use Monte Carlo simulation to address the above questions. We will focus on estimates of means and variances-covariances by ML and MI. This is because means and variances-covariances serve as building blocks for almost all commonly used methods in social and behavioral sciences (e.g., ANOVA, regression, correlations, factor analysis, principal component analysis, SEM, growth curves, etc.). If a missing data method leads to estimates of means and variances-covariances with little bias, then it will result in little bias for parameter estimates that are continuous functions of means and variances-covariances. If substantial bias exists in the estimates of means and variances-covariances, then we have to be lucky enough to get a good estimate of a function of means and/or variances-covariances. Since essentially all the commonly used parameter estimates are continuous functions of means and covariance matrices, the obtained results will have wide practical implications.

We will study possible bias in the estimates of means and variances-covariances by normal-distribution-based ML and MI. We will also compare empirical SEs and formula-based SEs provided by ML and MI. We review the methods and design of the study in the next section. The following section presents the Monte Carlo results. We conclude the paper with some discussion and advice on proper use of the two methods.

Methods

To study the effect of sample size, missing data proportion and departure from normality on the normal-distribution-based ML and MI, for simplicity and also for a thorough study with a reasonable length, we will mainly consider the problem with two variables. Actually, two variables already allow us to tell the pros and cons of the two methods. Following the suggestion of a reviewer on an earlier version of the paper, we also include a model with five variables.

Study design with two variables

Let z_1 and z_2 be two independent and standardized random variables, and

$$y_1=\mu_1+\sigma_1 z_1, y_2=\mu_2+\sigma_2[\rho z_1+(1-\rho^2)^{1/2} z_2]. \quad (1)$$

Then $\mathbf{y} = (y_1, y_2)'$ follows a bivariate distribution with a mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ and a variance-covariance matrix

$$\mathbf{\Sigma}=\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{pmatrix},$$

where $\sigma_{11}=\sigma_1^2$, $\sigma_{12}=\sigma_1\sigma_2\rho$ and $\sigma_{22}=\sigma_2^2$. Consider the sample

$$\begin{matrix} y_{11},\ldots,y_{n1},y_{(n+1)1}\cdots,y_{N1}\\ y_{12},\ldots,y_{n2}\end{matrix} \quad (2)$$

from the population **y**, where the first variable is observed on all *N* cases while the second one is observed only on the first *n* cases. Suppose the missingness of y_{i2} is due to the value of y_{i1} being greater than certain values. Because the value of y_{i1} is observed, all the missing values are MAR. Actually, using model (1) to simulate missing values with the MAR mechanism was suggested by Little and Rubin (2002, p. 90). The same model has been used to generate bivariate complete data with desired population correlations (e.g., Lee & Rodgers, 1998).

We choose $\mu_1 = 1$, $\mu_2 = 2$, $\sigma_1 = 1$, $\sigma_2 = 1$, and $\sigma_{12} = \rho = .5$ in the population. Three distribution conditions on z_1 and z_2 are used: The standard normal distribution $N(0, 1)$, the standardized log-normal distribution[1] $LN_s(0, 1/2)$, and the standardized uniform distribution $U_s(0, 1)$. The combination of z_1 and z_2 results in nine distribution conditions for y_2. The population skewness and kurtosis of y_2 for the nine conditions are given in Table 1, where the skewness ranges from 0 to 2.276 and the kurtosis ranges from −.750 to 11.567. These are well-within the range of the skewness and kurtosis of a real data set as reported in Table 2 of Graham and Schafer (1999). The population skewness and kurtosis of y_1 are (0, 0) when $z_1 \sim N(0, 1)$, (2.939, 18.507) when $z_1 \sim LN_s(0, 1/2)$, and (0,−1.200) when $z_1 \sim U_s(0, 1)$.

Although we used the same number of variables as in Demirtas et al. (2008), the designs are different. In their study, except for the normally distributed population, they let the two marginals be uncorrelated/independent. It is not easy to generate missing values with an MAR mechanism[2] when the two variables are uncorrelated. Demirtas et al. also let the two marginals be identically distributed. Some of the substantial bias reported in their paper can be due to either/both the nonnormality of the variable having missing values or/and the nonnormality of the variable containing no missing values. The design in this paper, with different marginal distributions, allows us to locate the causes of possible problems.

For each distribution condition, we choose five sample sizes $N = 30, 50, 100, 200, 500$, which are intended to cover sample sizes from small to large[3]. For each combination of distribution and sample size, three missing data conditions are created by deleting the corresponding y_{i2} when y_{i1} is greater than its 90th, 70th and 50th population quantiles,

[1]A random variable *x* following the log-normal distribution $LN(\mu, \sigma^2)$ is obtained by $x = \exp(z)$ and $z \sim N(\mu, \sigma^2)$. For a given random variable *x*, its standardized version is obtained by $x_S = [x - E(x)]/\{\mathrm{Var}(x)\}^{1/2}$.

[2]When two variables are independent and the missingness of the second variable depends on the value of the first variable, all the observed values for each variable form a random sample from the corresponding marginal population. Thus, the MAR mechanism automatically becomes MCAR.

[3]Small to large sample sizes depend on the problem considered. While $N = 500$ may be considered as a small sample size when there are 50 variables, it is large enough for most practical purposes when only 2 variables are involved.

respectively. Since y_{i1} is observed on all the cases, the proportions of missing values for the whole sample are p_m = .05, .15 and .25, respectively. Because proportions of complete cases can range from 100% (no missing value) to less than 50% in practice (Daniels & Hogan, 2008), we will regard p_m = .05 as a small or trivial proportion[4] and p_m= .25 as a large proportion. In summary, a total of 9 × 5 × 3 = 135 conditions are studied.

With observations on y_1 being complete, the estimate of $\boldsymbol{\gamma} = (\mu_1, \sigma_{11})'$ by either ML or MI is just $\hat{\boldsymbol{\gamma}} = (\bar{y}_1, s_{11})'$, where $\bar{y}_1$ is the sample mean and s_{11} is the sample variance of y_1. Since $\bar{y}_1$ is known to be unbiased and the bias in s_{11} (= $-\sigma_{11}/N$) is well-known, we will not further study $\hat{\boldsymbol{\gamma}}$. The MLE $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta} = (\mu_2, \sigma_{12}, \sigma_{22})'$ can be obtained by the analytical formula of Anderson (1957). With 500 replications[5] for each combination of the three conditions (population distribution, missing data proportion, and sample size) the average $\bar{\theta}$ and the sample standard deviation of $\hat{\theta}$ are obtained, where $\hat{\theta}$ is an element of $\hat{\boldsymbol{\theta}}$. The empirical bias of $\hat{\theta}$ is subsequently obtained using

$$\text{Bias} = \bar{\theta} - \theta.$$

The sample standard deviation of $\hat{\theta}$ is also its empirical SE (SE_{EP}). Notice that it is unlikely for an empirical bias to be zero due to sampling error. We will evaluate the significance of each bias by referring

$$t = \frac{\sqrt{500}\text{Bias}}{\text{SE}_{EP}} \quad (3)$$

to the Student *t*-distribution with 499 degrees of freedom.

When the population distribution is correctly specified, consistent SEs of $\hat{\boldsymbol{\theta}}$ can be obtained from the inverse of the observed information matrix, which is just the matrix of the negative second derivatives of the log likelihood function. When the population is misspecified, consistent SEs of the MLEs are obtained from the sandwich-type covariance matrix

$$\boldsymbol{\Omega} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1},$$

where **A** is the observed information matrix and **B** is the summation of the cross-products of the first derivative of the log likelihood function. Thus, we have two formula-based SEs for $\hat{\boldsymbol{\theta}}$ in each replication, SE_{OI} based on the observed information matrix and SE_{SW} based on the sandwich-type covariance matrix. Corresponding to SE_{OI} and SE_{SW}, two averages $\bar{\text{SE}}_s$ of the SEs of $\hat{\boldsymbol{\theta}}$ are obtained across the 500 replications. We use the average of the absolute difference (AAD) between the empirical SE and each $\bar{\text{SE}}$ across the fifteen conditions of *N* and p_m to measure the performance of the two formula-based SEs. That is, for SE_{OI},

[4]According to Collins et al. (2001), with less than 10% of the cases containing missing values and the correlation between the two variables are greater than .4, the bias in parameter estimate is negligible even under MNAR mechanism.

[5]We initially also tried 1000 replications for several combinations of conditions and found that the results are essentially the same as with 500 replications. We decided to use 500 replications to save the time of simulation.

$$\text{AAD}=\sum_{N,p_m}|\bar{\text{SE}}_{OI}-\text{SE}_{EP}|/15.$$

A parallel AAD for SE_{SW} is also obtained to measure its performance.

