

The Pennsylvania State University  
The Graduate School  
The College of Health and Human Development

**HANDLING MISSING DATA IN THE MODELING OF INTENSIVE  
LONGITUDINAL DATA**

A Thesis in  
Human Development and Family Studies  
by  
Linying Ji

© 2016 Linying Ji

Submitted in Partial Fulfillment  
of the Requirements  
for the Degree of

Master of Science

May 2016

The thesis of Linying Ji was reviewed and approved\* by the following:

Sy-Miin Chow  
Associate Professor of Human Development and Family Studies  
Thesis Advisor

Zita Oravec  
Assistant Professor of Human Development and Family Studies

Eva S. Lefkowitz  
Professor of Human Development and Family Studies  
Professor-in-Charge

\*Signatures are on file in the Graduate School.

# Abstract

The availability of intensive longitudinal data has helped spur the use more sophisticated methods for studying change. Unfortunately, missing data issues also arise frequently in such studies. Conventional missing data approaches are fraught with additional computational challenges when applied to intensive longitudinal data, and may not always be applicable due to the broad-ranging measurement characteristics of the covariates. In this study, we consider and illustrate the use of two approaches for implementing multiple imputing (MI) to cope with the missingness in fitting multivariate time series models, including a full MI approach, in which all missing cases are imputed simultaneously, and a partial MI approach, in which missing covariates are imputed with multiple imputation, while missingness in dependent variables are handled with full information maximum likelihood estimation. The performance of these approaches was examined under assumptions of missing completely at random, missing at random, and nonignorable missingness. The advantages and limitations of each approach are evaluated using a simulation study. We further demonstrate the implementation of the procedure in R using empirical data, involving  $n=111$  families in which children's influences on parental conflicts are modeled as covariates over the course of 15 days.

# Table of Contents

List of Tables	v
Acknowledgments	vi
Chapter 1	
<b>Introduction</b>	<b>1</b>
1.1 Missing Data in Intensive Longitudinal Data Analysis . . . . .	1
1.2 Motivating Example . . . . .	5
Chapter 2	
<b>Simulation Study</b>	<b>7</b>
2.1 Simulation Design . . . . .	7
2.1.1 Multiple Imputation Procedures and Evaluation of Simulation Results . . . . .	9
2.2 Simulation Results . . . . .	11
Chapter 3	
<b>Empirical Illustration</b>	<b>21</b>
Chapter 4	
<b>Discussion</b>	<b>24</b>
4.1 Limitations . . . . .	26
Appendix	
<b>R Code for Full/Partial MI, and Pooling Parameter Estimates</b>	<b>28</b>
References	<b>32</b>

# List of Tables

2.1	Summary Statistics of MCAR condition across 500 MC Replications ( $N = 100, T = 15$ )	16
2.2	Summary Statistics of MCAR condition across 500 MC Replications ( $N = 100, T = 75$ )	17
2.3	Summary Statistics of MAR condition across 500 MC Replications ( $N = 100, T = 15$ )	18
2.4	Summary Statistics of MAR condition across 500 MC Replications ( $N = 100, T = 75$ )	18
2.5	Summary Statistics of NMAR condition across 500 MC Replications ( $N = 100, T = 15$ )	19
2.6	Summary Statistics of NMAR condition across 500 MC Replications ( $N = 100, T = 75$ )	20
3.1	Parameter Estimates for Empirical Data . . . . .	23

# Acknowledgments

Special thanks go to Dr. Sy-Miin Chow, who supports, helps, and encourages me all the way towards the completion of this master thesis. I appreciate all the efforts and time that she has invested in this project.

# Chapter 1 | Introduction

## 1.1 Missing Data in Intensive Longitudinal Data Analysis

With new advances in statistical modeling techniques as well as data collection techniques, intensive longitudinal studies are becoming increasingly popular in the behavioral and social sciences. Typical intensive longitudinal data in the social and behavioral sciences encompass both intraindividual changes over time as well as interindividual differences in intraindividual change. They allow researchers to analyze inter-relations among change processes and identify possible antecedents and determinants of interindividual differences in the change processes (Baltes & Nesselroade, 1979).

Despite the promise of intensive longitudinal data, measuring participants repeatedly over many occasions also increases the occurrence of missing data, either due to the non-compliance of participants or study design. For instance, some participants may fail to report on some of the measurement occasions. Also, with event-contingent study designs, data would only be collected when a predefined event occurs, thus resulting in irregularly spaced data and substantial missingness in the data when used with discrete-time or regularly spaced models.

In order to handle missing data appropriately, it is essential to clarify the three basic types of missing mechanisms as defined by Rubin (1976). When missingness does not depend on any data, either observed or missing, we say that the missing is completely at random (MCAR). In this case, complete case estimator is usually unbiased for cross sectional data. But dropping cases with missing values in some of the variables will lead to inefficient use of data, which might result in low power. Missing at random (MAR) refers to the case when missingness does not depend on the unobserved data, conditional on the observed data (Fahrenberg & Myrtek, 2001). In contrast, non-ignorable missing data, or not missing at random (NMAR) means missingness depends on the value that would have been observed, but currently missing. NMAR is also called nonignorable missing data because parameter estimation is very likely to be biased if we do not take into consideration the missing mechanism in the estimation process. In practice, even though we usually assume MAR, it is impossible to test this assumption. This is because the true missing data mechanism is usually unknown, and the information needed to test the nature of the missing data mechanism may be

absent (Allison, 2003). With intensive longitudinal data, as time dependency of the data points play an important role in estimation, it is especially crucial to handle the missing data appropriately.

Inappropriate treatment of missing data is known to lead to estimation problems, including bias in parameter point estimates (Jones, 1996; Allison, 2003) and standard error estimates (Glasser, 1964), and reduction in power of the model (Afifi & Elashoff, 1966). Listwise deletion, or complete case analysis, is perhaps the most commonly used method in handling missing data. It is also the default method in most statistical software packages (Ibrahim, Chen, Lipsitz, & Herring, 2005). With this method, only those subjects whose data are completely observed are included in the analysis. Model fitting procedures would be performed as if the data set after the removal of listwise missingness is the complete data set. As noted above, in the context of cross-sectional models, parameter point estimates are usually unbiased with listwise deletion if the missingness occurs completely at random. However, listwise deletion is also potentially wasteful and inefficient: with fewer observations in the data set, the standard error estimates may be biased, and power may also be reduced. Listwise deletion is even less ideal for longitudinal data models, even if we are confident that the missingness occurs completely at random. Time dependency is important in modeling longitudinal data, as we are usually interested in examining the trajectories of the variables of interest and how they are predicted by previous (lagged) data from earlier time points. If listwise deletion method is applied, the time intervals between observations will be incorrectly altered, which will change the trajectories and time dependency between each observations. As a result, both point estimates and standard error estimates will likely be biased, especially for parameters that affect the lagged and time dependencies among successive occasions.

Another widely implemented approach to handling missing data in longitudinal studies is to use pattern-mixture models. Pattern-mixture models specify:

$$p(Y, R|X) = p(Y|R, X)p(R|X),$$

where  $Y$  represents the dependent variable,  $R$  represents the missing-data indicator matrix and  $X$  represents covariates. With this method, the observed data are grouped by missing data patterns (Little, 1993). That is, a model is fit to each group of observed data, resulting in one set of parameter estimations for each group. Those parameter estimates are averaged (or weighted averaged) over groups to obtain the overall parameter estimates for all observed data. This method has been applied to structural equation modeling of longitudinal data (Muthén, Kaplan, & Hollis, 1987; Allison, 1987), latent growth structural equation models (McArdle & Hamagami, 1992), and random-effects models (Hedeker & Gibbons, 1997). Verbeke and Molenberghs (2000/2009) provided detailed example of pattern-mixture modeling. Even though this method is based on a very intuitive way of decomposing  $p(Y, R|X)$  and is very popular in the structural equation modeling and random-effects literature, it is cumbersome to use with intensive longitudinal data because long time series may result in too many possible missing patterns to model. For instance, three measurement occasions will generate eight ( $2^3$ ) possible missing-data patterns, and the number of missing-data patterns will increase substantially as the number of measurement time points increases. There-



fore, it is often not practical to use pattern-mixture modeling in intensive longitudinal data analysis.

Interpolating the missing data is one option when a SEM-based approach is used for fitting dynamic models to longitudinal data with intensive measurement occasions (Molenaar, 1985). With this approach, missed observations are filled in using spline or other models deemed flexible enough to approximate a variety of change trajectories. Then, a block Toeplitz matrix, which serves as the observed covariance matrix in SEM software (Chow, Ho, Hamaker, & Dolan, 2010), can be created based on the interpolated data. The `na.approx()` function in zoo package (Zeileis & Grothendieck, 2005) is a tool to perform linear interpolation, which replaces the missing value with the linear interpolated data. Though practical and easy to implement this interpolation function only takes into consideration the times series variable itself and does not account for influences of covariates. It also assumes linearity and does not have higher order options. The `akima()` function in akima package (Akima, 1970, 1991) is another option to perform data interpolation. This function can be used to interpolate any specified points between two observed data points using polynomials up to degree three. Alternatively, Liu and Molenaar (2014) proposed to use `iVAR`, which is an R program for handling missingness in multivariate time series data by interpolating on the basis of VAR models. This program is helpful when the time series process is stationary, the beginning of the time series is fully observed, and the times series variables follow multivariate normal distribution. All these interpolation methods would only generate one value for each missing observation, and the interpolated data set would be used as if it were the fully observed data set in model fitting. These procedures do not account for the uncertainty in the interpolated data, and may lead to under estimation of parameter standard error. In addition, these interpolation methods assume MCAR or MAR, and thus may not work for NMAR cases, as all these methods don't take into consideration the missingness mechanism and perform interpolation only based on the observed data points in the time series variable, while under NMAR condition, missingness mechanism depends on unobserved data. Therefore, parameter point estimates and standard error estimates are likely to be biased under NMAR condition with interpolation.

Other well-known methods for dealing with missing data include full-information maximum likelihood (FIML) and multiple imputation (MI) (Collins, Schafer, & Kam, 2001). Both methods have been shown to produce consistent estimates under certain conditions (i.e. MAR) for regression models (Thoemmes & Rose, 2014), but most of the work was targeted toward cross-section data (Little & Rubin, 2014; Schafer, 1997; Allison, 2002; Ibrahim, 1990; Ibrahim et al., 2005; Horton & Kleinman, 2007). For instance, Ibrahim et al. (2005) reviewed and compared four ways of handling missing missing covariate data in generalized linear models, including maximum likelihood (ML), multiple imputation (MI), fully Bayesian (FB) methods, and weighted estimating equations (WEE). The authors concluded that in general, all four methods yielded comparable results and outperformed listwise deletion method, which tends to produce inflated standard errors and mean squared errors (MSEs). Based on simulation study results, when correctly specified models were used, standard errors, MSEs, and coverage probabilities in WEE were worse than those for ML, MI and FB; however, the WEE procedure was slightly more robust than the other three when the

models for the covariate distribution were misspecified. In the context of longitudinal models, we would expect estimation bias in both parameter estimates and parameter standard error estimates with complete case analysis, but whether the above examined methods still work well remain unclear.

FIML approach can be used to fit dynamic models with state-space models, as performed by programs such as the `mkfm6` and `Ox` (and its state-space library `SSfpack`) (Chow et al., 2010). This approach remains one of the “gold standard” approaches for handling missing data in the dependent variables when the model is correctly specified, the data are missing at random or completely at random, and the likelihood function has a close analytic form. However, this approach does not handle missingness in the covariates. One possible way to circumvent this problem is to include the covariates as dependent variables in the model, in which case the missingness can be handled via standard FIML estimation. Unfortunately, this method is not always applicable or practically feasible when many covariates are involved and some of them are non-normal or categorical in nature. Also, in many cases, it may not be desirable to include covariates as additional dependent variables in the dynamic model due to the added computational costs and the fact that the dynamics of the covariates may not be the focus of direct modeling interest.