For the normal-distribution-based MI, the so-called Jeffreys noninformative prior is used in data augmentation (see e.g., Schafer, 1997, p. 154). We also need to determine the number of cycles/iterations needed for the Markov chain to converge to its equilibrium distribution. Following the suggestion of Schafer (1997) and Schafer and Olsen (1998), we first used the EM-algorithm (Dempster, Laird, & Rubin, 1977) to estimate the MLE on one random sample from each of the nine distribution conditions. Let the starting values of $\boldsymbol{\mu}^{(0)}$ and $\boldsymbol{\Sigma}^{(0)}$ be respectively the sample means and sample covariance matrix of the complete cases (after performing listwise deletion), $\boldsymbol{\theta}^{(j)}$ be the parameter value of $\boldsymbol{\theta}$ at the *j*th iteration, and the convergence criterion be $\max_{1\le i\le 3}|\theta_i^{(j+1)}-\theta_i^{(j)}|<10^{-4}$, we found that the EM-algorithm converged in less than 50 iterations[6] for all the samples. For example, for the sample from the worst condition $z_1 \sim LN_s(0, 1/2)$ & $z_2 \sim LN_s(0, 1/2)$, $N = 30$, $p_m = .25$, the EM-algorithm converged in 7 iterations. To be conservative, we choose 100 iterations and use the MLE as the starting value for each imputation. For the same sample ($z_1 \sim LN_s(0, 1/2)$, $z_2 \sim LN_s(0, 1/2)$, $N = 30$, $p_m = .25$), we also calculated the autocorrelation of μ_2 with lag= 100 using 200 independent draws from the posterior distribution, and found that it is not significant at the .05 level when the standardized autocorrelation[7] is referred to $N(0, 1)$. We also replicated the above process of calculating the autocorrelation of μ_2 100 times and found that 4 of the auto-correlations are significant at .05 level. Based on all the evidence, we decided to use 100 cycles/iterations to obtain one set of imputed values for all the simulation conditions, with the MLEs $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\theta}}$ as the starting value of each Markov chain.

We also need to determine the number of imputations n_I for each missing value. We tried $n_I = 10$, 30, and 50 for $z_1 \sim N(0, 1)$ and z_2 following the three conditions in Table 1. We could not notice any systematic difference on empirical biases and SEs corresponding to the three n_Is, we ended up choosing $n_I = 30$. Schafer and Olsen (1998) and von Hippel (2007) noted that $n_I = 10$ is enough for most practical purposes while Graham, Olchowski and Gilreath (2007) found that a greater n_I may be needed to achieve a better power when the effect size is small.

According to equation (4.21) of Schafer (1997, p. 109) or equation (5.17) of Little and Rubin (2002, p. 86), $\tilde{\boldsymbol{\theta}}$, the MI parameter estimates for each replication are obtained by the average of the MLEs[8] across the $n_I = 30$ completed (with the imputed values) samples. Parallel to those for MLE, the average of the $\tilde{\boldsymbol{\theta}}$ as well as the sample standard deviation (denoted as SE_{EP}) for each element of $\tilde{\boldsymbol{\theta}}$ across the 500 replications are obtained. Empirical biases for parameter estimates by MI are subsequently evaluated as well as the

[6]The number of iterations depends on a defined convergence criterion. The convergence criterion, 10^{-4}, is from the consideration that we only report each parameter estimate to the 3rd decimal place.
[7]The standard error of the autocorrelation was calculated using the formula provided in equation (4.50) of Schafer (1997).
[8]With complete data, the MLEs are just the sample means and sample variances-covariance.

corresponding *t*-statistic parallel to (3). The formula for calculating the covariance matrix of $\tilde{\boldsymbol{\theta}}$ is given by (see eq. 5.1 of Allison, 2001; eq. 4.22 to 4.24 of Schafer, 1997; eq. 5.18 to 5.20 of Little & Rubin, 2002)

$$\mathbf{V}=\overline{\mathbf{V}}_c+(1+\frac{1}{n_I})\mathbf{V}_s,$$

, where $\mathbf{V}_c$ represents the formula-based estimate of the covariance matrix of the ML estimates of $\boldsymbol{\theta}$ for each sample with imputed values, $\bar{\mathbf{V}}_c$ is the average of $\mathbf{V}_c$ across the n_I imputations, and $\mathbf{V}_s$ is the sample covariance matrix of the ML estimates across the n_I completed samples. Parallel to that in ML, we have two $\mathbf{V}_c$s for each completed sample, one is the inverse of the observed information matrix and the other is the sandwich-type covariance matrix. Thus, we also have two formula-based SEs for each $\tilde{\theta}$, SE_{OI} and SE_{SW}. The average of each of the two SEs across 500 replications is obtained. Each of them is compared against SE_{EP} using the average of their absolute difference (AAD) across the fifteen conditions of N and p_m to measure the performance of the two formula-based SEs in MI.

For each distribution condition, the average of the SE_{EP}s across the fifteen conditions of N and p_m is also calculated for each element of $\hat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$, respectively. These averages are used to compare the relative efficiency of $\hat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$.

Study design with five variables

Parallel to the population model in (1), the population of the five variables is formulated by

$$\mathbf{y}=\boldsymbol{\mu}+\mathbf{A}\mathbf{z},$$

where $\boldsymbol{\mu} = (1, 2, 3, 4, 5)'$; $\mathbf{A}$ is a lower-triangular matrix such that $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}' = (\sigma_{ij})$ with $\sigma_{ii} = 1$ and $\sigma_{ij} = .5$ when $i \neq j$; and $\mathbf{z} = (z_1, z_2, z_3, z_4, z_5)'$ with z_js being standardized independent random variables. Four distribution conditions are chosen on $\mathbf{z}$: (I) $z_j \sim N(0, 1)$, $j = 1, 2, 3, 4, 5$; (II) $z_1, z_2 \sim N(0, 1)$ & $z_3, z_4, z_5 \sim LN_s(0, 1/2)$; (III) $z_1, z_2 \sim LN_s(0, 1/2)$ & $z_3, z_4, z_5 \sim N(0, 1)$; (IV) $z_j \sim LN_s(0, 1/2)$, $j = 1, 2, 3, 4, 5$. The skewenesses and kurtoses of y_3, y_4 and y_5 for the four conditions are within the range of those reported in Table 1, and are not reported here to save space. For each distribution condition, the missing data schemes are created by removing (y_3, y_4, y_5) when $y_1 + y_2$ is greater than its 90th, 70th and 50th population quantiles, respectively. Because y_1 and y_2 are both completely observed, the missing data are MAR and their proportions are p_m =6%, 18% and 30%, respectively. Same as for the two-variable design, the sample sizes are 30, 50, 100, 200 and 500; and the number of replications is 500. The number of imputations as well as the number of iterations to obtain an imputation are also the same as for the two-variable design.

The vector of parameters associated with the complete data is $\boldsymbol{\gamma} = (\mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22})'$ and that associated with missing data is

$$\boldsymbol{\theta}=(\mu_3,\mu_4,\mu_5,\sigma_{31},\sigma_{41},\sigma_{51},\sigma_{32},\sigma_{42},\sigma_{52},\sigma_{33},\sigma_{43},\sigma_{53},\sigma_{44},\sigma_{54},\sigma_{55})'.$$

Parallel to the two-variable design, for each parameter in $\boldsymbol{\theta}$ we will evaluate the empirical bias in its estimates by ML and MI as well as the corresponding SE_{EP}, SE_{OI} and SE_{SW}.

A note on imputed values

To better understand the results of MI in the next section, we would like to note that, for the sample in (2), each imputed value of y_2 is obtained by the regression equation

$$y_2=a+by_1+e, \quad (4)$$

where $e \sim N(0, \sigma^2)$, a, b and σ^2 are determined by $\mathbf{y} = (y_1, y_2)' \sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ and the Jeffreys prior. While the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are obtained by sampling from the posterior distribution, conditional on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, e and y_1 are independent. When substituting z_1 in equation (1) by $(y_1 - \mu_1)/\sigma_1$, we may rewrite the y_2 in (1) as

$$y_2=\mu_2+\frac{\sigma_2\rho(y_1-\mu_1)}{\sigma_1}+\sigma_2(1-\rho^2)^{1/2}z_2=[\mu_2-\frac{\sigma_2\rho\mu_1}{\sigma_1}]+\frac{\sigma_2\rho y_1}{\sigma_1}+\sigma_2(1-\rho^2)^{1/2}z_2. \quad (5)$$

Notice that the last term in (5), call it the error term, has a mean zero and variance $\sigma_{22}(1-\rho^2)$. Obviously, equations (4) and (5) are parallel. Actually, for a given $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$,

$$a=\mu_2-b\mu_1, b=\frac{\sigma_{12}}{\sigma_{11}}, \sigma^2=\mathrm{Var}(e)=\sigma_{22}=\frac{\sigma_{21}^2}{\sigma_{11}},$$

which are identical to the intercept, slope and error variance in (5). Regardless of the distribution of z_2 in (5), MI substitutes each missing y_2 by (4) with a normally distributed e. When z_2 has heavier or lighter tails than that of a normal distribution, neither $\mathbf{V}_c$ nor $\mathbf{V}_s$ is consistent with that corresponding to (1) or (5). Thus, we would expect the normal-distribution- based MI not to work well when z_2 or the conditional distribution of the missing variables given the observed ones is substantially different from normal. Also notice that the covariance of y_1 and y_2 stems from y_1 being on the right side of equation (4), not related to e. We would expect that the estimates of σ_{12} and their SEs to be less related to the distribution of e or z_2.