MI, a commonly adopted approach for handling missingness, is first proposed by Rubin (1977), and elaborated in his book (Rubin, 2004). It is based on the premise that in MI,  $m > 1$  versions of multiple imputed data sets are first created by filling in the missing values using a missing data model of choice. These data sets are then analyzed as if they were complete data, using standard complete-data methods. Parameter estimations for the model of interest are then pooled to produce inferential statement, which incorporates the missing-data uncertainty. This method has the advantage of preserving the relations of the variables in the data and meanwhile accounting for the uncertainty about these relations (van Buuren & Groothuis-Oudshoorn, 2011). It has been widely applied in cross-sectional survey data (Rubin, 1996) and is one of the most popular missing data approaches (Schafer & Graham, 2002; Rubin, 1996; Allison, 2000; Sinharay, Stern, & Russell, 2001; Zhang, 2003; Harel & Zhou, 2007). Even though the underlying missing data mechanism may or may not conform to the MAR mechanism, inclusion of appropriate auxiliary variables – namely, observed variables that are not of substantive interest but may be related to aspects of the missing data mechanism – helps approximate MAR. The auxiliary variables are included in the imputation model only for the purpose of improving estimates of the missing data, reducing error variance and increasing the precision of the parameter estimates (Thoemmes & Rose, 2014). Van Buuren, Boshuizen, Knook, et al. (1999) recommended including a variable as auxiliary variable if the absolute value of its correlation with either missingness or the variable with missing data exceeds .1. Collins et al. (2001) suggested a threshold of .4. The validity of including auxiliary variables has been assessed and proven both theoretically (Schafer, 1997; Rubin, 1996; Meng, 1994) and through simulation studies (Collins et al., 2001).

MI can be implemented with different techniques and software programs (e.g. Schafer, 1999;

King, Honaker, Joseph, & Scheve, 2001; van Buuren & Groothuis-Oudshoorn, 2011; Raghunathan, Lepkowski, Van Hoewyk, & Solenberger, 2001). An extensive review of statistical computing software for MI packages for regression models has been provided by Horton and Lipsitz (2001), and a list of software for MI can also be found at <http://www.stefvanbuuren.nl/mi/Software.html>. In this study, we use MI by chained equation (MICE; van Buuren & Groothuis-Oudshoorn, 2011), as implemented in the R package, MICE. With this method, imputations are drawn by iterating over the conditional densities on a variable-by-variable basis by Markov chain Monte Carlo (MCMC) techniques. This program provides considerable flexibility in customizing imputation models for different data characteristics and modeling purposes, and extensive graphical summaries of the MCMC process.

Both MI and FIML are widely applied in handling missing data, but subtle differences between the two in dealing with multivariate time series models have not been tested, particularly in conditions involving data that are not missing at random. The goal of this paper is to critically examine the use of these two procedures under different missingness conditions and compare their estimation results. We evaluated the performance of two MI approaches for imputing missingness in a vector autoregressive (VAR) model, including (1) a full MI approach, in which all missing variables are imputed simultaneously and the vector autoregressive model is then fitted and pooled over multiple sets of complete (imputed) data sets, and (2) a partial MI approach, in which MI is only performed on the missing covariates while missingness in the dependent variables is handled via FIML estimation. Using a series of simulation studies, we evaluate and demonstrate the use of the full MI and partial MI approaches, and compare estimation results to listwise deletion methods under different missingness conditions. Following that, we illustrate the use of MI with a set of empirical data. We conclude with a brief discussion on the strengths and limitations of these approaches, as well as practical guidelines for researchers interested in implementing these two MI-based approaches.

## 1.2 Motivating Example

The current study was inspired by an empirical study that explored the dynamics of interparental emotion states after conflicts and associations with child emotions and behaviors during conflicts (Schermerhorn, Chow, & Cummings, 2010). Longitudinal data were collected from 111 cohabiting couples with a child of 8-16 years of age over 15 days. The study conformed to an event-contingent design in which the parents were asked to respond whenever or shortly after a conflict arose. In particular, the parents were asked to rate their own as well as their children's emotional states and behaviors at each conflict. To analyze over-time and lagged dependencies in the couples' dynamics and possible associations with child-related variable, Schermerhorn et al. (2010) considered the model:

$$\begin{aligned}
\begin{bmatrix} w_{it} \\ h_{it} \end{bmatrix} &= \begin{bmatrix} a_{w \rightarrow w} & b_{h \rightarrow w} \\ b_{w \rightarrow h} & a_{h \rightarrow h} \end{bmatrix} \begin{bmatrix} w_{i,t-1} \\ h_{i,t-1} \end{bmatrix} + \begin{bmatrix} c_{x_1 \rightarrow w} & d_{x_2 \rightarrow w} \\ c_{x_1 \rightarrow h} & d_{x_2 \rightarrow h} \end{bmatrix} \begin{bmatrix} x_{1it} \\ x_{2it} \end{bmatrix} + \begin{bmatrix} e_{w,i,t} \\ e_{h,i,t} \end{bmatrix} \\
\begin{bmatrix} e_{w,i,t} & e_{h,i,t} \end{bmatrix} &\sim N\left(\begin{bmatrix} 0 & 0 \end{bmatrix}, \begin{bmatrix} \sigma_{e_w}^2 & \sigma_{e_{wh}} \\ \sigma_{e_{wh}} & \sigma_{e_h}^2 \end{bmatrix}\right)
\end{aligned} \tag{1.1}$$

The model is a vector autoregressive (VAR) model of order 1, namely, with dependent variables at the current time point predicted by dependent variables at the immediately preceding time point (i.e., a lag of 1). Specifically, the emotional states of the  $i$ th family’s wife and husband ( $i = 1, \dots, n$ ) at the end of the  $t$ th ( $t = 1, \dots, T$ ) conflict ( $w_{it}$  and  $h_{it}$ , respectively) were posited to be influenced by their own emotional states at the  $t - 1$ th conflict, the strengths of which are captured in the auto-regression parameters,  $a_{w \rightarrow w}$  and  $a_{h \rightarrow h}$ . In addition, the couple’s previous emotional states at the previous conflict were also assumed to affect the partner’s emotional state at the  $t$ th conflict, as governed by the cross regression parameters,  $b_{h \rightarrow w}$  and  $b_{w \rightarrow h}$ . Two covariates were included in the dynamic model. The covariate,  $x_1$ , represents the child’s agentic behavior, which includes actions like helping out, taking sides, comforting the parents, trying to make peace. The other covariate,  $x_2$ , is a measure of child’s negative emotions, including anger, sadness, fear, etc.

In this empirical study, a large portion (>67% for all child-related variables) of the child-related covariates was missing because the children were typically not present when their parents were having conflicts (Schermerhorn et al., 2010). In order to handle the missingness, the authors recoded the child-related covariates originally measured on a 10-point scale into a dummy variable such that a child’s value on each covariate was coded as 0 both when the child did not display the behavior during the conflict, and when the child was “missing” (i.e., not present at the conflict); the covariate was coded as 1 when the child showed any level of that behavior. In other words, occasions on which the child was absent were treated to be the same as the occasions on which the child was present but did not display any of the specified behaviors (i.e., agentic behavior and negative emotion). This way of handling the missingness in the covariates may eliminate subtle differences in the child’s influences when he/she was absent vs. present and consequently, alter substantive interpretations of the results. In addition, recoding the originally “continuous” covariates on a 10-point scale into binary variables bears the risk of obscuring other interesting dynamics at the family level. Motivated by these empirical concerns, our goal in this study is to evaluate and compare the previous modeling results with new results obtained from fitting the model in Equation 1.1 when the covariates are included in their original metric units, and their missingness is dealt with via the two MI approaches considered in the present study.

# Chapter 2 | Simulation Study

## 2.1 Simulation Design

Extending previous results of our motivating example, we were interested in comparing the strengths and weaknesses of the two MI approaches under different missing data mechanisms. We simulated two conditions with different sample configurations, namely,  $T = 15$ ,  $n = 100$ , and  $T = 75$ ,  $n = 100$ . The fewer measurement occasions in the first condition were selected to mirror the characteristics of the empirical study described under the motivating example, and is similar in sample size configuration to many other momentary assessment studies in the social and behavioral sciences. The second condition was selected to provide a large-sample comparison to the first condition and specifically, the effects of the number of time points on parameter and standard error estimation in the presence of missingness. The true parameter values of the dynamic model used in the simulation were set to typical ranges of auto- and cross-regression parameters observed in the motivating example as well as other empirical studies in psychology utilizing variations of the VAR model (Chow, Nesselrode, Shifren, & McArdle, 2004; Chow, Hamagami, & Nesselrode, 2007). The values of the model (Equation 1.1) were set as follows:

$$a_{w \rightarrow w} = 0.4, a_{h \rightarrow h} = 0.3, b_{h \rightarrow w} = -0.3, b_{w \rightarrow h} = -0.2, c_{x_1 \rightarrow w} = 0.3, c_{x_1 \rightarrow h} = 0.3, d_{x_2 \rightarrow w} = -0.5, d_{x_2 \rightarrow h} = -0.4,$$

and 
$$\begin{bmatrix} \sigma_{e_w}^2 & \sigma_{e_{wh}} \\ \sigma_{e_{wh}} & \sigma_{e_h}^2 \end{bmatrix} = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}.$$

We were interested in generating both continuous as well as categorical time-varying covariates that might be governed by their own intrinsic dynamics, but whose exact nature of the dynamics were unknown and thus not modeled by the researcher. To do so, we generated two covariates,  $x_{1it}$  and  $x_{2it}$ , whose values depended on two completely observed auxiliary variables,  $x_{3it}$  and  $x_{4it}$ . Both  $x_{3it}$  and  $x_{4it}$  were uniformly distributed over  $[-3, 3]$ . The covariate  $x_{1it}$  was a binary variable generated based on a continuous but unobserved time-varying variable,  $\theta_{it}$ , whose value at time  $t$  for child  $i$  was obtained as  $\theta_{it} = .8 * \theta_{i,t-1} + .4 * x_{4it} + \epsilon_{it}$ ; the value of  $x_{1it}$  was set to 1 when  $\theta$  was larger than a pre-specified threshold and 0 otherwise. In our simulation study, this threshold value was set to be the median of all generated  $\theta_{it}$  values. The covariate  $x_{2it}$  was a continuous variable predicted by one of the external variables  $x_3$  as  $x_{2it} = .6 * x_{3it} + \epsilon_{it}$ , where  $\epsilon$  was distributed as  $N(0, .01)$ .

To evaluate the performance of the MI approaches under different missingness conditions, missing data were generated following three possible missing data mechanisms: MCAR, MAR and NMAR. Let  $r_{w_{it}}, r_{h_{it}}, r_{x_{1it}}, r_{x_{2it}}$  be vectors of missingness indicators for the dependent variables and the covariates, respectively, such that  $r_{it}=1$  if the corresponding variable for person  $i$  at time  $t$  is missing and 0 otherwise. Whether the probability distribution of  $r_{it}$  is conditioned on observed or unobserved data determines the nature of the missingness conditions:

$$pr(r_{it} = 1|Y, X),$$

where  $Y$  represents dependent variables, and  $X$  represents covariates. Details of the missing data generation procedures are described as follows.

For MCAR condition,  $r_{it}$  did not depend on any information, observed or unobserved. We introduced 30% of missingness to all of the dependent variables as well as the covariates completely at random.

Under MAR condition, the probability of observing a missing value in the covariates were conditioned on  $x_{4it}$ , where  $x_{4it}$  was a completely observed external/auxiliary input. To mirror characteristics of the motivating example wherein a child was either present or absent during a conflict, the two child variables,  $x_1$  and  $x_2$ , were either both missing for a certain time point for one family or both not missing. The  $pr(r_{x_{2it}} = 1|x_{4it})$  was modeled by the following logistic regression model  $logit\{pr(r_{x_{2it}} = 1|x_{4it})\} = \varphi_{x_{2_0}} + \varphi_{x_{2_1}} * x_{4it}$ , where  $\varphi_{x_{2_0}} = -1.1$  and  $\varphi_{x_{2_1}} = .7$ . Values of  $\varphi$ s were chosen to achieve desired percentage of missingness and we did not have substantive interest in these values. This yielded approximately 30% of missingness in the covariates.

For NMAR condition, around 30% of missing data were generated for both dependent variables and covariates, following below missingness mechanism models:

- missingness mechanism for wife and husband variables:

$$logit\{pr(r_{w_{it}} = 1|w_{it})\} = \varphi_0 + \varphi_w * w_{it}, \text{ where } \varphi_0 = -0.8 \text{ and } \varphi_w = -.2$$

$$logit\{pr(r_{h_{it}} = 1|h_{it})\} = \varphi_0 + \varphi_h * h_{it}, \text{ where } \varphi_0 = -.8 \text{ and } \varphi_h = -.4$$

The missingness of wife ( $r_{w_{it}}$ ) and husband ( $r_{h_{it}}$ ) variables depended on the values of wife and husband emotion state variables ( $w_{it}, h_{it}$ ). The lower the values were for wife and husband emotion state variables, the more likely the self-reported values would be missing.

- missingness mechanism for child variables:

$$logit\{pr(r_{x_{1it}} = 1|r_{w_{it}}, r_{h_{it}}, x_{1it}, w_{t-1}, h_{t-1}, r_{x_{2it}})\} = \varphi_0 + \varphi_{rw} * r_{w_{it}} + \varphi_{rh} * r_{h_{it}} + \varphi_{rx_2} * r_{x_{2it}} + \varphi_{x_1} * x_{1it} + \varphi_{lagw} * w_{i,t-1} + \varphi_{lagh} * h_{i,t-1}, \text{ where } \varphi_0 = -0.8, \varphi_{rw} = .3, \varphi_{rh} = .1, \varphi_{x_1} = -.2, \varphi_{rx_2} = .2, \varphi_{lagw} = .1, \text{ and } \varphi_{lagh} = .6$$

$$logit\{pr(r_{x_{2it}} = 1|r_{w_{it}}, r_{h_{it}}, x_{2it}, w_{t-1}, h_{t-1})\} = \varphi_0 + \varphi_{rw} * r_{w_{it}} + \varphi_{rh} * r_{h_{it}} + \varphi_{x_2} * x_{2it} +$$

$\varphi_{lagw} * w_{i,t-1} + \varphi_{lagh} * h_{i,t-1}$ , where  $\varphi_0 = -0.8$ ,  $\varphi_{rw} = .3$ ,  $\varphi_{rh} = .1$ ,  $\varphi_{x_2} = .3$ ,  $\varphi_{lagw} = .1$ , and  $\varphi_{lagh} = .6$

The probability of having missing values for child variables, including child agentic behavior ( $r_{x_{1it}}$ ) and child negativity ( $r_{x_{2it}}$ ), would be larger when the parents failed to report their own emotion state scores ( $r_{w_{it}}, r_{h_{it}}$ ), the child had more negative behavior ( $x_{2it}$ ), the child showed less effort in helping out (low score in child agentic behavior,  $x_{1it}$ ) and the parents had lower scores for emotion state after their previous fight ( $w_{t-1}, h_{t-1}$ ).

### 2.1.1 Multiple Imputation Procedures and Evaluation of Simulation Results

We tested two MI approaches in this simulation study: full MI, where all missing values of the dependent variables as well as the covariates were imputed with MICE; and partial MI, where only covariates were imputed with MICE and missingness in dependent variables was handled by the FIML procedure.

For both full MI and partial MI, we included all variables, with or without missing values, as predictors into the imputation model. This is in line with the general advice of including as many relevant variables as possible in MI (Collins et.al. 2001). To be specific, for MCAR and NMAR conditions, both dependent variables,  $w$  and  $h$ , lagged dependent variables, both covariates,  $x_1$  and  $x_2$ , lagged  $x_2$ , and both fully observed variables,  $x_3$  and  $x_4$  were included in the imputation model; for MAR conditions, all the above mentioned variables except lagged dependent variables were included in the imputation model. We used more general models instead of the data generation models to specify the imputation models because the true missingness model is typically unknown in practice. Five imputations were created by calling the `mice()` function in the MICE package. In order to test how number of imputations might influence the imputation results and results of model fitting, we performed the same procedure but increased the number of imputations to ten. The model fitting results were very close to the results when five imputed data set were generated. Therefore, we will only present results from the set of analysis where number of imputations equals to five in this paper. We used default method to generate the imputed data set, that is, predictive mean matching method was used for numerical data and logistic regression method was used for binary variables.

For the full MI approach, the imputed data set, where all missing values were filled in with imputed data, were used for model fitting. In particular, ML parameter estimates were obtained by optimizing a loglikelihood function known as prediction error decomposition – computed using by-products of applying the linear Kalman filter to Equation 1.1 in state-space form (Chow et al., 2010; Harvey, 2001). This procedure was performed with a Kalman filter-based Fortran program, called `mkfm6` (Dolan, 2002), which handles missingness in indicators for the endogenous (dependent) latent variables via FIML. In addition, we wrote “interface” R wrappers to alternate between the MICE routine for performing MI and the call to `mkfm6` to obtain parameter point estimates and standard error estimates. Each of the five imputed data set went through the same model fitting

procedure and five sets of parameter estimates were obtained. The five sets of estimates were pooled into one set of estimates according to Rubin (1996)'s standard procedures:

$$\begin{pmatrix} \text{Posterior mean} \\ \text{of parameter estimates} \end{pmatrix} = \text{Average} \begin{bmatrix} \text{repeated complete-data} \\ \text{posterior means of parameter estimates} \end{bmatrix},$$

meaning the final estimates of the parameters equal the average of the parameter estimates over the repeated imputations.

$$\begin{pmatrix} \text{Posterior} \\ \text{variance of parameter estimates} \end{pmatrix} = \text{Average} \begin{bmatrix} \text{repeated complete-data} \\ \text{variances of parameter estimates} \end{bmatrix} +$$

$$\text{Variance} \begin{bmatrix} \text{repeated complete-data} \\ \text{posterior means of parameter estimates} \end{bmatrix},$$

meaning the final variances of the parameter estimates equal the sum of the average of the variances of parameter estimates over the repeated imputations, and the variances of parameter estimates across imputations.

R codes for pooling the parameter estimates with each imputed data set were provided in the Appendix.

For the partial MI procedure, missingness in dependent variables were handled via FIML by `mkfm6`, while missingness in the covariates were filled in with MI method prior to fitting the model in the `mkfm6`. The model fitting procedures as well as the procedure to pool the parameter estimates were the same as full MI approach. Estimation results of the full MI and partial MI approaches were compared in the 2.2 Simulation Results section.

We considered 2 ( $T = 15$  and  $75$ )  $\times$  3 (MAR vs. NMAR vs. MCAR) = 6 conditions in our simulation study with  $n = 100$  for all conditions. A total of 500 Monte Carlo replications were run for each condition. We employed a range of standard criteria commonly used in simulation studies to analyze the results of our simulation study. In particular, to assess characteristics of the point estimates, we computed both the root mean squared errors (RMSEs) and biases. RMSE for a particular parameter was defined as the square root of the average squared difference between estimates for that parameter ( $\hat{\theta}_h$ ) and the true parameter value ( $true\theta$ ) across Monte Carlo runs (RMSE =  $\sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - true\theta)^2}$ ). Bias was defined as the average difference between estimates of that parameter and the true parameter value across Monte Carlo runs (Bias =  $\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - true\theta)$ ). To evaluate the standard error estimates, we computed difference of standard error (DSE), which was the deviance of the standard deviation of each estimated parameter across Monte Carlo runs (i.e., the empirical standard error), and the average standard error estimate for that parameter across Monte Carlo runs ( $\widehat{SE}$ ). In addition, we also compared the quality of the SE estimates to the ‘‘benchmark’’ standard error estimates obtained from the full data set by computing DSEfull, which represented the differences between the average standard error estimates (across Monte Carlo runs) for the full data set against those obtained from using (1) imputed data based on the full MI approach, (2) imputed data based on the partial MI approach, and (3) listwise deleted. Finally, we also calculated power and coverage rates. Power was defined as the proportion of 95% confidence



intervals that did not contain 0 across the Monte Carlo replications. Coverage rates were defined as the percentages of replications whose 95% confidence intervals for the parameters included the true parameter values. Simulation results with power close to 1.0 and coverage rates close to the nominal rate of .95 would be considered as ideal.

## 2.2 Simulation Results

Key results from the simulations are summarized in Table 2.1 through Table 2.6. Key performance measures with each missing data technique, including listwise deletion, full MI and partial MI were compared against each other as well as full data estimation results under the six simulation conditions.

As expected, compared with the listwise deletion method, the partial as well as full MI approaches improved the accuracy of the parameters' point estimates under MAR and NMAR conditions, but not much for MCAR condition. It can be seen from Table 2.1 to Table 2.6 that biases and RMSE were smaller with both of the MI approaches for most of the parameters under MAR and NMAR conditions. Under MCAR and MAR conditions, even though the point estimates for the covariate parameters were less biased with listwise deletion data than the imputed data (see Figure 2.2), the point estimates of the dynamic parameters (see Figure 2.1), including auto-regressive and cross-regressive parameters were noticeably biased, because listwise deletion altered the spacing between time points and led to biased estimates in the dynamic parameters. With partial MI approach, the missing covariates were filled in with imputed data, and the original spacing between the data points was preserved. Even though some of the dependent variables were still missing, FIML method was able to provide rather accurate point estimates with such missingness under MCAR and MAR conditions.

Among the dynamic parameters, the point estimates of the auto-regressive parameters improved more than the cross-regressive parameters under both MI approaches. With MI, either partial MI or full MI, under the MAR and NMAR conditions, but not the MCAR conditions, the RMSEs of auto-regressive parameters decreased, as compared with the RMSEs of the estimation of auto-regressive parameters with the listwise deletion data. In particular, under the MAR condition, the RMSEs of the point estimates of the auto-regressive parameters were almost the same as the RMSEs of auto-regressive parameter estimates when fitting the model to the full data.

Under MCAR condition, however, when the full MI was applied, biases and RMSEs were reduced for some of the parameters but not others. It can be seen from Figure 2.1 and Figure 2.2 that the average RMSEs for the four dynamic parameters with full MI were very close to that obtained with the listwise deletion method. The reason why full MI did not improve the point estimates might be that under the MCAR condition, the missingness was completely at random and did not depend on any other information, observed or not. Therefore, imputing missing values based on other observed variables provided only spurious information. Also, to mimic the real-life scenario

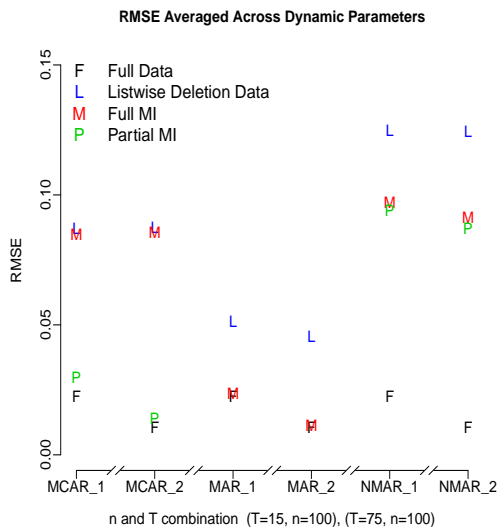


Figure 2.1: A comparison of RMSE for dynamic parameter estimates

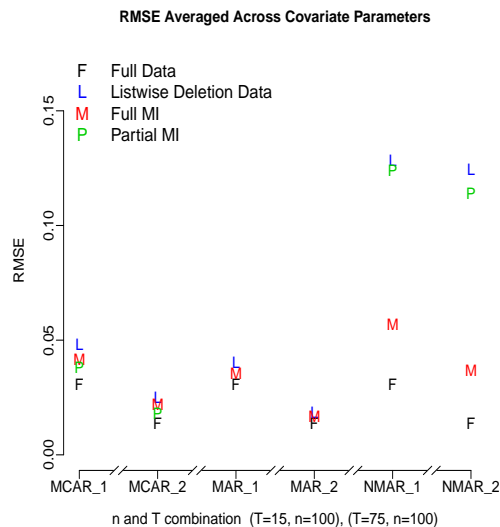


Figure 2.2: A comparison of RMSE for covariate parameter estimates

where the true missingness model was never known, we imputed the missing values with a general imputation model, which was misspecified compared with the data generation model. Thus, under MCAR, the full MI approach actually introduced more sources of model misspecification into the model estimation process than the partial MI approach, and the inclusion of such spurious information also did not lead to any advantages in estimating the dynamic parameters when compared to the listwise deletion method. Under the MAR and NMAR conditions, missingness either depended on observed or unobserved information, therefore imputing all missing variables based on all information in the non-missing data did help improve the accuracy of the point estimates in most cases, with the full MI approach producing biases and RMSEs that were either comparable to, or lower than those from the partial MI approach; in some cases (e.g., the dynamic parameters under MAR conditions), the RMSEs were even close to those obtained from using the full data.

Results were more complex with the standard error estimates. With listwise deletion, standard error estimates were more susceptible to changes in the missing data generation conditions. Figure 2.3 shows that under MCAR and MAR conditions, DSEs with listwise deletion method were very small, indicating the average standard error estimates of the parameters across Monte Carlo runs were very close to the empirical standard deviations of the parameters. However, as shown in Figure 2.4, positive biases in SE estimates compared to those from the full data set were observed with listwise deletion data under all MCAR and MAR conditions, especially in the smaller  $T$  condition, suggests inflation of standard error estimates when compared to the estimation results using the full data set. A possible reason might be that listwise deletion resulted in smaller sample size, which in turn led to more uncertainty in the estimates, thus leading to “over-estimation” of the SEs compared to variability that existed in the full data. The extent of over-estimation was reduced

with both partial MI and full MI. We can see from the Table 2.1 through Table 2.4 that DSEfull is smaller with imputed data for both MCAR and MAR conditions, meaning that the standard error estimations obtained when fitting model to MI data deviate less to that obtained when fitting model to full data.