Results

We will first present the results for the two-variable design before turning to the results for the five-variable design. Since our main interest is in the performance of ML and MI when the population distribution varies, we arrange the results according to the 9 distribution conditions as reported in Table 1. Due to space limitation, for each population distribution of the two-variable design, we will only include in the paper the empirical bias of individual parameter estimates at each studied condition. Results of SEs for individual parameter estimates at each missing data proportion and sample size are put on the web at www.nd.edu/~kyuan/ML-MI/ and will be referred to in the discussion. Instead, we will

include in the paper the average of the empirical SEs (SE_{EP}), and the average of the absolute difference (AAD) between SE_{EP} and the formula-based SEs (SE_{OI} or SE_{SW}) across the 15 combinations of sample size and missing data proportion. Tables for all the results of the five-variable design are put on the same web address, and will be referred to when discussing the results in the paper.

Bias and SE of $\hat{\theta}$ and $\tilde{\theta}$ with the two-variable design

We will first report the empirical bias. In each of the tables of empirical bias, significant ones at .05 level are put in boldface, and the number of significance n_s for each estimate is also reported at the end of each table. Notice that a significant bias can be the effect of type I error, and we will summarize the findings for each estimate after presenting the results for all the conditions.

Table 2 contains the empirical biases for the estimates of $\boldsymbol{\theta} = (\mu_2, \sigma_{12}, \sigma_{22})'$ by ML and MI when $z_1 \sim N(0, 1)$ and $z_2 \sim N(0, 1)$. None of the biases for the estimates for μ_2 is statistically significant while 7 of the 15 biases corresponding to $\tilde{\sigma}_{22}$ are significant; 2 of the 15 biases for each of $\hat{\sigma}_{12}$, $\hat{\sigma}_{22}$ and $\tilde{\sigma}_{12}$ are also significant. At $p_m = .25$, $N = 30$, and 50, the empirical bias in $\tilde{\sigma}_{22}$ by MI is about 30% and 15% of the value of $\sigma_{22} = 1$, respectively. All the others in Table 2 are less than 10% of the parameter value, estimates with relatively large bias are $\tilde{\sigma}_{22}$ at $p_m = .15$ and $N = 30$, 50 and $\hat{\sigma}_{22}$ at $p_m = .25$ and $N = 30$. Comparing the numbers under $\tilde{\sigma}_{22}$ at $p_m = .25$ against those at $p_m = .05$ in Table 2, we may notice that the empirical biases at $p_m = .25$ and $N = 100$, 200 are greater than those at $p_m = .05$ and $N = 30$, 50, although more data are available at $p_m = .25$ and $N = 100$, 200. Thus, when the sample size is not large enough, a large proportion of missing values can bring substantial bias to the variance parameter estimates by MI even when the population is normally distributed.

Empirical biases of $\hat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$ when $z_1 \sim N(0, 1)$ and $z_2 \sim \log N_s(0, 1/2)$ are presented in Table 3. Although the distribution condition is not what MI or ML is designed for, the biases in Table 3 are only slightly larger than those in Table 2, with only 5 of the 15 biases corresponding to $\tilde{\sigma}_{22}$ being statistically significant at .05 level. Similar to those in Table 2, the largest biases are with $\tilde{\sigma}_{22}$ at $p_m = .25$ and $N = 30$, 50 and at $p_m = .15$ and $N = 30$.

Table 4 contains the empirical biases when $z_1 \sim N(0, 1)$ and $z_2 \sim U_s(0, 1)$. Again, the values in Table 4 are very comparable to those in Table 2, and relatively larger ones are associated with $\tilde{\sigma}_{22}$ at $p_m = .25$, $N = 30$, 50, and $p_m = .15$, $N = 30$.

Table 5 contains the empirical biases when $z_1 \sim \log N_s(0, 1/2)$ and $zs_2 \sim N(0, 1)$, where none of the entries under $\hat{\mu}_2$ or $\tilde{\mu}_2$ is statistically significant at .05 level. However, the biases for both $\tilde{\sigma}_{22}$ and $\tilde{\sigma}_{22}$ in Table 5 are obviously a lot worse than in any of the previous conditions. That for MI at $p_m = .25$ and $N = 30$ is about 2.5 times of the value of the parameter itself. Although smaller compare to MI, the MLE $\tilde{\sigma}_{22}$ also contains substantial bias with smaller N at $p_m = .25$ and .15. Actually, the nonnormality of y_2 is solely due to z_1, whose information is observed through y_1. So it is the nonnormality of the observed variable that creates substantial biases. As N increases, the biases in both $\tilde{\sigma}_{22}$ and $\tilde{\sigma}_{22}$ decrease. In particular, at $N = 500$, the bias in $\tilde{\sigma}_{22}$ becomes .044, due to the consistency of the MLE. The bias in $\tilde{\sigma}_{22}$ also decreases quite faster.

Table 6 contains the empirical biases of $\hat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$ when $z_1 \sim \log N_s(0, 1/2)$ and $z_2 \sim \log N_s(0, 1/2)$, a condition departs most from normality among those in Table 1. The biases in Table 6 are comparable to those in Table 5, implying that it is mainly the interaction of nonnormality of y_1 and missing data in y_2 that caused the biases. Although the entries under $\hat{\mu}_2$, $\hat{\sigma}_{12}$, $\tilde{\mu}_2$ and $\tilde{\sigma}_{12}$ in Table 6 are on average larger than those in Table 2, none of them in Table 6 is statistically significant. This is because the distribution of y_1 as well as the covariance between y_1 and y_2 are determined by the distribution of z_1 in the population. When z_1 has heavy tails, the corresponding SEs of $\hat{\mu}_2$, $\hat{\sigma}_{12}$, $\tilde{\mu}_2$ and $\tilde{\sigma}_{12}$ are also greater (see Table A5(a) and (b)), and their relatively larger values are mostly due to sampling errors.

Table 7 contains the empirical bias when $z_1 \sim \log N_s(0, 1/2)$ and $z_2 \sim U_s(0, 1)$. As in the previous tables with $z_1 \sim \log N_s(0, 1/2)$, $\tilde{\sigma}_{22}$ contains large bias at $p_m = .25$ and smaller Ns. The bias is still more than 10% even when $N = 500$. The MLE $\hat{\sigma}_{22}$ also contains substantial bias at smaller Ns. The biases in both $\hat{\sigma}_{22}$ and $\tilde{\sigma}_{22}$ drop quickly as N increases. Once again, the results imply that the bias is caused mainly by the interaction of the nonnormal distribution of the observed variable and missing data, not the distribution of the random component z_2 that solely belongs to the variable with missing values (see equation 1).

Empirical biases of $\hat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$ when $z_1 \sim U_s(0, 1)$ and $z_2 \sim N(0, 1)$ are reported in Table 8. Obviously, the biases are much smaller when compared to those under the condition with $z_1 \sim \log N_s(0, 1/2)$. But the empirical biases in $\hat{\sigma}_{22}$ and $\tilde{\sigma}_{22}$ are still greater than those in Table 2, especially when $p_m = .25$.

Table 9 contains the empirical bias when $z_1 \sim U_s(0, 1)$ and $z_2 \sim \log N_s(0, 1/2)$. With $\tilde{\sigma}_{22}$ having a relative bias of 34% at $p_m = .25$ and $N = 30$, the results in Table 9 imply once again that the magnitude of the bias associated with MI is closely related to the distribution of y_1, not y_2. Actually, the y_2 for the condition in this table departs from the normal distribution much more than that in Table 8 while the empirical bias in this table is smaller on average.

Table 10 contains the empirical bias when both z_1 and z_2 follow $U_s(0, 1)$. Both y_1 and y_2 have zero skewness and tails lighter than that of the normal distribution. Again, $\tilde{\sigma}_{22}$ has substantial empirical bias at $p_m = .25$ and smaller Ns; $\hat{\sigma}_{22}$ also has a bias of 14% of the parameter value at $p_m = .25$ and $N = 30$. The biases in both $\hat{\sigma}_{22}$ and $\tilde{\sigma}_{22}$ drop quickly as N increases. The empirical biases for other estimates are comparable to those for normally distributed data in Table 2.

As pointed out earlier, a significant bias can be due to sampling errors. Each parameter estimate is evaluated at 135 conditions in Tables 2 to 10, and the percentage of significant empirical biases across the 9 tables are respectively

ML			MI		
$\hat{\mu}_2$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{22}$	$\tilde{\mu}_2$	$\tilde{\sigma}_{12}$	$\tilde{\sigma}_{22}$
6.6%	14.1%	37.0%	5.2%	13.3%	62.2%

Thus, the estimates of μ_2 by ML and MI contain essentially no bias while $\tilde{\sigma}_{22}$ is most biased.

Table 11 contains the average of the SE_{EP}s of $\hat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$ across the 15 combined conditions of p_m and N, which imply that the MI estimates are not as efficient as the MLEs in general. The lack of efficiency in $\tilde{\boldsymbol{\theta}}$ can also be observed from individual SE_{EP}s reported at www.nd.edu/~kyuan/ML-MI/TableA1-A9.pdf, where SE_{EP}s of individual MLEs are in Tables A1(a) to A9(a) and those of individual MI estimates are in Tables A1(b) to A9(b).

Table 12 contains the AADs between the empirical SEs and the averaged SE_{OI} and SE_{SW} for the MLE $\hat{\boldsymbol{\theta}}$. We may notice that SE_{OI} predicts SE_{EP} slightly better than SE_{SW} for normally distributed data; SE_{OI} also predicts the SE_{EP} of $\hat{\mu}_2$ better for other distribution conditions. However, SE_{SW} predicts the SE_{EP} of $\tilde{\sigma}_{12}$ better when $z_1 \sim LN_s(0, 1/2)$ and predicts the SE_{EP} of $\tilde{\sigma}_{22}$ better when $z_1 \sim N(0, 1)$ & $z_2 \sim LN_s(0, 1/2)$, $z_1 \sim N(0, 1)$ & $z_2 \sim U_s(0, 1)$, $z_1 \sim LN_s(0, 1/2)$ & $z_2 \sim N(0, 1)$, $z_1 \sim LN_s(0, 1/2)$ & $z_2 \sim LN_s(0, 1/2)$, $z_1 \sim U_s(0, 1)$ & $z_2 \sim LN_s(0, 1/2)$, and $z_1 \sim U_s(0, 1)$ & $z_2 \sim U_s(0, 1)$. In particular, for a given p_m, SE_{SW} and SE_{EP} are closer in general as N increases. This clearly does not hold for SE_{OI}. For example, even at $N = 500$, SE_{OI} remains to be substantially below SE_{EP} in Tables A2(a), A5(a) and A8(a), and substantially above SE_{EP} in Tables A3(a) and A9(a). Results in Tables A1(a) to A9(a) imply that SE_{SW} tends to under-predict SE_{EP} when the population has heavier tails, and the under-prediction is especially obvious when N is small. Comparing the results at $p_m = .05$ against those at $p_m = .25$ in Tables A1(a) to A9(a) implies that the under-prediction of SE_{EP} by SE_{SW} is mostly due to smaller sample sizes as well as the nature of the underlying population distribution, and it has little to do with the percentage of missing data. When the population distribution is only slightly heavier than that of the normal distribution (e.g., $z_1 \sim LN_s(0, 1/2)$ & $z_2 \sim U_s(0, 1)$), SE_{OI} may perform slightly better than SE_{SW} due to the under-prediction of SE_{EP} by SE_{SW} for the variance of $\tilde{\sigma}_{22}$.

Parallel to Table 12, the AADs corresponding to the MI estimate $\tilde{\boldsymbol{\theta}}$ are given in Table 13. Notice that the numbers under SE_{OI} and SE_{SW} for $\tilde{\mu}_2$ are identical, which is because the covariance matrix of $\bar{\mathbf{y}}$ is estimated by $\mathbf{S}/N$ when evaluating both SE_{SW} and SE_{OI}, where $\mathbf{S}$ is the sample covariance matrix of the completed sample by imputation. Most AADs for the estimates $\tilde{\sigma}_{12}$ and $\tilde{\sigma}_{22}$ under SE_{SW} are smaller. The results in Tables A1(b), A4(b) and A7(b) indicate that SE_{SW} is very close to SE_{EP} when $z_2 \sim N(0, 1)$ and N is large. However, a large N does not make SE_{SW} and SE_{EP} closer when $z_2 \sim LN_s(0, 1/2)$, as indicated in Tables A2(b), A5(b) and A8(b). In particular, SE_{SW} tends to under-predict the corresponding SE_{EP} when $z_2 \sim LN_s(0, 1/2)$, which is because a heavy-tailed random component is replaced by a normally distributed one in the imputation process. In parallel, SE_{SW} tends to over-predict SE_{EP} when $z_2 \sim U_s(0, 1)$, and a large N does not alleviate the over-prediction as long as p_m remains to be a constant. An apparently odd phenomenon in Table A2(b) is that SE_{SW}s for $\tilde{\sigma}_{22}$ at $p_m = .15$, $N = 200$ and 500 are smaller than the corresponding ones at $p_m = .05$. This is because at $p_m = .15$ more heavy-tailed y_2s are replaced by normally distributed y_2s in the completed samples by imputation. The empirical results, together with the note at the end of the previous section, may imply that SE_{SW} for MI cannot be consistent unless y_2 is conditionally normally distributed given y_1.

For the two-variable design, we have also studied conditions with different correlations and when both variables contain missing values as well as when z_1 and z_2 each follows a

standardized gamma distribution. Except that the MLEs and parameter estimates by MI associated with the missing values are less biased when the correlation of y_1 and y_2 increases, the patterns are the same as those in Tables 2 to 13 and A1 to A9. That is, MI estimates remain to be less efficient and more biased for the variance parameter than the corresponding MLEs. For example, with y_2 missing and p_m = .25 for normally distributed z_1 and z_2, corresponding to N = 30 and 50 the empirical biases in $\tilde{\sigma}_{22}$ are .173 and .113 at ρ = .70, and .343 and .223 at ρ = .30, respectively; while all the empirical biases in $\tilde{\sigma}_{22}$ are in the 2nd decimal place or smaller.

Biases and SEs of $\hat{\theta}$ and $\tilde{\theta}$ with the five-variable design

Results for the five-variable design provide essentially the same information as that for the two-variable design. Tables containing the empirical biases and SEs of $\hat{\theta}$ and $\tilde{\theta}$ are in Tables A10 to A13 at www.nd.edu/~kyuan/ML-MI/TableA10-A13.pdf, where Table A10 contains the empirical biases and SEs when z_1, z_2, z_3, z_4, $z_5 \sim N(0, 1)$, Table A11 contains the results when z_1, $z_2 \sim N(0, 1)$ & z_3, z_4, $z_5 \sim LN_s(0, 1/2)$; Table A12 contains those when z_1, $z_2 \sim LN_s(0, 1/2)$ & z_3, z_4, $z_5 \sim N(0, 1)$, and Table A13 contains those when z_1, z_2, z_3, z_4, $z_5 \sim LN_s(0, 1/2)$. In the two-variable design, we did not notice much bias on $\tilde{\sigma}_{12}$ because y_1 is solely responsible for the covariance and is fully observed. The biases in $\tilde{\sigma}_{43}$, $\tilde{\sigma}_{53}$ and $\tilde{\sigma}_{54}$ in Tables A12(a) and A13(a) are about the size of their population values of .5. Like for the two-variable design, the number of significant empirical biases in Tables A10(a) to A13(a) for variance parameters tend to be more than that for the covariance parameters, and that for mean parameters is the smallest. By averaging across the four tables for each kind of parameters, the percentages of significant entries of empirical bias are respectively

ML			MI		
means	covariances	variances	means	covariances	variances
5.0%	15.2%	34.4%	4.4%	23.7%	68.9%

Again, the estimates of mean parameters by ML and MI are essentially not biased, while the estimates of variance parameters by MI are most biased.

In summary, the Monte Carlo results imply that estimates of mean parameters by both ML and MI contain little bias even with distribution violations. For variance-covariance parameters, when the population distribution of the observed variables is normally distributed, MLEs contain little bias regardless of the population distribution of the missing variables. If the distribution of the observed variables is nonnormal, especially with heavier tails, MLEs of variance-covariance parameters can contain substantial bias at smaller N together with a nontrivial proportion of missing values. On the other hand, the variance-covariance estimates by MI can contain substantial bias when sample size is small and the proportion of missing values is not trivial even when the population is normally distributed. When the distribution of the observed variables has heavier tails than those of a normal distribution, the empirical biases in variance estimates by MI can be more than twice of the parameter values. In particular, the empirical bias in $\tilde{\theta}$ is about 2 to 4 times of that in $\hat{\theta}$ on average.

When data are normally distributed, SE_{OI}s and SE_{SW}s predict the empirical SEs of $\tilde{\theta}$ about equally well; they also perform well for the SEs of $\hat{\theta}$ while SE_{OI}s perform slightly better. When the population is not normally distributed, the formula-based SEs do not perform as well as with a normally distributed population in general, especially when the population distribution has heavy tails and sample size N is small. Under the MAR mechanism, the distribution of the observed variables mainly affects the SEs of estimates of covariances between the observed and missing variables; the population distribution of the missing variables mainly affects the SEs of variance-covariance estimates among the missing variables. When the observed variables are normally distributed or close to being normally distributed, SE_{OI}s and SE_{SW}s are very close for the covariance estimates between the observed variables and missing variables. When either the observed or missing variables are nonnormally distributed, SE_{SW}s for estimates of variance parameters can predict SE_{EP} s a lot better than SE_{OI}s with a medium or large N. Comparing ML with MI, the SE_{EP}s of $\tilde{\theta}$ are slightly better predicted by SE_{SW}s than those of $\hat{\theta}$ when the conditional distribution of the missing values given the observed variables is close to normally distributed or when the sample size is small. When the conditional distribution of the missing variables given the observed variables is nonnormal, SE_{SW}s of variance-covariance estimates in $\tilde{\theta}$ work poorly; and a large sample size does not help. Although the SE_{SW}s for $\hat{\theta}$ are consistent, they tend to under-predict the empirical SEs, especially when sample size is small and the population has heavier tails.