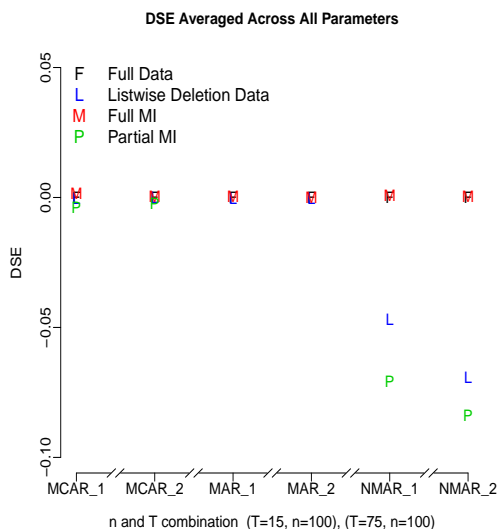


Figure 2.3: A comparison of DSE averaged across all parameter estimates

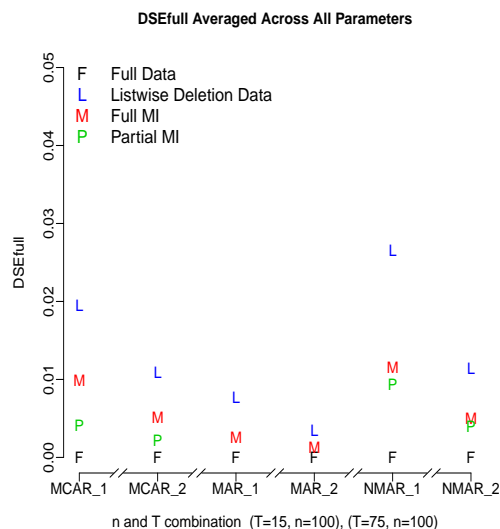


Figure 2.4: A comparison of DSEfull averaged across all parameter estimates

As distinct from MCAR and MAR conditions, fitting model with listwise deletion data led to under-estimation of the standard errors under the NMAR conditions, resulting in negative DSEs, see Figure 2.3. This is because we used FIML method in the estimation of the parameters, assuming all missingness was MAR. This method does not take nonignorable missingness into account and under-estimates the standard error of the parameters. Such under-estimation was corrected when all missing values were imputed with the full MI method. It can be seen from Table 2.5 to Table 2.6 that the DSEs of full MI data under NMAR conditions were very close to zero, especially with larger number of time points. However, if only covariates were imputed with partial MI, as shown in Figure 2.3, the magnitude of DSEs increased as compared with listwise deletion data, indicating larger underestimation of standard errors. Again, this was probably due to the fact that the dynamic model with missing dependent variables were estimated with FIML, which does not handle NMAR missingness. Nevertheless, DSEfull still decreased after using partial MI, see Figure 2.4, suggesting the standard error estimates with partial MI data were still relatively close to the standard error estimates obtained with full data.

Even though compared with partial MI method, the full MI approach produced better standard error estimates especially under NMAR conditions, better coverage rates were obtained with partial than full MI in general. Figure 2.5 shows that under MCAR condition, full MI hardly improved

the coverage of the dynamic parameters. But with partial MI, coverage of almost all parameters were close to .95. Under MAR condition, the coverage of parameter estimates for covariates were satisfactory under listwise deletion, but very low for dynamic parameters (see Figure 2.5 and 2.6). Performing MI procedures helped with the coverage of dynamic parameter estimates as compared with listwise deletion. After imputation, the coverage rates of most parameter estimates were close to the nominal rate of .95. Specifically, the partial MI approach also improved the coverage rates of the dynamic parameters under NMAR conditions, especially for the dynamic parameters in comparison to both the listwise deletion and the full MI methods. When the full MI method was used, coverage of the parameters – particularly the dynamic parameters but also to some extent, the covariate-related parameters – remained low and close to the coverage rates from using the listwise deletion method. Thus, despite some advantages of the full MI approach in providing more accurate standard error estimates than the partial MI approach, the latter may actually be preferred over the full MI approach as a general missing data handling approach based on the coverage and point estimate properties of the parameters across all missing data conditions. Of course, these results are restricted to the situations evaluated in the simulation study - that is, the true dynamic model is known and correctly specified, even though the missing data mechanisms may range from MCAR to NMAR. However, these simulation results helped highlight some lesser known advantages of the partial MI approach compared to other full MI approaches.

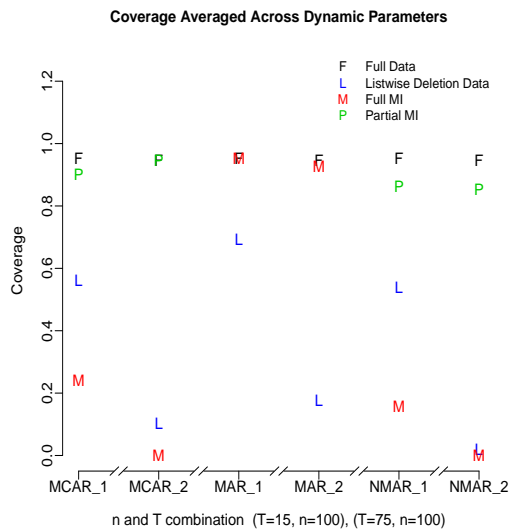


Figure 2.5: A comparison of coverage for dynamic parameter estimates

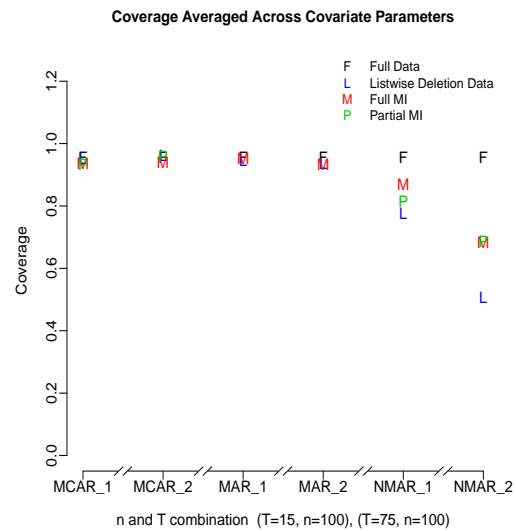


Figure 2.6: A comparison of coverage for covariate parameter estimates

With regard to power, for the sample sizes consider in the current simulation, the power estimates based on the 95% confidence level were well above .8 for all conditions and using any of the three approaches considered (partial MI, full MI and listwise deletion methods). However, we shall be aware that with the two MI approaches, the power for all parameter estimates were very close to 100% across all conditions. The advantages offered by the MI approaches in terms of power would

likely be even more pronounced in conditions with smaller sample sizes and/or larger percentages of missingness, given the expected decrements in performance of the listwise deletion method under such scenarios.

Table 2.1: Summary Statistics of MCAR condition across 500 MC Replications ( $N = 100, T = 15$ )

Parameter	Full data						Listwise deletion						
	Bias	RMSE	SD	DSE	Power	Coverage	Bias	RMSE	SD	DSE	DSEfull	Power	Coverage
$a_{w \rightarrow w}$	-0.003	0.021	0.021	0.001	1.000	0.962	-0.112	0.119	0.040	0.000	0.018	1.000	0.184
$b_{h \rightarrow w}$	-0.001	0.024	0.024	0.000	1.000	0.938	0.064	0.078	0.045	-0.001	0.020	1.000	0.680
$b_{w \rightarrow h}$	0.000	0.021	0.021	0.001	1.000	0.956	0.041	0.057	0.039	0.000	0.017	0.984	0.838
$a_{h \rightarrow h}$	0.000	0.024	0.024	0.000	1.000	0.954	-0.083	0.094	0.043	-0.001	0.019	0.998	0.540
$c_{ca \rightarrow w}$	0.002	0.038	0.038	0.000	1.000	0.954	0.014	0.058	0.056	0.002	0.021	1.000	0.954
$d_{cn \rightarrow w}$	-0.002	0.025	0.025	0.000	1.000	0.950	-0.004	0.039	0.039	0.001	0.015	1.000	0.960
$c_{ca \rightarrow h}$	0.002	0.036	0.036	0.001	1.000	0.964	0.013	0.058	0.056	0.001	0.020	1.000	0.942
$d_{cn \rightarrow h}$	-0.001	0.024	0.025	0.000	1.000	0.952	0.001	0.037	0.037	0.001	0.014	1.000	0.950
$\sigma_{e_w}^2$	0.016	0.041	0.038	-0.001	1.000	0.932	0.140	0.156	0.069	-0.005	0.027	1.000	0.424
$\sigma_{e_h}^2$	0.006	0.038	0.037	0.000	1.000	0.938	0.066	0.089	0.060	0.000	0.022	1.000	0.830
$\sigma_{e_{wh}}$	-0.013	0.031	0.029	-0.002	1.000	0.910	-0.092	0.105	0.051	-0.002	0.021	0.994	0.476

Parameter	Full Multiple imputation						Partial Multiple imputation							
	Bias	RMSE	SD	DSE	DSEfull	Power	Coverage	Bias	RMSE	SD	DSE	DSEfull	Power	Coverage
$a_{w \rightarrow w}$	-0.101	0.104	0.025	0.005	0.008	1.000	0.044	-0.011	0.030	0.028	-0.003	0.003	1.000	0.884
$b_{h \rightarrow w}$	0.089	0.093	0.026	0.006	0.008	1.000	0.156	0.008	0.032	0.031	-0.004	0.003	1.000	0.912
$b_{w \rightarrow h}$	0.055	0.060	0.023	0.006	0.006	1.000	0.502	0.004	0.027	0.027	-0.003	0.002	1.000	0.912
$a_{h \rightarrow h}$	-0.079	0.083	0.025	0.006	0.008	1.000	0.260	-0.004	0.031	0.031	-0.005	0.002	1.000	0.898
$c_{ca \rightarrow w}$	0.002	0.049	0.049	0.000	0.011	1.000	0.932	0.003	0.045	0.045	-0.001	0.006	1.000	0.944
$d_{cn \rightarrow w}$	-0.011	0.035	0.034	0.001	0.009	1.000	0.932	-0.001	0.032	0.032	-0.002	0.005	1.000	0.926
$c_{ca \rightarrow h}$	0.007	0.049	0.049	-0.001	0.011	1.000	0.942	0.003	0.044	0.044	-0.002	0.005	1.000	0.942
$d_{cn \rightarrow h}$	-0.006	0.033	0.033	0.001	0.009	1.000	0.946	0.001	0.032	0.032	-0.003	0.004	1.000	0.920
$\sigma_{e_w}^2$	0.124	0.135	0.055	-0.003	0.015	1.000	0.348	0.026	0.057	0.051	-0.007	0.007	1.000	0.894
$\sigma_{e_h}^2$	0.060	0.078	0.049	-0.002	0.011	1.000	0.776	0.008	0.047	0.046	-0.005	0.004	1.000	0.924
$\sigma_{e_{wh}}$	-0.101	0.109	0.041	-0.001	0.013	0.998	0.292	-0.028	0.048	0.038	-0.008	0.004	1.000	0.776

Bias =  $\frac{1}{H} \sum_h^H (\hat{\theta}_h - true\theta)$ , where  $H$  = total number of MC replications,  $\hat{\theta}_h$  = estimate of  $\theta$  from the  $h$ th MC replications, true  $\theta$  = true value of a parameter; RMSE =  $\sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - true\theta)^2}$ ; SD = standard deviation of  $\hat{\theta}$  across Monte Carlo runs; DSE = deviance of average standard error estimate across Monte Carlo runs from SD; DSEfull = difference between average standard error estimate with missing data under each method and average standard error estimate with full data; Power = 1 - the proportion of 95% confidence intervals (CIs) that contain 0 across the Monte Carlo replications; Coverage = the proportion of 95% confidence intervals (CIs) that contain true  $\theta$  across the Monte Carlo replications.

Table 2.2: Summary Statistics of MCAR condition across 500 MC Replications ( $N = 100, T = 75$ )

Parameter	Full data						Listwise deletion						
	Bias	RMSE	SD	DSE	Power	Coverage	Bias	RMSE	SD	DSE	DSEfull	Power	Coverage
$a_{w \rightarrow w}$	0.000	0.010	0.010	0.000	1.000	0.946	-0.121	0.122	0.020	-0.001	0.009	1.000	0.000
$b_{h \rightarrow w}$	0.000	0.011	0.011	0.000	1.000	0.940	0.073	0.076	0.021	0.000	0.011	1.000	0.102
$b_{w \rightarrow h}$	0.001	0.010	0.010	0.000	1.000	0.942	0.050	0.053	0.018	0.001	0.008	1.000	0.306
$a_{h \rightarrow h}$	0.000	0.011	0.011	0.000	1.000	0.952	-0.097	0.099	0.021	-0.001	0.010	1.000	0.008
$c_{ca \rightarrow w}$	0.000	0.016	0.016	0.001	1.000	0.960	0.014	0.031	0.027	0.001	0.012	1.000	0.950
$d_{cn \rightarrow w}$	0.000	0.011	0.011	0.000	1.000	0.952	0.001	0.019	0.019	0.001	0.009	1.000	0.970
$c_{ca \rightarrow h}$	0.001	0.017	0.017	0.000	1.000	0.952	0.016	0.031	0.027	0.001	0.011	1.000	0.960
$d_{cn \rightarrow h}$	0.001	0.011	0.011	0.000	1.000	0.964	0.000	0.019	0.019	0.000	0.008	1.000	0.976
$\sigma_{e_w}^2$	0.003	0.017	0.016	0.000	1.000	0.944	0.143	0.147	0.033	-0.001	0.016	1.000	0.006
$\sigma_{e_h}^2$	0.001	0.016	0.016	0.000	1.000	0.954	0.071	0.077	0.030	0.000	0.013	1.000	0.384
$\sigma_{e_{wh}}$	-0.003	0.013	0.012	0.000	1.000	0.946	-0.091	0.094	0.024	0.001	0.013	1.000	0.068

Parameter	Full Multiple imputation						Partial Multiple imputation							
	Bias	RMSE	SD	DSE	DSEfull	Power	Coverage	Bias	RMSE	SD	DSE	DSEfull	Power	Coverage
$a_{w \rightarrow w}$	-0.103	0.104	0.011	0.002	0.003	1.000	0.000	-0.003	0.013	0.013	-0.002	0.001	1.000	0.962
$b_{h \rightarrow w}$	0.092	0.093	0.012	0.002	0.004	1.000	0.000	0.003	0.015	0.014	-0.003	0.002	1.000	0.942
$b_{w \rightarrow h}$	0.060	0.061	0.011	0.002	0.003	1.000	0.002	0.003	0.013	0.013	-0.003	0.000	1.000	0.946
$a_{h \rightarrow h}$	-0.084	0.085	0.012	0.002	0.004	1.000	0.000	-0.003	0.015	0.014	-0.003	0.001	1.000	0.934
$c_{ca \rightarrow w}$	0.005	0.024	0.023	-0.001	0.006	1.000	0.962	0.004	0.021	0.021	-0.001	0.003	1.000	0.956
$d_{cn \rightarrow w}$	-0.012	0.020	0.016	0.000	0.005	1.000	0.928	0.001	0.015	0.015	-0.001	0.003	1.000	0.954
$c_{ca \rightarrow h}$	0.009	0.025	0.023	-0.001	0.005	1.000	0.946	0.004	0.022	0.021	-0.002	0.002	1.000	0.956
$d_{cn \rightarrow h}$	-0.011	0.019	0.016	0.000	0.005	1.000	0.922	-0.001	0.015	0.015	-0.001	0.003	1.000	0.968
$\sigma_{e_w}^2$	0.122	0.124	0.025	-0.001	0.008	1.000	0.000	0.006	0.023	0.022	-0.002	0.004	1.000	0.942
$\sigma_{e_h}^2$	0.058	0.062	0.022	0.000	0.006	1.000	0.322	0.001	0.021	0.021	-0.002	0.003	1.000	0.956
$\sigma_{e_{wh}}$	-0.103	0.105	0.019	0.000	0.007	1.000	0.004	-0.009	0.019	0.017	-0.003	0.002	1.000	0.922

Bias= $\frac{1}{H} \sum_h^H (\hat{\theta}_h - true\theta)$ , where  $H$ =total number of MC replications,  $\hat{\theta}_h$  =estimate of  $\theta$  from the  $h$ th MC replications, true  $\theta$  = true value of a parameter; RMSE= $\sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - true\theta)^2}$ ; SD=standard deviation of  $\hat{\theta}$  across Monte Carlo runs; DSE = deviance of average standard error estimate across Monte Carlo runs from SD; DSEfull = difference between average standard error estimate with missing data under each method and average standard error estimate with full data; Power = 1 - the proportion of 95% confidence intervals (CIs) that contain 0 across the Monte Carlo replications; Coverage = the proportion of 95% confidence intervals (CIs) that contain true  $\theta$  across the Monte Carlo replications.

Table 2.3: Summary Statistics of MAR condition across 500 MC Replications ( $N = 100, T = 15$ )

Parameter	Full data						Listwise deletion						Multiple imputation							
	Bias	RMSE	SD	DSE	Power	Coverage	Bias	RMSE	SD	DSE	DSEfull	Power	Coverage	Bias	RMSE	SD	DSE	DSEfull	Power	Coverage
$a_{w \rightarrow w}$	-0.003	0.021	0.021	0.001	1.000	0.962	-0.062	0.068	0.029	-0.001	0.006	1.000	0.424	-0.011	0.024	0.021	0.002	0.001	1.000	0.960
$b_{h \rightarrow w}$	-0.001	0.024	0.024	0.000	1.000	0.938	0.034	0.046	0.030	0.000	0.006	1.000	0.806	0.004	0.025	0.025	0.001	0.001	1.000	0.950
$b_{w \rightarrow h}$	0.000	0.021	0.021	0.001	1.000	0.956	0.021	0.034	0.026	0.001	0.005	1.000	0.898	-0.007	0.021	0.020	0.002	0.001	1.000	0.956
$a_{h \rightarrow h}$	0.000	0.024	0.024	0.000	1.000	0.954	-0.048	0.057	0.030	0.000	0.006	1.000	0.644	0.004	0.024	0.024	0.000	0.001	1.000	0.948
$c_{ca \rightarrow w}$	0.002	0.038	0.038	0.000	1.000	0.954	0.011	0.052	0.051	0.000	0.012	1.000	0.944	0.002	0.044	0.044	-0.001	0.005	1.000	0.952
$d_{cn \rightarrow w}$	-0.002	0.025	0.025	0.000	1.000	0.950	-0.002	0.032	0.032	-0.001	0.006	1.000	0.942	0.007	0.030	0.029	0.000	0.004	1.000	0.950
$c_{ca \rightarrow h}$	0.002	0.036	0.036	0.001	1.000	0.964	0.010	0.049	0.049	0.001	0.012	1.000	0.936	0.003	0.040	0.040	0.002	0.005	1.000	0.960
$d_{cn \rightarrow h}$	-0.001	0.024	0.025	0.000	1.000	0.952	-0.002	0.029	0.029	0.001	0.005	1.000	0.970	-0.006	0.028	0.027	0.001	0.003	1.000	0.952
$\sigma_{\epsilon_w}^2$	0.016	0.041	0.038	-0.001	1.000	0.932	0.089	0.100	0.047	0.000	0.011	1.000	0.548	0.027	0.049	0.041	0.000	0.003	1.000	0.908
$\sigma_{\epsilon_h}^2$	0.006	0.038	0.037	0.000	1.000	0.938	0.042	0.063	0.047	-0.002	0.009	1.000	0.852	0.002	0.039	0.039	0.000	0.002	1.000	0.954
$\sigma_{\epsilon_{wh}}$	-0.013	0.031	0.029	-0.002	1.000	0.910	-0.058	0.068	0.035	-0.001	0.007	1.000	0.578	-0.011	0.032	0.031	-0.001	0.002	1.000	0.914

Bias =  $\frac{1}{H} \sum_h^H (\hat{\theta}_h - true\theta)$ , where  $H$  = total number of MC replications,  $\hat{\theta}_h$  = estimate of  $\theta$  from the  $h$ th MC replications, true  $\theta$  = true value of a parameter; RMSE =  $\sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - true\theta)^2}$ ; SD = standard deviation of  $\hat{\theta}$  across Monte Carlo runs; DSE = deviance of average standard error estimate across Monte Carlo runs from SD; DSEfull = difference between average standard error estimate with missing data under each method and average standard error estimate with full data; Power = 1 - the proportion of 95% confidence intervals (CIs) that contain 0 across the Monte Carlo replications; Coverage = the proportion of 95% confidence intervals (CIs) that contain true  $\theta$  across the Monte Carlo replications.

Table 2.4: Summary Statistics of MAR condition across 500 MC Replications ( $N = 100, T = 75$ )

Parameter	Full data						Listwise deletion						Multiple imputation							
	Bias	RMSE	SD	DSE	Power	Coverage	Bias	RMSE	SD	DSE	DSEfull	Power	Coverage	Bias	RMSE	SD	DSE	DSEfull	Power	Coverage
$a_{w \rightarrow w}$	0.000	0.010	0.010	0.000	1.000	0.946	-0.063	0.065	0.012	0.000	0.002	1.000	0.002	-0.008	0.013	0.010	0.000	0.000	1.000	0.882
$b_{h \rightarrow w}$	0.000	0.011	0.011	0.000	1.000	0.940	0.036	0.038	0.013	0.000	0.003	1.000	0.200	0.004	0.011	0.010	0.001	0.001	1.000	0.960
$b_{w \rightarrow h}$	0.001	0.010	0.010	0.000	1.000	0.942	0.025	0.027	0.012	-0.001	0.002	1.000	0.476	-0.006	0.011	0.010	0.000	0.000	1.000	0.926
$a_{h \rightarrow h}$	0.000	0.011	0.011	0.000	1.000	0.952	-0.052	0.053	0.013	-0.001	0.003	1.000	0.034	0.004	0.011	0.010	0.000	0.001	1.000	0.942
$c_{ca \rightarrow w}$	0.000	0.016	0.016	0.001	1.000	0.960	0.009	0.023	0.021	0.001	0.005	1.000	0.934	0.002	0.019	0.019	0.000	0.002	1.000	0.952
$d_{cn \rightarrow w}$	0.000	0.011	0.011	0.000	1.000	0.952	0.000	0.014	0.014	0.000	0.003	1.000	0.940	0.010	0.016	0.013	0.000	0.002	1.000	0.886
$c_{ca \rightarrow h}$	0.001	0.017	0.017	0.000	1.000	0.952	0.011	0.025	0.022	0.000	0.005	1.000	0.916	0.004	0.020	0.019	-0.001	0.002	1.000	0.942
$d_{cn \rightarrow h}$	0.001	0.011	0.011	0.000	1.000	0.964	0.001	0.013	0.013	0.000	0.003	1.000	0.954	-0.004	0.013	0.012	0.001	0.002	1.000	0.958
$\sigma_{\epsilon_w}^2$	0.003	0.017	0.016	0.000	1.000	0.944	0.081	0.084	0.021	0.000	0.005	1.000	0.030	0.014	0.022	0.018	0.000	0.002	1.000	0.884
$\sigma_{\epsilon_h}^2$	0.001	0.016	0.016	0.000	1.000	0.954	0.041	0.045	0.020	0.000	0.004	1.000	0.524	-0.004	0.017	0.017	0.000	0.001	1.000	0.956
$\sigma_{\epsilon_{wh}}$	-0.003	0.013	0.012	0.000	1.000	0.946	-0.050	0.052	0.016	-0.001	0.003	1.000	0.104	-0.001	0.014	0.013	0.000	0.001	1.000	0.942

Bias =  $\frac{1}{H} \sum_h^H (\hat{\theta}_h - true\theta)$ , where  $H$  = total number of MC replications,  $\hat{\theta}_h$  = estimate of  $\theta$  from the  $h$ th MC replications, true  $\theta$  = true value of a parameter; RMSE =  $\sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - true\theta)^2}$ ; SD = standard deviation of  $\hat{\theta}$  across Monte Carlo runs; DSE = deviance of average standard error estimate across Monte Carlo runs from SD; DSEfull = difference between average standard error estimate with missing data under each method and average standard error estimate with full data; Power = 1 - the proportion of 95% confidence intervals (CIs) that contain 0 across the Monte Carlo replications; Coverage = the proportion of 95% confidence intervals (CIs) that contain true  $\theta$  across the Monte Carlo replications.