Conclusion and Discussion

Because it is nearly impossible to check the underlying population distribution behind a sample with missing values, a desirable missing data method needs to be robust to distribution violations. Although our studies on the normal-distribution-based ML and MI are limited, the results show a clear picture of pros and cons of the two most promising missing data procedures. Estimates of variance parameters by MI tend to contain substantial bias when the percentage of missing data is not trivial and the sample size is small, which replicates what Demirtas et al. (2008) have found. Our results further indicate that non-normal distribution of the observed variables is mainly responsible for bias in MI estimates. MLEs of variance parameters can also have substantial bias when the population distribution of the observed variables is nonnormal, but the bias is a lot smaller compared to that in estimates by MI. In addition to having smaller bias, MLEs are also more efficient than parameter estimates by MI. Thus, ML is generally preferred in practice, especially at smaller sample sizes. With respect to standard errors, SE_{SW}s are recommended although they tend to under-estimate the true SEs at smaller Ns.

Comparing the results at $p_m = .05$ against those at $p_m = .25$ in Tables 2 to 10 or Table A10(a) to A13(a) suggests that bias in estimates of variance parameters is mainly caused by the percentage of missing values. Comparing the results at $p_m = .05$ against those at $p_m = .25$ in Tables A1 to A13 suggests that biases in SE_{OI}s are mainly caused by departure of the underlying distribution from normality, and those in SE_{SW}s are mostly due to smaller sample sizes.

As we pointed out earlier, the results obtained have direct consequences on many statistical models that are commonly used in social science research. For example, the estimate of the regression coefficient in (4) is $\hat{b}=\hat{\sigma}_{21}/\hat{\sigma}_{11}$ by ML or the average of s_{21}/s_{11} by MI, where s_{11} and s_{21} are the sample variance-covariance of the complete data after imputation. The results in the previous section suggest that σ_{11} can be severely over-estimated by MI when y_1 contains a substantial proportion of missing values and y_2 has heavy tails. Then b will be severely under-estimated. When all the variables have heavier tails than that of the normal distribution, as for the condition with $z_1 \sim LN_s(0, 1/2)$ & $z_2 \sim LN_s(0, 1/2)$ in the two-variable design, the SE of the regression coefficient $\tilde{b}$ or $\hat{b}$ will be substantially under-estimated by SE_{OI} since both the SEs of $\hat{\sigma}_{21}$ and $\hat{\sigma}_{11}$ or $\tilde{\sigma}_{21}$ and $\tilde{\sigma}_{11}$ are substantially under-estimated. A biased parameter estimate plus an under-estimated SE will lead a researcher to believe that the predictor has a much smaller covariate value than really is the case. Similarly, when variance parameters are severely over-estimated, one would have little power in testing an existing mean difference. While the implications of the finding on other models can also be deduced similarly, the actual results in a given analysis depend on the particular distribution of the population, the sample size, how the MAR values are created as well as the proportion of missing values.

Although estimates of variances-covariances by MI can have substantial bias at a smaller N, all the biases decrease as N increases. Since we are not aware of any consistency results of MI with distribution violations, we would like to offer some rationale towards its existence. We have observed in equation (4) that, during the iterations of Markov chain, y_2 is obtained by an independent draw of $e \sim N(0, \boldsymbol{\sigma}^2)$ conditional on a, b, $\boldsymbol{\sigma}^2$. Once y_2 is obtained, a, b and $\boldsymbol{\sigma}^2$ are obtained by a random draw from the posterior distribution of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ conditional on (2) together with the y_2s obtained from (4) (see Schafer, 1997). Thus, conditional on a, b and $\boldsymbol{\sigma}^2$, regardless of the true distribution of $\mathbf{y} = (y_1, y_2)'$, we have

$$\begin{aligned}
&\mathrm{Cov}(y_1, y_2)=b\mathrm{Var}(y_1)=(\sigma_{12}/\sigma_{11})\sigma_{11}=\sigma_{12},\\
&E(y_2)=a+bE(y_1)=(\mu_2-b\mu_1)+bE(y_1)=\mu_2-b[\mu_1-E(y_1)]=\mu_2,\\
&\mathrm{Var}(y_2)=b^2\mathrm{Var}(y_1)+\sigma^2=(\sigma_{12}/\sigma_{11})^2\sigma_{11}+(\sigma_{22}-\sigma_{21}^2/\sigma_{11})=\sigma_{22}.
\end{aligned}$$

For the normal distribution based MI with Jeffreys prior, the posterior distribution of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ only involves the first- and second-order moments of the complete data (with imputation). This suggests that, except for sampling errors, parameter estimates by the normal-distribution- based MI will not depend on the underlying population distribution. If the parameter estimates by MI are consistent when $\mathbf{y} \sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, they will still be consistent when the underlying population is nonnormally distributed. Of course, consistency alone does not tell much whether the method is preferred at a given sample size. Because SEs of $\tilde{\boldsymbol{\mu}}$ only involves the second-order moments, we expected SEs of a mean parameter by MI to be consistent. However, the variance of $\tilde{\boldsymbol{\sigma}}_{22}$ by MI involves the fourth-order moments of the simulated e in (4). Unless the distribution of e matches that of y_2 given y_1 in the population, it is unlikely for MI to generate consistent SEs for an estimator of a variance parameter. Actually, empirical results in Tables A2(b) and A8(b) suggest that SE_{SW}s cannot be consistent for the SEs of parameter estimates by MI.

With the results reported in this paper, we may doubt the value of MI or even ML when the population distribution is unknown, p_m is not trivial and N is not large. Remember that the missing values in this paper are created by removing the y_2 corresponding to the largest y_1. In practice, missing values may occur corresponding to all ranges of values of the observed variables. Then the biases associated with estimates of variances-covariances by either ML or MI should not be as severe as reported in this paper. While being cautious with the use of ML and MI with violation of conditions, these are still the most promising methods before we know the underlying population distribution. If known, ML or MI based on the true underlying population is always preferred. In particular, MI allows a researcher to choose informative priors. With small samples, if prior information is available and properly included, then MI may outperform ML.

A reviewer noted that the distributions of log-transformed variance estimates will be better approximated by normal distributions. This is true when data are normally distributed without any missing values, because the log-transformation stabilizes the variance of the transformed statistic. With either nonnormally distributed data or missing values, the log transformation does not stabilize the variance of $\hat{\sigma}_{jj}$ or $\tilde{\sigma}_{jj}$ anymore. Actually, the results reported in this paper are biases and SEs of the normal-distribution-based MLEs and MI estimates, not their confidence intervals or distributions. Let $\boldsymbol{\beta} = \mathbf{g}(\boldsymbol{\theta})$ be the transformed parameters. Because the MLEs of $\boldsymbol{\beta}$ are given by $\hat{\boldsymbol{\beta}}=\mathbf{g}(\hat{\boldsymbol{\theta}})$, we would get the same biases and SEs for the variance estimates when transforming $\hat{\boldsymbol{\beta}}$ back to $\hat{\boldsymbol{\theta}}$. The same is true if we apply the log transformation to $\tilde{\boldsymbol{\theta}}$. Biases and SEs for variance parameter estimates by MI might be different if one reparameterizes the likelihood function and the prior distributions using $\beta_j = \log \sigma_{jj}$. But the resulting posterior distributions involving β_j will be different from the popular normal-distribution-based MI. For example, the posterior distribution of the covariance matrix involving β_j will not be the same as those given in Little and Rubin (2002, p. 228) or Schafer (1997, p. 184). Further study is needed in this direction.

We hope we have made it clear that the purpose of the paper is to compare ML with the MI methodology, as presented in Rubin (19878), Schafer (1997) and Little and Rubin (2002). Since, to our knowledge, no MI package generates SE_{SW} automatically, we had to write our own codes for the Monte Carlo study[9]. Although we have no doubt with our codes in correctly implementing the MI methodology, we are interested in whether standard software generates similar results. Tables A14 to A15 at www.nd.edu/~kyuan/ML-MI/TableA14-A15.pdf contain the empirical biases and SEs for our two-variable design with $z_1 \sim N(0, 1)$ & $z_2 \sim N(0, 1)$ and $z_1 \sim LN_s(0, 1)$ & $z_2 \sim N(0, 1)$, obtained using SAS Proc MI and SAS Macro language. Since Proc MI does not yield standard errors based on the sandwich-type covariance matrix, SE_{SW} is not included in the two tables. We would like to note that 27 replications out of 500 do not contain any missing cases at $p_m = .05$ and $N = 30$; and 3 replications do not contain any missing cases at $p_m = .05$ and $N = 50$. These cases are discarded automatically by Proc MI. Clearly, results in Table A14 are comparable to those in Table 2 and Table A1, and those in Table A15 are comparable to those in Table 5 and

[9]Results in Tables 2 to 9 and A1 to A13 were obtained using SAS IML. Readers interested in replicating the results are welcome to obtain the codes from the first author of the paper.