Table 2.5: Summary Statistics of NMAR condition across 500 MC Replications ( $N = 100, T = 15$ )

Parameter	Full data						Listwise deletion						
	Bias	RMSE	SD	DSE	Power	Coverage	Bias	RMSE	SD	DSE	DSEfull	Power	Coverage
$a_{w \rightarrow w}$	-0.003	0.021	0.021	0.001	1.000	0.962	-0.145	0.165	0.079	-0.032	0.025	0.938	0.234
$b_{h \rightarrow w}$	-0.001	0.024	0.024	0.000	1.000	0.938	0.096	0.121	0.073	-0.020	0.029	0.906	0.644
$b_{w \rightarrow h}$	0.000	0.021	0.021	0.001	1.000	0.956	0.069	0.088	0.054	-0.008	0.024	0.808	0.716
$a_{h \rightarrow h}$	0.000	0.024	0.024	0.000	1.000	0.954	-0.104	0.125	0.069	-0.019	0.027	0.914	0.560
$c_{ca \rightarrow w}$	0.002	0.038	0.038	0.000	1.000	0.954	0.062	0.129	0.113	-0.046	0.029	0.938	0.722
$d_{cn \rightarrow w}$	-0.002	0.025	0.025	0.000	1.000	0.950	0.031	0.130	0.126	-0.081	0.021	0.938	0.904
$c_{ca \rightarrow h}$	0.002	0.036	0.036	0.001	1.000	0.964	0.093	0.148	0.116	-0.051	0.028	0.938	0.576
$d_{cn \rightarrow h}$	-0.001	0.024	0.025	0.000	1.000	0.952	0.028	0.107	0.103	-0.059	0.019	0.938	0.902
$\sigma_{e_w}^2$	0.016	0.041	0.038	-0.001	1.000	0.932	0.100	0.200	0.174	-0.104	0.033	1.000	0.520
$\sigma_{e_h}^2$	0.006	0.038	0.037	0.000	1.000	0.938	0.008	0.150	0.150	-0.085	0.028	1.000	0.876
$\sigma_{e_{wh}}$	-0.013	0.031	0.029	-0.002	1.000	0.910	-0.111	0.129	0.065	-0.010	0.029	0.934	0.562

Parameter	Full Multiple imputation						Partial Multiple imputation							
	Bias	RMSE	SD	DSE	DSEfull	Power	Coverage	Bias	RMSE	SD	DSE	DSEfull	Power	Coverage
$a_{w \rightarrow w}$	-0.114	0.117	0.026	0.004	0.008	1.000	0.022	-0.044	0.123	0.115	-0.083	0.010	0.906	0.864
$b_{h \rightarrow w}$	0.101	0.104	0.028	0.006	0.009	1.000	0.130	0.032	0.094	0.089	-0.054	0.011	0.906	0.874
$b_{w \rightarrow h}$	0.070	0.074	0.024	0.005	0.007	0.998	0.304	0.027	0.065	0.059	-0.029	0.008	0.906	0.860
$a_{h \rightarrow h}$	-0.089	0.093	0.027	0.004	0.008	1.000	0.174	-0.031	0.094	0.089	-0.056	0.009	0.906	0.858
$c_{ca \rightarrow w}$	0.032	0.066	0.057	-0.005	0.014	1.000	0.878	-0.003	0.105	0.105	-0.057	0.010	0.906	0.826
$d_{cn \rightarrow w}$	-0.010	0.038	0.036	0.000	0.011	1.000	0.936	0.049	0.154	0.146	-0.112	0.009	0.906	0.866
$c_{ca \rightarrow h}$	0.070	0.087	0.051	0.000	0.014	1.000	0.712	0.023	0.113	0.111	-0.064	0.010	0.906	0.706
$d_{cn \rightarrow h}$	-0.004	0.036	0.036	0.000	0.011	1.000	0.950	0.041	0.124	0.117	-0.084	0.009	0.906	0.858
$\sigma_{e_w}^2$	0.122	0.134	0.055	0.000	0.017	1.000	0.380	-0.039	0.161	0.156	-0.109	0.010	1.000	0.864
$\sigma_{e_h}^2$	0.032	0.061	0.051	-0.003	0.012	1.000	0.896	-0.069	0.162	0.146	-0.103	0.007	1.000	0.824
$\sigma_{e_{wh}}$	-0.121	0.128	0.043	0.000	0.016	0.984	0.232	-0.039	0.076	0.065	-0.028	0.010	1.000	0.820

Bias =  $\frac{1}{H} \sum_h^H (\hat{\theta}_h - true\theta)$ , where  $H$  = total number of MC replications,  $\hat{\theta}_h$  = estimate of  $\theta$  from the  $h$ th MC replications,  $true\theta$  = true value of a parameter; RMSE =  $\sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - true\theta)^2}$ ; SD = standard deviation of  $\hat{\theta}$  across Monte Carlo runs; DSE = deviance of average standard error estimate across Monte Carlo runs from SD; DSEfull = difference between average standard error estimate with missing data under each method and average standard error estimate with full data; Power = 1 - the proportion of 95% confidence intervals (CIs) that contain 0 across the Monte Carlo replications; Coverage = the proportion of 95% confidence intervals (CIs) that contain true  $\theta$  across the Monte Carlo replications.

Table 2.6: Summary Statistics of NMAR condition across 500 MC Replications ( $N = 100, T = 75$ )

Parameter	Full data						Listwise deletion						
	Bias	RMSE	SD	DSE	Power	Coverage	Bias	RMSE	SD	DSE	DSEfull	Power	Coverage
$a_{w \rightarrow w}$	0.000	0.010	0.010	0.000	1.000	0.946	-0.152	0.168	0.071	-0.052	0.010	0.924	0.000
$b_{h \rightarrow w}$	0.000	0.011	0.011	0.000	1.000	0.940	0.106	0.120	0.057	-0.035	0.012	0.924	0.008
$b_{w \rightarrow h}$	0.001	0.010	0.010	0.000	1.000	0.942	0.075	0.084	0.038	-0.019	0.009	0.924	0.064
$a_{h \rightarrow h}$	0.000	0.011	0.011	0.000	1.000	0.952	-0.113	0.126	0.054	-0.033	0.011	0.924	0.002
$c_{ca \rightarrow w}$	0.000	0.016	0.016	0.001	1.000	0.960	0.050	0.113	0.102	-0.072	0.013	0.924	0.250
$d_{cn \rightarrow w}$	0.000	0.011	0.011	0.000	1.000	0.952	0.038	0.137	0.131	-0.111	0.009	0.924	0.878
$c_{ca \rightarrow h}$	0.001	0.017	0.017	0.000	1.000	0.952	0.084	0.139	0.111	-0.082	0.012	0.924	0.016
$d_{cn \rightarrow h}$	0.001	0.011	0.011	0.000	1.000	0.964	0.034	0.109	0.104	-0.084	0.009	0.924	0.886
$\sigma_{e_w}^2$	0.003	0.017	0.016	0.000	1.000	0.944	0.099	0.201	0.175	-0.144	0.015	1.000	0.004
$\sigma_{e_h}^2$	0.001	0.016	0.016	0.000	1.000	0.954	0.008	0.149	0.149	-0.120	0.013	1.000	0.606
$\sigma_{e_{wh}}$	-0.003	0.013	0.012	0.000	1.000	0.946	-0.109	0.114	0.036	-0.011	0.013	1.000	0.018

Parameter	Full Multiple imputation						Partial Multiple imputation							
	Bias	RMSE	SD	DSE	DSEfull	Power	Coverage	Bias	RMSE	SD	DSE	DSEfull	Power	Coverage
$a_{w \rightarrow w}$	-0.110	0.110	0.011	0.002	0.003	1.000	0.000	-0.038	0.116	0.110	-0.097	0.003	0.912	0.854
$b_{h \rightarrow w}$	0.099	0.099	0.011	0.003	0.004	1.000	0.000	0.029	0.087	0.082	-0.067	0.005	0.912	0.880
$b_{w \rightarrow h}$	0.068	0.069	0.011	0.001	0.002	1.000	0.000	0.025	0.058	0.053	-0.040	0.003	0.912	0.816
$a_{h \rightarrow h}$	-0.087	0.088	0.012	0.002	0.004	1.000	0.000	-0.028	0.087	0.083	-0.069	0.004	0.912	0.868
$c_{ca \rightarrow w}$	0.028	0.038	0.025	-0.002	0.006	1.000	0.766	-0.005	0.091	0.091	-0.070	0.004	0.912	0.754
$d_{cn \rightarrow w}$	-0.011	0.020	0.016	0.000	0.005	1.000	0.890	0.045	0.146	0.139	-0.124	0.004	0.912	0.864
$c_{ca \rightarrow h}$	0.070	0.074	0.023	-0.001	0.005	1.000	0.134	0.024	0.102	0.100	-0.079	0.004	0.912	0.262
$d_{cn \rightarrow h}$	-0.003	0.016	0.015	0.001	0.005	1.000	0.944	0.040	0.117	0.110	-0.095	0.004	0.912	0.864
$\sigma_{e_w}^2$	0.117	0.120	0.025	-0.001	0.008	1.000	0.002	-0.043	0.150	0.144	-0.123	0.005	1.000	0.860
$\sigma_{e_h}^2$	0.029	0.036	0.022	0.000	0.006	1.000	0.734	-0.069	0.152	0.136	-0.116	0.004	1.000	0.686
$\sigma_{e_{wh}}$	-0.114	0.116	0.019	0.000	0.007	1.000	0.000	-0.028	0.063	0.056	-0.040	0.004	1.000	0.818

Bias= $\frac{1}{H} \sum_h^H (\hat{\theta}_h - true\theta)$ , where  $H$ =total number of MC replications,  $\hat{\theta}_h$  =estimate of  $\theta$  from the  $h$ th MC replications, true  $\theta$  = true value of a parameter; RMSE= $\sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{\theta}_h - true\theta)^2}$ ; SD=standard deviation of  $\hat{\theta}$  across Monte Carlo runs; DSE = deviance of average standard error estimate across Monte Carlo runs from SD; DSEfull = difference between average standard error estimate with missing data under each method and average standard error estimate with full data; Power = 1 - the proportion of 95% confidence intervals (CIs) that contain 0 across the Monte Carlo replications; Coverage = the proportion of 95% confidence intervals (CIs) that contain true  $\theta$  across the Monte Carlo replications.

# Chapter 3 |

## Empirical Illustration

We used data collected for the Schermerhorn et al. (2010) study on the dynamic of interparental emotion states after conflicts with child emotions and behavior as covariates for empirical illustration of the missing data handling methods considered in the simulation study. The goal of this illustration was to compare results based on the three different missing data methods to the method used in the published paper, wherein the covariates were coded as either 0 or 1 with both non-presents of the behavior and unreported behavior coded as 0. The results from using the full and partial MI approaches were very close since only 0.4% of the dependent variables were missing. Therefore, only results from the full MI approach were reported. The construct of interest in this illustration was conflict resolution, which was reported by both husband and wife after each conflict. Two child variables, child agentic behavior and child negative emotions, were included in the model as two covariates.

Under the full MI method, 10 imputations were obtained with the `mice()` function in MICE package in R. Even though the default number of imputations was 5, we generated 10 imputations because the percentages of missingness in the covariates were quite large. All variables we considered as relevant were included as predictors in the MI model, including husband's and wife's negative emotion after each conflict and after each previous conflict, husband's and wife's positive emotion after each conflict and after each previous conflict, husband's and wife's conflict resolution score after each conflict and after each previous conflict, child agentic behavior, child negative emotion, husband's and wife's reported time length of the interaction, who initiated the problem, whether the problem is old or new, husband's and wife's reported hostility in marital relations, husband's and wife's self-reported Symptom Checklist-90 depression score, husband's and wife's self-reported depressive symptom score, husband's and wife's reported marital satisfaction score, and child age. Among these variables, whether the problem was new or old and whether husband or wife initiated the problem were dichotomous variables. The default imputation methods in the MICE package,

namely, predictive mean matching method for numerical data and logistic regression method for binary variables, were used to generate the imputed data sets.

We used `mkfm6` to fit the state space model (Equation 3.1) with each of the 10 imputed data sets as if they were fully observed data. Ten sets of parameter estimates were obtained and pooled parameter estimates and parameter standard error estimates were calculated using R according to the method described in Section 3.2.3. The same state-space model and estimation algorithm was used for the data with covariates re-coded as in the published study, as well as using the data with missing-covariate-cases dropped (listwise deletion). Parameter estimation results are shown in Table 3.1.

As we can see from Table 3.1, the parameter estimates using the data re-coding procedure in the published study and the full MI approach were very close, while listwise deletion method seemed to produce biased estimates. For the dynamic parameters, the auto-regressive parameter for wives' conflict resolution was observed to be statistically significant with both the original re-coding method and MI, but not for listwise deletion data. The significant negative auto-regressive parameter indicates that wife's conflict resolution score was influenced by her conflict resolution score at the end of the immediately preceding conflict, but in the opposite direction. That is to say, when wife's conflict resolution score was high at the end of one conflict, her conflict resolution score would tend to be low at the end of the next conflict. This significant auto-regressive parameter was not detected with listwise deletion method. The auto-regressive parameter for husbands was not significantly different from zero using all three missing data handling approaches, meaning husbands' conflict resolution scores were not predictable from his conflict resolution scores at the preceding time point. No significant cross-lagged regression parameters were found using all the missing data handling methods considered. Therefore, husbands' conflict resolution scores were not predicted by wives' conflict resolution scores at the end of immediate preceding time point and vice versa.

All three methods found significant influence of child variables on the dependent variables. To be specific, child agentic behavior was found to predict increase in both husbands' and wives' conflict resolution scores, while child negative emotion was associated with decrease in both husbands' and wives' conflict resolution scores. Even though all three methods yielded similar statistically significant findings and hence conclusions, the magnitudes of the parameter estimates were different. For instance, we can see from Table 3.1 that for child agentic behavior ( $x_1$ ), both the magnitude of the parameter point estimates and the parameter standard error estimates are larger with multiple imputation compared with those when the original method was applied. This may suggest stronger influence of child agentic behavior over parents' emotion states, meanwhile reflecting uncertainties of the parameter estimates due to missingness in the child variable. The reason for the difference is that the original method re-coded the child variables in a way that equated instances with no agentic behavior or negative emotion to instances with missing child covariate information. Process noises parameter estimates were similar with all three methods.

Table 3.1: Parameter Estimates for Empirical Data

Parameter	Original method			Listwise deletion			Multiple imputation		
	$\hat{\theta}$	$SE_{\theta}$	$t_{value}$	$\hat{\theta}$	$SE_{\theta}$	$t_{value}$	$\hat{\theta}$	$SE_{\theta}$	$t_{value}$
$a_{w \rightarrow w}$	-0.091	0.043	-2.116*	-0.130	0.086	-1.512	-0.088	0.044	-2.000*
$b_{h \rightarrow w}$	0.056	0.043	1.302	0.134	0.085	1.576	0.055	0.043	1.279
$b_{w \rightarrow h}$	0.011	0.044	0.250	-0.076	0.089	-0.854	0.013	0.044	0.295
$a_{h \rightarrow h}$	-0.043	0.043	-1.000	0.034	0.088	0.386	-0.043	0.043	-1.000
$c_{x_1 \rightarrow w}$	0.753	0.215	3.502*	2.311	0.634	3.645*	1.549	0.555	2.791*
$d_{x_2 \rightarrow w}$	-0.733	0.166	-4.416*	-0.287	0.057	-5.035*	-0.506	0.089	-5.685*
$c_{x_1 \rightarrow h}$	0.608	0.218	2.789*	1.470	0.655	2.244*	1.167	0.509	2.293*
$d_{x_2 \rightarrow h}$	-0.698	0.168	-4.155*	-0.254	0.059	-4.305*	-0.429	0.087	-4.931*
$\sigma_{e_w}^2$	7.652	0.233	32.841*	7.319	0.397	18.436*	7.476	0.239	31.280*
$\sigma_{e_h}^2$	7.836	0.239	32.787*	7.805	0.424	18.408*	7.705	0.240	32.104*
$\sigma_{e_{wh}}$	6.688	0.221	30.262*	6.722	0.388	17.325*	6.540	0.223	29.327*

\*  $p < .05$

# Chapter 4

## Discussion

In this study, we illustrated and examined the performance of partial MI and full MI in the context of intensive longitudinal data analysis in fitting a bivariate vector autoregressive model with covariates. We evaluated the relative strengths and limitations of the two approaches under different missingness conditions and number of measurement occasions conditions with simulation studies. Model parameter estimation results were compared between the partial MI and the full MI approach, as well as with the listwise deletion approach. Four main findings emerged. First, consistent with previous findings (Little & Rubin, 2014), doing MI using a mildly misspecified imputation model still led to better performance than listwise deletion. By retaining the original spacing (ie.g., the correct time intervals) between adjacent observations, both imputation approaches outperformed listwise deletion by yielding smaller biases and RMSEs in the point estimates. Both MI approaches also performed better than listwise deletion in standard error estimation under all missingness conditions. Second, point estimates from the partial MI approach, especially those associated with the time series parameters, were found to have better accuracy, precision and coverage properties in general compared to the full MI approach. In contrast, the full MI approach was found to yield higher accuracy in standard error estimation, particularly in situations involving nonignorable missingness. Both approaches were found to yield reasonable results in handling percentages of missingness commonly encountered in empirical applications. Third, even though the full MI approach yielded more accurate standard error estimates than the partial MI approach under the NMAR condition, standard error estimates with the partial MI approach improved and became closer to the empirical standard errors under this condition with larger  $T$ . Finally, larger  $T$  did not help improve the accuracy of the point estimates of the listwise deletion approach. For smaller  $T$ , the advantages of MI in improving parameter standard error estimates are especially salient.

In this particular study, when implementing MI to the empirical data, the key results and conclusions were consistent with the results reported earlier wherein the missing covariates were simply coded as 0. However, listwise deletion of observations with missing covariates did change the magnitudes of the point and standard error estimates, and may lead to some discrepancies in results in other applications with high instances of missingness in covariates. In addition, in other studies, it may not always be theoretically plausible to recode the missing covariates as 0.

In general, results from this article demonstrated advantages of the MI approaches in handling missing data in intensive longitudinal studies. First, unlike listwise deletion, which alters the spacing between the measurement time points, MI approach preserves the original observed time intervals, thus leading to more accurate parameter estimation in time series models. Meanwhile, MI also takes into consideration the uncertainty of the imputed values by assimilating such uncertainty through the generation of multiple imputed data sets and including between-imputation variations in parameter estimates in the overall standard error estimates. Second, MI approaches – including the full MI and partial MI approaches considered in the present study – are more flexible from an implementation as well as estimation standpoints compared with a full FIML approach or pattern-mixture modeling. Statistical packages doing MI usually allow different imputation models to accommodate various data characteristics. In our simulation study, we included one continuous variable and one binary variable as covariates of the dynamic model, and estimation results were reasonable under both MI approaches. Number of measurement time points is also not a constraint with MI-based approaches. The two conditions we used in our simulation study have 15 and 75 time points, which would have yielded  $2^{15}$  and  $2^{75}$  different missingness patterns if pattern-mixture modeling was used. Clearly, this would pose great computational challenges. Third, it is easier to include auxiliary variables with the MI approaches than a FIML approach (Collins et al., 2001). With MI, it is straightforward to include as many auxiliary variables as appropriate as part of the imputation model. In our simulation study, under NMAR condition, we included two fully observed auxiliary variables in the imputation model, resulting in nine variables in total. It took around 50 seconds to finish five imputations with a Windows 8 PC using Rstudio version 3.1.2 ( $N = 100$ ,  $T = 75$ ).

Some discrepancies between the results reported in this study and those from the literature should be noted. For instance, Liu and Molenaar (2014) claimed that imputing missing data in multivariate time series using VAR models yielded better estimates for the cross-lagged coefficients than a MI approach they considered. However, their true model was a time series model and they did not incorporate lagged information in the imputation model for their MI procedure and neither did they include auxiliary variables. We argue that the MI approaches would perform equally well, if not better, if appropriate lagged dependent variables and appropriate auxiliary variables were included in the imputation model. Further simulation studies are needed to verify this.

It is also important to point out that auxiliary variables should be carefully chosen when we perform MI, or biases may actually increase as opposed to decrease with the inclusion of auxiliary variable under certain conditions (Thoemmes & Rose, 2014). Kano (2015) demonstrated mathematically that including auxiliary variable will enlarge the estimating bias under following condition:

$$|\rho_{yr}| < |\rho_{yr} - \rho_{yx}\rho_{xr}|,$$

where  $|\rho_{yr}|$  represents the absolute value of the correlation between a dependent variable ( $y$ ) and a missingness indicator ( $r$ ). And  $\rho_{yx}\rho_{xr}$  represents the product of the correlation between an auxil-

iary variable ( $x$ ) and the dependent variable, and that between auxiliary variable and missingness indicator. In empirical applications, it would be helpful to include sensitivity tests to evaluate how inclusion of various auxiliary variables would change the model estimation results (Thoemmes & Rose, 2014).

In addition to auxiliary variables, another important decision researchers need to make is how many replications are needed for the MI. Rubin (2004) demonstrated theoretically that 3 to 5 imputations would be sufficient under most realistic circumstances, which has been used as a general guidance by many researchers in deciding the number of imputations to use in empirical applications. However, with more complicated models and longitudinal data, some researchers have also proposed using a larger number of imputations to improve the stability of confidence interval estimation (Royston et al., 2004), obtain more accountable conclusions with empirical data (Spratt et al., 2010), and to get more reliable model selection results with longitudinal data (Shen & Chen, 2013). In the current study, we used 5 replications for the MI procedures in the simulation study. This appeared sufficient for the particular settings considered in this study as increasing the number of replications to 10 did not affect the results substantially. However, in other studies involving higher percentages or more complex patterns of missingness, smaller sample sizes, and weaker correlations, more number of replications may be needed (Lu, 2014).

## 4.1 Limitations

Results of this study only pertain to the current model (i.e., bivariate order one autoregressive model) and sample size configurations. Findings of the current study about the performance of partial MI and full MI may not generalize to studies involving other sample size configurations. For example, with percentage of missingness larger than 30%, more imputations may be necessary for MI to achieve stable estimation. Smaller sample sizes and shorter time series may also affect the result of MI. In addition, the two MI approaches considered in this study are both two-step procedures (i.e., with imputation followed by estimation of modeling parameters as if the imputed data were observed). Thus, further simulation studies are warranted to investigate the performance of the partial and full MI approaches under more complicated models (e.g. clustered data), and in comparison to other one-step approaches that perform the imputation and parameter estimation simultaneously. Finally, we did not address the effect of mis-specification of imputation models on the results from MI in our simulation study. Even though missing data mechanism in general is not directly falsifiable especially in cases involving non-ignorable missingness, a seriously flawed imputation model which deviates substantially from the true missingdata mechanism will very likely lead to biased estimates and improper interpretations (Barnard & Meng, 1999). Sensitivity test will be helpful in evaluating the reliability of MI under different scenarios.

Moving forward, it would be of interest to test and evaluate the performance of MI approaches with other dynamic models, such as dynamic models involving latent variables and categorical



dependent variables, and fully Bayesian approach for one-step model-fitting and imputation.