Table A4. A reviewer noted that Proc MI and the program Amelia II (Honaker, King & Blackwell, 2009) generate different results. Since MI is a simulation-based methodology, certain differences caused by sampling errors are expected. Actually, different results from two MI programs can be due to different seeds to start the random process, different random number generators or algorithms, different numbers of burning cycles/iterations in performing the Markov chains, and/or different numbers of imputations for each missing value. These factors also make the results in Tables 2 & A1 and 5 & A4 different from those in Tables A14 and A15 even after averaging over 500 replications. However, the same systematic results or patterns are observed between Tables 2 & A1 and A14, and between Tables 5 & A4 and A15. For example, at $p_m = .25$ and $N = 30$, the empirical bias in $\tilde{\sigma}_{22}$ in Table A15 is 2.758, more than twice of the population value $\sigma_{22} = 1.0$.

Acknowledgments

We would like to thank Dr. Wei Zhang at the SAS Institute for conducting the simulation using SAS Proc MI and the SAS Macro language. We also would like to thank the three reviewers for constructive comments on earlier versions of the paper.

References

Allison PD. Multiple imputation for missing data: A cautionary tale. Sociological Methods & Research. 2000; 28:301–309.

Allison, PD. Missing data. Thousand Oaks, CA: Sage; 2001.

Allison PD. Missing data techniques for structural equation modeling. Journal of Abnormal Psychology. 2003; 112:545–557. [PubMed: 14674868]

Anderson TW. Maximum likelihood estimates for the multivariate normal distribution when some observations are missing. Journal of the American Statistical Association. 1957; 52:200–203.

Arminger G, Sobel ME. Pseudo-maximum likelihood estimation of mean and covariance structures with missing data. Journal of the American Statistical Association. 1990; 85:195–203.

Buhi ER, Goodson P, Neilands T. Out of sight, not out of mind: Strategies for handling missing data. American Journal of Health Behavior. 2008; 32:83–92. [PubMed: 18021036]

Choi Y, Golder S, Gilimore MR, Morrison DM. Analysis with missing data in social work research. Journal of Social Service Research. 2005; 31:23–48.

Collins LM, Schafer JL, Kam CK. A comparison of inclusive and restrictive strategies in modern missing-data procedures. Psychological Methods. 2001; 6:330–351. [PubMed: 11778676]

Croy CD, Novins DK. Methods for addressing missing data in psychiatric and developmental research. Journal of the American Academy of Child & Adolescent Psychiatry. 2005; 44:1230–124. [PubMed: 16292114]

D'Agostino RB, Belanger A, D'Agostino RB Jr. A suggestion for using powerful and informative tests of normality. American Statistician. 1990; 44:316–321.

Daniels, MJ.; Hogan, JW. Missing data in longitudinal studies. Boca Raton, FL: Chapman & Hall/CRC; 2008.

Dempster AP, Laird NM, Rubin DB. Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society B. 1977; 39:1–38.

Demirtas H, Freels SA, Yucel RM. Plausibility of multivariate normality assumption when multiply imputing non-Gaussian continuous outcomes: A simulation assessment. Journal of Statistical Computation and Simulation. 2008; 78:69–84.

Enders CK. The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. Psychological Methods. 2001; 6:352–370. [PubMed: 11778677]

Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prevention Science. 2007; 8:206–213. [PubMed: 17549635]

Graham, JW.; Schafer, JL. On the performance of multiple imputation for multivariate data with small sample size. In: Hoyle, R., editor. Statistical strategies for small sample research. Thousand Oaks, CA: Sage; 1999. p. 1-29.

Harel O, Zhou X-H. Multiple imputation: Review of theory, implementation and software. Statistics in Medicine. 2007; 26:3057–3077. [PubMed: 17256804]

Honaker, J.; King, G.; Blackwell, M. Amelia II: A program for missing data. 2009. (http://gking.harvard.edu/amelia/).

Horton NJ, Kleinman KP. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. American Statistician. 2007; 61:79–90. [PubMed: 17401454]

Jamshidian M, Bentler PM. Using complete data routines for ML estimation of mean and covariance structures with missing data. Journal of Educational and Behavioral Statistics. 1999; 23:21–41.

Kenward MG, Carpenter J. Multiple imputation: Current perspectives. Statistical Methods in Medical Research. 2007; 16:199–218. [PubMed: 17621468]

King G, Honaker J, Joseph A, Scheve K. Analyzing incomplete political science data: An alternative algorithm for multiple imputation. American Political Science Review. 2001; 95:49–69.

Lee S-Y, Song X-Y. A unified maximum likelihood approach for analyzing structural equation models with missing nonstandard data. Sociological Methods & Research. 2007; 35:352–381.

Lee W-C, Rodgers JL. Bootstrapping correlation coefficients using univariate and bivariate sampling. Psychological Methods. 1998; 3:91–103.

Little, RJA.; Rubin, DB. Statistical analysis with missing data. 2nd ed.. New York: Wiley; 2002.

Micceri T. The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin. 1989; 105:156–166.

Molenberghs, G.; Kenward, MG. Missing data in clinical studies. Chichester, England: Wiley; 2007.

Olinsky A, Chen S, Harlow L. The comparative efficacy of imputation methods for missing data in structural equation modeling. European Journal of Operational Research. 2003; 151:53–79.

Peng C-YJ, Zhu J. Comparison of two approaches for handling missing covariates in logistic regression. Educational & Psychological Measurement. 2008; 68:58–77.

Peugh JL, Enders CK. Missing data in educational research: A review of reporting practices and suggestions for improvement. Review of Educational Research. 2004; 74:525–556.

Rubin DB. Inference and missing data (with discussions). Biometrika. 1976; 63:581–592.

Rubin, DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.

Schafer, JL. Analysis of incomplete multivariate data. London: Chapman & Hall; 1997.

Schafer JL, Graham JW. Missing data: Our view of the state of the art. Psychological Methods. 2002; 7:147–177. [PubMed: 12090408]

Schafer JL, Olsen MK. Multiple imputation for multivariate missing-data problems: A data analyst's perspective. Multivariate Behavioral Research. 1998; 33:545–571.

Taylor L, Zhou XH. Multiple imputation methods for treatment noncompliance and nonresponse in randomized clinical trials. Biometrics. 2009; 65:88–95. [PubMed: 18397338]

Thomas N. Assessing model sensitivity of the imputation methods used in the national assessment of educational progress. Journal of Educational and Behavioral Statistics. 2000; 25:351–371.

von Hippel PT. Regression with missing Ys: An improved strategy for analyzing multiply imputed data. Sociological Methodology. 2007; 37:83–117.

Yuan K-H. Normal distribution based pseudo ML for missing data: With applications to mean and covariance structure analysis. Journal of Multivariate Analysis. 2009; 100:1900–1918.

Yuan K-H, Bentler PM. Consistency of normal distribution based pseudo maximum likelihood estimates when data are missing at random. American Statistician. 2010; 64:263–267. [PubMed: 21197422]

Yuan K-H, Lambert PL, Fouladi RT. Mardia's multivariate kurtosis with missing data. Multivariate Behavioral Research. 2004; 39:413–437.

Table 1

The population skewness and kurtosis of y_2 in the nine distribution conditions.

	z_2					
	N (0,1)		log N_s(0,1/2)		U_s(0,1)	
z_1	skewness	kurtosis	skewness	kurtosis	skewness	kurtosis
N(0,1)	.000	.000	1.909	10.410	.000	−.675
LN_s(0,1/2)	.367	1.157	2.276	11.567	.367	.482
U_s(0,1)	.000	−.075	1.909	10.335	.000	−.750

Table 2

Empirical biases of the MLEs ($\hat{\boldsymbol{\theta}}$) and MI estimates ($\tilde{\boldsymbol{\theta}}$) when $z_1 \sim N(0, 1)$ & $z_2 \sim N(0, 1)$; entries in boldface are statistically significant at .05 level.

		ML			MI		
p_m	n	$\hat{\mu}_2$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{22}$	$\tilde{\mu}_2$	$\tilde{\sigma}_{12}$	$\tilde{\sigma}_{22}$
.05	30	−.005	**−.025**	−.019	−.003	**−.023**	.010
	50	.006	**−.026**	**−.027**	.006	**−.026**	−.014
	100	−.003	−.003	−.005	−.003	−.002	.003
	200	−.001	.001	.003	−.001	.000	.007
	500	.000	.000	.002	.000	.000	.004
.15	30	.018	−.018	−.007	.017	−.019	**.089**
	50	.002	−.006	.012	.001	−.008	**.066**
	100	.010	−.009	−.003	.008	−.011	**.020**
	200	−.006	−.005	−.008	−.005	−.004	.006
	500	−.005	.000	−.002	−.005	.001	.004
.25	30	−.018	−.013	**.063**	−.021	−.015	**.302**
	50	−.018	−.011	.014	−.015	−.008	**.145**
	100	−.016	−.017	−.019	−.016	−.017	**.040**
	200	.006	.004	.007	.006	.003	**.035**
	500	−.005	−.004	−.002	−.005	−.004	.010
	n_s	0	2	2	0	2	7

Table 3

Empirical biases of the MLEs ($\hat{\boldsymbol{\theta}}$) and MI estimates ($\tilde{\boldsymbol{\theta}}$) when $z_1 \sim N(0, 1)$ & $z_2 \sim LN_s(0, 1/2)$; entries in boldface are statistically significant at .05 level.

		ML			MI		
p_m	n	$\hat{\mu}_2$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{22}$	$\tilde{\mu}_2$	$\tilde{\sigma}_{12}$	$\tilde{\sigma}_{22}$
.05	30	−.005	**−.029**	**−.058**	−.004	**−.027**	−.032
	50	.005	**−.024**	−.014	.005	**−.025**	−.002
	100	−.002	−.001	−.009	−.002	−.000	−.001
	200	.001	.003	−.002	.001	.003	.001
	500	.001	.002	.006	.001	.002	.007
.15	30	.019	−.015	.018	.018	−.017	**.117**
	50	.003	−.013	.008	.002	−.014	**.063**
	100	.006	−.011	−.009	.004	−.013	.014
	200	**−.008**	−.006	**−.031**	−.007	−.004	−.017
	500	**−.006**	−.001	−.014	**−.006**	−.000	−.008
.25	30	−.008	−.006	.095	−.011	−.008	**.339**
	50	−.015	−.006	−.011	−.014	−.004	**.116**
	100	−.018	−.013	−.024	−.018	−.012	.033
	200	.006	.005	.011	.006	.005	**.039**
	500	−.006	−.005	−.015	−.005	−.005	−.003
	n_s	2	2	2	1	2	5

Table 4

Empirical biases of the MLEs ($\hat{\boldsymbol{\theta}}$) and MI estimates ($\tilde{\boldsymbol{\theta}}$) when $z_1 \sim N(0, 1)$ & $z_2 \sim U_s(0, 1)$; entries in boldface are statistically significant at .05 level.

		ML			MI		
p_m	n	$\hat{\mu}_2$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{22}$	$\tilde{\mu}_2$	$\tilde{\sigma}_{12}$	$\tilde{\sigma}_{22}$
.05	30	−.007	−.007	−.011	−.007	−.006	.013
	50	.000	**−.027**	**−.021**	.001	**−.027**	−.006
	100	−.003	−.010	**−.013**	−.003	−.010	−.005
	200	−.003	.001	.001	−.003	.001	.004
	500	.000	.001	−.002	.000	.001	−.000
.15	30	−.001	−.012	.009	−.001	−.012	**.101**
	50	−.002	−.018	−.005	−.003	−.018	**.047**
	100	.004	**−.014**	**−.015**	.004	**−.014**	.009
	200	.006	.009	**.011**	.005	.008	**.022**
	500	−.001	−.000	−.001	−.001	−.000	.004
.25	30	−.036	−.019	**.046**	−.038	−.021	.282
	50	**.035**	**−.034**	.009	**.036**	**−.036**	**.138**
	100	−.005	−.009	−.004	−.005	−.009	**.056**
	200	−.006	−.009	.002	−.007	−.009	**.030**
	500	−.006	−.004	−.002	−.006	−.004	**.010**
	n_s	1	3	5	1	3	8

Table 5

Empirical biases of the MLEs ($\hat{\boldsymbol{\theta}}$) and MI estimates ($\tilde{\boldsymbol{\theta}}$) when $z_1 \sim LN_s(0, 1/2)$ & $z_2 \sim N(0, 1)$; entries in boldface are statistically significant at .05 level.

		ML			MI		
p_m	n	$\hat{\mu}_2$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{22}$	$\tilde{\mu}_2$	$\tilde{\sigma}_{12}$	$\tilde{\sigma}_{22}$
.05	30	−.004	−.008	**.064**	−.003	−.007	**.171**
	50	.002	**−.040**	−.002	.002	**−.039**	**.049**
	100	−.002	−.018	.006	−.001	−.017	**.031**
	200	−.001	−.006	.008	−.002	−.008	**.019**
	500	.001	−.002	.006	.001	−.002	.010
.15	30	.024	.014	**.312**	.024	.011	**.725**
	50	.009	.018	**.208**	.008	.018	**.445**
	100	.009	−.014	**.081**	.007	−.016	**.190**
	200	−.010	−.016	**.027**	−.008	−.013	**.085**
	500	−.006	−.006	.010	−.006	−.006	**.032**
.25	30	−.010	−.015	**1.264**	−.014	−.021	**2.464**
	50	−.042	−.062	**.635**	−.040	−.061	**1.307**
	100	−.032	−.056	**.251**	−.033	−.055	**.581**
	200	.008	.009	**.150**	.008	.009	**.320**
	500	−.012	−.019	**.044**	−.010	−.016	**.113**
	n_s	0	1	10	0	1	14

Table 6

Empirical biases of the MLEs ($\hat{\boldsymbol{\theta}}$) and MI estimates ($\tilde{\boldsymbol{\theta}}$) when $z_1 \sim LN_s(0, 1/2)$ & $z_2 \sim LN_s(0, 1/2)$; entries in boldface are statistically significant at .05 level.

		ML			MI		
p_m	n	$\hat{\mu}_2$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{22}$	$\tilde{\mu}_2$	$\tilde{\sigma}_{12}$	$\tilde{\sigma}_{22}$
.05	30	−.005	−.016	.009	−.004	−.017	**.102**
	50	.002	−.031	.023	.002	−.031	**.075**
	100	−.001	−.018	−.001	−.001	−.016	.025
	200	.002	−.001	.004	.001	−.003	.015
	500	.002	−.000	.010	.002	−.001	.014
.15	30	.025	.020	**.341**	.026	.017	**.778**
	50	.005	−.003	**.186**	.004	−.005	**.422**
	100	.002	−.022	**.065**	.001	−.023	**.176**
	200	−.011	−.017	.002	−.010	−.013	**.059**
	500	−.007	−.007	−.003	−.007	−.007	.018
.25	30	.012	.030	**1.413**	.009	.025	**2.643**
	50	−.032	−.038	**.626**	−.032	−.037	**1.275**
	100	−.029	−.046	**.235**	−.031	−.046	**.555**
	200	.010	.013	**.154**	.010	.014	**.322**
	500	−.014	−.019	**.033**	−.012	−.017	**.102**
	n_s	0	0	8	0	0	11

Table 7

Empirical biases of the MLEs ($\hat{\boldsymbol{\theta}}$) and MI estimates ($\tilde{\boldsymbol{\theta}}$) when $z_1 \sim LN_s(0, 1/2)$ & $z_2 \sim U_s(0, 1)$; entries in boldface are statistically significant at .05 level.

		ML			MI		
p_m	n	$\hat{\mu}_2$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{22}$	$\tilde{\mu}_2$	$\tilde{\sigma}_{12}$	$\tilde{\sigma}_{22}$
.05	30	−.008	−.041	.041	−.008	−.040	**.126**
	50	−.002	−.022	.020	−.001	−.019	**.077**
	100	−.003	−.011	.006	−.003	−.009	**.033**
	200	−.002	−.005	.009	−.002	−.006	**.021**
	500	.001	.000	.002	.001	.000	.007
.15	30	.004	−.019	**.349**	.004	−.016	**.762**
	50	−.007	**−.050**	**.135**	−.007	**.049**	**.351**
	100	−.000	−.009	**.061**	−.001	−.010	**.172**
	200	.009	.017	**.061**	.008	.015	**.115**
	500	−.001	−.004	.014	.000	−.003	**.036**
.25	30	−.047	−.093	**1.033**	−.048	−.094	**2.128**
	50	**−.066**	**−.102**	**.550**	**−.073**	**−.116**	**1.223**
	100	−.022	−.029	**.257**	−.021	−.020	**.613**
	200	−.010	−.016	**.155**	−.009	−.013	**.316**
	500	−.009	−.007	**.055**	−.009	−.008	**.121**
	n_s	1	2	9	1	2	14

Table 8

Empirical biases of the MLEs ($\hat{\boldsymbol{\theta}}$) and MI estimates ($\tilde{\boldsymbol{\theta}}$) when $z_1 \sim U_s(0, 1)$ & $z_2 \sim N(0, 1)$; entries in boldface are statistically significant at .05 level.