# Appendix |

## R Code for Full/Partial MI, and Pooling Parameter Estimates

```
# create data set for MI with variables need to be imputed and auxiliary variables
nmarformice=cbind(nmartempmisswp,nmartempmisshp,nmartempca,nmartempcn,
  nmartempwplag1,nmartempwplag2,nmartempwplag3,nmartempwplag4,nmartempwplag5,
  nmartempwplag6,nmartempwplag7,nmartempwplag8,nmartempwplag9,nmartempwplag10,
  nmartempwplag11,nmartempwplag12,nmartempwplag13,nmartempwplag14,nmartempwplag15,
  nmartempwplag16,nmartempwplag17,nmartempwplag18,nmartempwplag19,nmartempwplag20,
  nmartempwplag21,nmartempwplag22,nmartempwplag23,nmartempwplag24,nmartempwplag25,
  nmartempwplag26,nmartempwplag27,nmartempwplag28,nmartempwplag29,nmartempwplag30,
  nmartempwplag31,nmartempwplag32,nmartempwplag33,nmartempwplag34,nmartempwplag35,
  nmartempwplag36,nmartempwplag37,nmartempwplag38,nmartempwplag39,nmartempwplag40,
  nmartempwplag41,nmartempwplag42,nmartempwplag43,nmartempwplag44,nmartempwplag45,
  nmartempwplag46,nmartempwplag47,nmartempwplag48,nmartempwplag49,nmartempwplag50,
  nmartempwplag51,nmartempwplag52,nmartempwplag53,nmartempwplag54,nmartempwplag55,
  nmartempwplag56,nmartempwplag57,nmartempwplag58,nmartempwplag59,nmartempwplag60,
  nmartempwplag61,nmartempwplag62,nmartempwplag63,nmartempwplag64,nmartempwplag65,
  nmartempwplag66,nmartempwplag67,nmartempwplag68,nmartempwplag69,nmartempwplag70,
  nmartempwplag71,nmartempwplag72,nmartempwplag73,nmartempwplag74,nmartempwplag75,
  nmartempwplag76,nmartempwplag77,nmartempwplag78,nmartempwplag79,nmartempwplag80,
  nmartempwplag81,nmartempwplag82,nmartempwplag83,nmartempwplag84,nmartempwplag85,
  nmartempwplag86,nmartempwplag87,nmartempwplag88,nmartempwplag89,nmartempwplag90,
  nmartempwplag91,nmartempwplag92,nmartempwplag93,nmartempwplag94,nmartempwplag95,
  nmartempwplag96,nmartempwplag97,nmartempwplag98,nmartempwplag99,nmartempwplag100)
nmarformicef=data.frame(nmarformice)
nmarformicef[,3]=as.factor(nmarformicef[,3])
levels(nmarformicef[,3])

# perform imputation
nmarimp=mice(nmarformicef,m=m)

##### Full MI #####

# initial list to store outputs
k=11 # number of parameters estimated
qhat=matrix(NA, nrow=m,ncol=k) # parameter estimates from each imputation
u=array(NA,dim=c(k,k,m)) # vcov of par estimates from each imputation

for (i in 1:m) {

  data.impute=complete(nmarimp,action=i) # retrieve the imputed data sets
  wpforfit=matrix(data.impute[,1],nrow=n,byrow=TRUE)
  hpforfit=matrix(data.impute[,2],nrow=n,byrow=TRUE)
  cnforfit=matrix(data.impute[,4],nrow=n,byrow=TRUE)
  caforfit=matrix(data.impute[,3],nrow=n,byrow=TRUE)
  datafile=sprintf(paste("nardata%i", ".txt",sep=""),i)
  datafile1=sprintf(paste("nardata%i", ".txt",sep=""),i)
```

```

fileKF =sprintf(paste("mk%i", ".txt", sep="", collapse=""),i)
filebat=sprintf(paste("run%i", ".bat", sep="", collapse=""),i)

# rearrange the data set to mkfm6 format
temp=cbind(wpforfit, hpforfit, caforfit, cnforfit)
writemkfm(temp, ne, nt, datafile)

# compile script for mkfm6
source(paste("compileKFscript.R", sep=""))
source(paste("compilebat.R", sep=""))

# call mkfm6
system(sprintf(paste("run%i", ".bat", sep="", collapse=""),i), wait = TRUE, intern=TRUE)

# read in mkfm6 estimation results
if (file.info("pars.out")$size > 0){
  nmarpars=scan("pars.out")
  if (length(nmarpars)==22){
    noRun=c(noRun, run)
  }else {
    qhat[i,]=nmarpars[seq((k*k+1), length(nmarpars), 2)]
    u[, ,i]=matrix(nmarpars[1:(k*k)], ncol=k)
  }
  }else {
  noRun = c(noRun, run)
}
}

##### Partial MI #####

# initial list to store outputs
k=11 # number of parameters estimated
qhat=matrix(NA, nrow=m, ncol=k) # parameter estimates from each imputation
u=array(NA, dim=c(k,k,m)) # vcov of par estimates from each imputation

for (i in 1:m) {

  data.impute=complete(nmarimp, action=i) # retrieve the imputed data sets
  cnforfit=matrix(data.impute[,4], nrow=n, byrow=TRUE)
  caforfit=matrix(data.impute[,3], nrow=n, byrow=TRUE)
  datafile=sprintf(paste("nardata%i", ".txt", sep=""),i)

```

```

datafile1=sprintf(paste("nardata%i",".txt",sep=""),i)
fileKF =sprintf(paste("mk%i",".txt",sep=""),collapse=""),i)
filebat=sprintf(paste("run%i",".bat",sep=""),collapse=""),i)

# rearrange the data set to mkfm6 format
temp=cbind(wpforfit,hpforfit,caforfit,cnforfit)
temp[is.na(temp)]=-999
writemkfm(temp,ne,nt,datafile)

# compile script for mkfm6
source(paste("compileKFscript.R",sep=""))
source(paste("compilebat.R",sep=""))

# call mkfm6
system(sprintf(paste("run%i",".bat",sep=""),collapse=""),i,wait = TRUE,intern=TRUE)

# read in mkfm6 estimation results
if (file.info("pars.out")$size > 0){
  nmarpars=scan("pars.out")
  if (length(nmarpars)==22){
    noRun=c(noRun,run)
  }else {
    qhat[i,]=nmarpars[seq((k*k+1),length(nmarpars),2)]
    u[, ,i]=matrix(nmarpars[1:(k*k)],ncol=k)
  }
  }else {
    noRun = c(noRun,run)
  }
}

##### Pool Results across Imputations #####

# qhat represents M sets of parameter point estimates
# u is a list of M sets of variance covariance matrix of the parameter estimates

# Calculate average parameter point estimates across M sets of model fitting results

qbarimpute <- apply(qhat, 2, mean)

# Calculate pooled standard error estimates
ubarimpute <- apply(u, 1:2, mean)

```

```
e <- qhat - matrix(qbarimpute, nrow = m, ncol = k, byrow = TRUE)
b <- (t(e) %*% e)/(m - 1)
vcov <- ubarimpute + (1 + 1/m) * b
se=sqrt(diag(vcov))
```

# References

- Affi, A., & Elashoff, R. (1966). Missing observations in multivariate statistics i. review of the literature. *Journal of the American Statistical Association*, *61*(315), 595–604.
- Akima, H. (1970). A new method of interpolation and smooth curve fitting based on local procedures. *Journal of the ACM (JACM)*, *17*(4), 589–602. doi: 10.1145/321607.321609
- Akima, H. (1991). A method of univariate interpolation that has the accuracy of a third-degree polynomial. *ACM Transactions on Mathematical Software (TOMS)*, *17*(3), 341–366. doi: 10.1145/114697.116810
- Allison, P. D. (1987). Estimation of linear models with incomplete data. *Sociological methodology*, *17*(1), 71–103.
- Allison, P. D. (2000, 02). Multiple imputation for missing data: A cautionary tale. *Sociological Methods & Research*, *28*(3). doi: 10.1177/0049124100028003003
- Allison, P. D. (2002). Missing data: Quantitative applications in the social sciences. *British Journal of Mathematical and Statistical Psychology*, *55*(1), 193–196.
- Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of abnormal psychology*, *112*(4), 545. doi: 10.1037/0021-843X.112.4.545
- Baltes, P. B., & Nesselroade, J. R. (1979). *Longitudinal research in the study of behavior and development*. Academic Press.
- Barnard, J., & Meng, X.-L. (1999). Applications of multiple imputation in medical studies: from aids to nhanes. *Statistical Methods in Medical Research*, *8*(1), 17–36.
- Chow, S.-M., Hamagami, F., & Nesselroade, J. R. (2007). Age differences in dynamical cognition-emotion linkages. *Psychology and Aging*, *22*(4), 765–780.
- Chow, S.-M., Ho, M.-h. R., Hamaker, E. L., & Dolan, C. V. (2010). Equivalence and differences between structural equation modeling and state-space modeling techniques. *Structural Equation Modeling*, *17*(2), 303–332. doi: 10.1080/10705511003661553
- Chow, S.-M., Nesselroade, J. R., Shifren, K., & McArdle, J. J. (2004). Dynamic structure of emotions among individuals with Parkinson’s disease. *Structural Equation Modeling*, *11*, 560–582. doi: 10.1207/s15328007sem1104\_4
- 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*(4), 330.
- Fahrenberg, J., & Myrtek, M. (2001). *Progress in ambulatory assessment: Computer-assisted psychological and psychophysiological methods in monitoring and field studies*. Hogrefe & Huber Publishers.
- Glasser, M. (1964). Linear regression analysis with missing observations among the independent variables. *Journal of the American Statistical Association*, *59*(307), 834–844.
- Harel, O., & Zhou, X.-H. (2007). Multiple imputation: review of theory, implementation and software. *Statistics in medicine*, *26*(16), 3057–3077.
- Harvey, A. C. (2001). *Forecasting, structural time series models and the kalman filter*. Cambridge: Cambridge University Press.

- Hedeker, D., & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological methods*, 2(1), 64. doi: 10.1037/1082-989X.2.1.64
- Horton, N. J., & Kleinman, K. P. (2007). Much ado about nothing. *The American Statistician*, 61(1). doi: 10.1198/000313007X172556
- Horton, N. J., & Lipsitz, S. R. (2001). Multiple imputation in practice: comparison of software packages for regression models with missing variables. *The American Statistician*, 55(3), 244–254. doi: 10.1198/000313001317098266
- Ibrahim, J. G. (1990). Incomplete data in generalized linear models. *Journal of the American Statistical Association*, 85(411), 765–769.
- Ibrahim, J. G., Chen, M.-H., Lipsitz, S. R., & Herring, A. H. (2005). Missing-data methods for generalized linear models: A comparative review. *Journal of the American Statistical Association*, 100(469), 332–346. doi: 10.1198/016214504000001844
- Jones, M. P. (1996). Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of the American Statistical Association*, 91(433), 222–230.
- Kano, Y. (2015). *Developments in multivariate missing data analysis*. Paper presented at the meeting of Psychometric Society 2015, Beijing China. Retrieved from <http://www.sigmath.es.osaka-u.ac.jp/kano/research/paper/IMPS2015.pdf>
- King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. In *American political science association* (Vol. 95, pp. 49–69). Retrieved from [http://journals.cambridge.org/article\\_S0003055401000235](http://journals.cambridge.org/article_S0003055401000235)
- Little, R. J. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88(421), 125–134.
- Little, R. J., & Rubin, D. B. (2014). *Statistical analysis with missing data*. John Wiley & Sons.
- Liu, S., & Molenaar, P. C. (2014). ivar: A program for imputing missing data in multivariate time series using vector autoregressive models. *Behavior research methods*, 46(4), 1138–1148. doi: 10.3758/s13428-014-0444-4
- Lu, K. (2014). Number of imputations needed to stabilize estimated treatment difference in longitudinal data analysis. *Statistical methods in medical research*, 0962280214554439.
- McArdle, J. J., & Hamagami, F. (1992). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. *Experimental aging research*, 18(3), 145–166.
- Meng, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science*, 538–558.
- Molenaar, P. C. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, 50(2), 181–202.
- Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, 52(3), 431–462.
- Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey methodology*, 27(1), 85–96.
- Royston, P., et al. (2004). Multiple imputation of missing values. *Stata journal*, 4(3), 227–41.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592.
- Rubin, D. B. (1977). Formalizing subjective notions about the effect of nonrespondents in sample surveys. *Journal of the American Statistical Association*, 72(359), 538–543.
- Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American statistical Association*, 91(434), 473–489.
- Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys* (Vol. 81). John Wiley & Sons.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. CRC press.

- Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical methods in medical research*, 8(1), 3–15. doi: 10.1177/096228029900800102
- Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods*, 7(2), 147. doi: 10.1037/1082-989X.7.2.147
- Schermerhorn, A. C., Chow, S.-M., & Cummings, E. M. (2010). Developmental family processes and interparental conflict: patterns of microlevel influences. *Developmental psychology*, 46(4), 869–885.
- Shen, C.-W., & Chen, Y.-H. (2013). Model selection of generalized estimating equations with multiply imputed longitudinal data. *Biometrical Journal*, 55(6), 899–911.
- Sinharay, S., Stern, H. S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological methods*, 6(4), 317.
- Spratt, M., Carpenter, J., Sterne, J. A., Carlin, J. B., Heron, J., Henderson, J., & Tilling, K. (2010). Strategies for multiple imputation in longitudinal studies. *American journal of epidemiology*, 172(4), 478–487.
- Thoemmes, F., & Rose, N. (2014). A cautious note on auxiliary variables that can increase bias in missing data problems. *Multivariate behavioral research*, 49(5), 443–459. doi: 10.1080/00273171.2014.931799
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3), 1–67. Retrieved from <http://www.jstatsoft.org/v45/i03/>
- Van Buuren, S., Boshuizen, H. C., Knook, D. L., et al. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in medicine*, 18(6), 681–694.
- Verbeke, G., & Molenberghs, G. (2000/2009). *Linear mixed models for longitudinal data*. Springer Science & Business Media. (Original work published 2000)
- Zeileis, A., & Grothendieck, G. (2005). zoo: S3 infrastructure for regular and irregular time series. *Journal of Statistical Software* 14(6), 1-27.
- Zhang, P. (2003). Multiple imputation: theory and method. *International Statistical Review*, 71(3), 581–592.