		ML			MI		
p_m	n	$\hat{\mu}_2$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{22}$	$\tilde{\mu}_2$	$\tilde{\sigma}_{12}$	$\tilde{\sigma}_{22}$
.05	30	**−.020**	**−.019**	**−.026**	**−.020**	**−.019**	−.007
	50	.006	**−.020**	**−.036**	.006	**−.020**	**.025**
	100	.000	.002	−.004	.000	.002	.001
	200	.001	−.000	.008	.001	.000	**.011**
	500	−.001	−.004	−.006	−.001	−.004	−.005
.15	30	−.014	**−.031**	−.029	−.012	**−.030**	**.066**
	50	−.008	−.014	−.002	−.010	−.015	**.047**
	100	−.000	−.013	−.010	−.001	**−.014**	.015
	200	.003	−.001	.007	.002	−.002	**.019**
	500	−.000	.000	.001	−.001	.000	.006
.25	30	.007	−.002	**.127**	.007	−.000	**.445**
	50	.009	.005	**.108**	.009	.005	**.285**
	100	.016	.013	**.051**	.015	.011	**.127**
	200	**−.015**	**−.020**	−.005	−.014	**−.019**	**.034**
	500	−.002	−.002	.006	−.002	−.002	**.021**
	n_s	2	4	5	1	5	10

Table 9

Empirical biases of the MLEs ($\hat{\boldsymbol{\theta}}$) and MI estimates ($\tilde{\boldsymbol{\theta}}$) when $z_1 \sim U_s(0, 1)$ & $z_2 \sim LN_s(0, 1/2)$; entries in boldface are statistically significant at .05 level.

		ML			MI		
p_m	n	$\hat{\mu}_2$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{22}$	$\tilde{\mu}_2$	$\tilde{\sigma}_{12}$	$\tilde{\sigma}_{22}$
.05	30	**−.017**	**−.020**	−.044	**−.017**	**−.021**	−.025
	50	.004	−.013	−.034	.004	−.013	−.023
	100	.000	.005	−.005	.000	.005	.001
	200	.004	.000	−.001	.004	.000	.002
	500	−.002	−.003	**−.015**	−.001	−.003	−.013
.15	30	−.021	**−.036**	−.058	−.020	**−.035**	.032
	50	−.004	−.010	−.013	−.005	−.012	.036
	100	.001	−.011	.004	.001	−.012	.031
	200	.007	.003	.022	.006	.002	**.034**
	500	−.001	−.000	−.005	−.001	−.000	−.001
.25	30	−.007	−.017	.050	−.005	−.014	**.343**
	50	.008	.000	**.064**	.006	−.001	**.232**
	100	.008	.004	.041	.007	.002	**.116**
	200	−.007	−.011	−.003	−.006	−.010	**.035**
	500	−.002	−.002	.012	−.002	−.002	**.027**
	n_s	1	2	2	1	2	6

Table 10

Empirical biases of the MLEs ($\hat{\boldsymbol{\theta}}$) and MI estimates ($\tilde{\boldsymbol{\theta}}$) when $z_1 \sim U_s(0, 1)$ & $z_2 \sim U_s(0, 1)$; entries in boldface are statistically significant at .05 level.

		ML			MI		
p_m	n	$\hat{\mu}_2$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{22}$	$\tilde{\mu}_2$	$\tilde{\sigma}_{12}$	$\tilde{\sigma}_{22}$
.05	30	−.006	**−.020**	**−.029**	−.005	−.018	−.008
	50	−.001	**−.015**	**−.031**	−.000	−.014	**−.020**
	100	**.010**	−.006	−.010	**.010**	−.006	−.004
	200	.002	−.000	−.003	.002	−.001	−.001
	500	−.002	−.001	−.002	−.002	−.002	−.002
.15	30	**−.025**	−.004	.005	**−.029**	−.009	**.096**
	50	−.004	−.004	.004	−.004	−.004	**.057**
	100	−.002	−.011	−.012	−.002	−.012	.013
	200	.001	−.002	.000	.001	−.003	**.012**
	500	−.002	**−.006**	**−.007**	−.002	**−.006**	−.001
.25	30	−.021	−.033	**.136**	−.026	−.038	**.462**
	50	.018	.014	**.092**	.021	.018	**.264**
	100	.018	.008	**.046**	.019	.009	**.126**
	200	−.008	−.006	.011	−.008	−.006	**.049**
	500	.004	.005	**.011**	.004	.004	**.025**
	n_s	2	3	7	2	1	9

Table 11

Average empirical standard errors of the MLEs ($\hat{\theta}$) and MI estimates ($\tilde{\theta}$) across 15 conditions (5 missing data proportions times 3 sample sizes)

Distribution condition	ML			MI		
	$\hat{\mu}_2$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{22}$	$\tilde{\mu}_2$	$\tilde{\sigma}_{12}$	$\tilde{\sigma}_{22}$
$z_1 \sim N(0, 1)$ & $z_2 \sim N(0, 1)$	.170	.189	.232	.171	.191	.244
$z_1 \sim N(0, 1)$ & $z_2 \sim LN_s(0, 1/2)$	.169	.188	.458	.170	.190	.498
$z_1 \sim N(0, 1)$ & $z_2 \sim U_s(0, 1)$	.169	.186	.200	.170	.188	.209
$z_1 \sim LN_s(0, 1/2)$ & $z_2 \sim N(0, 1)$	.272	.521	.686	.275	.526	.768
$z_1 \sim LN_s(0, 1/2)$ & $z_2 \sim LN_s(0, 1/2)$	.272	.528	.988	.275	.533	1.170
$z_1 \sim LN_s(0, 1/2)$ & $z_2 \sim U_s(0, 1)$	.267	.499	.642	.270	.507	.709
$z_1 \sim U_s(0, 1)$ & $z_2 \sim N(0, 1)$	.187	.191	.246	.189	.194	.262
$z_1 \sim U_s(0, 1)$ & $z_2 \sim LN_s(0, 1/2)$	.185	.189	.450	.187	.191	.493
$z_1 \sim U_s(0, 1)$ & $z_2 \sim U_s(0, 1)$	.187	.190	.220	.189	.192	.231

Table 12

Average absolute difference between empirical standard errors and formula-based standard errors of MLEs ($\hat{\boldsymbol{\theta}}$) across 15 conditions (5 missing data proportions times 3 sample sizes)

Distribution condition	SE_{OI}			SE_{SW}		
	$\hat{\mu}_2$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{22}$	$\hat{\mu}_2$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{22}$
$z_1 \sim N(0,1)$ & $z_2 \sim N(0,1)$	.008	.008	.009	.009	.015	.020
$z_1 \sim N(0,1)$ & $z_2 \sim LN_s(0,1/2)$	.011	.012	.234	.015	.019	.136
$z_1 \sim N(0,1)$ & $z_2 \sim U_s(0,1)$	.007	.008	.023	.008	.013	.014
$z_1 \sim LN_s(0,1/2)$ & $z_2 \sim N(0,1)$	.015	.123	.122	.020	.085	.113
$z_1 \sim LN_s(0,1/2)$ &$z_2 \sim LN_s(0,1/2)$	.024	.143	.413	.033	.108	.290
$z_1 \sim LN_s(0,1/2)$ & $z_2 \sim U_s(0,1)$	.014	.116	.112	.019	.079	.125
$z_1 \sim U_s(0,1)$ & $z_2 \sim N(0,1)$	.011	.010	.012	.014	.014	.022
$z_1 \sim U_s(0,1)$ & $z_2 \sim LN_s(0,1/2)$	.015	.011	.218	.020	.020	.124
$z_1 \sim U_s(0,1)$ & $z_2 \sim U_s(0,1)$	.013	.009	.025	.015	.012	.018

Table 13

Average absolute difference between empirical standard errors and formula-based standard errors of MI estimates ($\tilde{\boldsymbol{\theta}}$) across 15 conditions (5 missing data proportions times 3 sample sizes)

Distribution condition	SE_{OI}			SE_{SW}		
	$\tilde{\mu}_2$	$\tilde{\sigma}_{12}$	$\tilde{\sigma}_{22}$	$\tilde{\mu}_2$	$\tilde{\sigma}_{12}$	$\tilde{\sigma}_{22}$
$z_1 \sim N(0,1)$ & $z_2 \sim N(0,1)$	.003	.005	.016	.003	.005	.015
$z_1 \sim N(0,1)$ & $z_2 \sim LN_s(0,1/2)$	.006	.006	.242	.006	.008	.201
$z_1 \sim N(0,1)$ & $z_2 \sim U_s(0,1)$	.004	.006	.045	.004	.006	.031
$z_1 \sim LN_s(0,1/2)$ &$z_2 \sim N(0,1)$	.013	.116	.063	.013	.051	.059
$z_1 \sim LN_s(0,1/2)$ & $z_2 \sim LN_s(0,1/2)$	.020	.138	.449	.020	.073	.337
$z_1 \sim LN_s(0,1/2)$ & $z_2 \sim U_s(0,1)$	.008	.108	.061	.008	.047	.073
$z_1 \sim U_s(0,1)$ & $z_2 \sim N(0,1)$	.005	.012	.016	.005	.006	.014
$z_1 \sim U_s(0,1)$ & $z_2 \sim LN_s(0,1/2)$	.008	.007	.223	.008	.009	.186
$z_1 \sim U_s(0,1)$ &$z_2 \sim U_s(0,1)$	.008	.013	.048	.008	.003	.029