

ABSTRACT

Title of dissertation: HANDLING OF MISSING
DATA WITH GROWTH MIXTURE MODELS

Daniel Y. Lee, Doctor of Philosophy, 2019

Dissertation directed by: Professor Jeffrey R. Harring
University of Maryland
Department of Human Development and
Quantitative Methodology

The recent growth of applications of growth mixture models for inference with longitudinal data has introduced a wide range of research dedicated to testing the different aspects of the model. One area of research that has not drawn much attention, however, is the performance of growth mixture models with missing data and when using the various methods for dealing with them. Missing data are usually an inconvenience that must be addressed in any data analysis scenario, and the use of growth mixture models is no less an exception to this. While the literature on various other aspects of growth mixture models has grown, not much research has been conducted on the consequences of mishandling missing data. Although the literature on missing data has generally accepted the use of modern missing data handling techniques, these techniques are not free of problems nor have they been comprehensively tested in the context of growth mixture models. The purpose of this dissertation is to incorporate the various missing data handling techniques on growth mixture models and, by using Monte Carlo simulation techniques, to pro-

vide guidance on specific conditions in which certain missing data handling methods will produce accurate and precise parameter estimates typically compromised when using simple, ad hoc, missing data handling approaches, or incorrect techniques.

HANDLING OF MISSING DATA
WITH GROWTH MIXTURE MODELS

by

Daniel Y. Lee

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor Jeffrey R. Harring, Chair/Advisor
Professor Xin He
Professor Laura M. Stapleton
Professor Tracy M. Sweet
Professor Ji Seung Yang

© Copyright by
Daniel Y. Lee
2019

Dedication

This dissertation is dedicated to my mom, dad, and Yehyun. They have been a constant reminder that no matter what happens, at the end of the day, everything will be okay.

Acknowledgments

An overwhelming number of people have been a part of this accomplishment. Among them, there are a select group of individuals I would especially like to acknowledge. Without their support, finishing this dissertation would have been impossible.

First, I would like to thank the EDMS family, especially my advisor, Jeffery R. Harring, who has gone above and beyond simply being my advisor. Throughout the years he has become a close mentor, colleague, and friend. I would also like to thank the members of my dissertation committee, Ji Seung Yang, Tracy Sweet, Laura Stapleton and Xin He, as well as other EDMS professors I've had the pleasure to collaborate with and learn from during my time in Maryland: Gregory Hancock, Hong Jiao and Robert Lissitz. I would also like to acknowledge the HDQM staff, especially Jannitta, Charm, and Cornelia, for their amazing administrative support.

Some of the best memories I made during my time in the program were with classmates I used to share the office space in Cole with. I would like to thank them for all the memories collaborating on assignments, talking smack about the undergrads, and introducing me to the Wednesdays Lamb shank special over at Stamp. I would especially like to thank my brother in Christ and colleague Kaiwen for his unwavering friendship since the beginning.

My last few years in the program would not have been possible if not for CAL. In particular, I would like to thank Dorry Kenyon, Shu Jing Yen, and the PQR team for making my job at CAL stress free so that I could keep my mind on finishing this dissertation. I am also very thankful for the group of colleagues who were constantly checking up on my progress.

Miriam, Andrew, Charlotte, Eugene, Ed, Joe, Chunky, and other family and friends in Maryland, New Jersey, New York, and Korea have also been a major influence in my life during this time. I would like to thank them for reminding me that there are more important things in life than work.

Lastly, above all, I would like to thank Jesus Christ, my lord and my savior, for blessing me with the health, motivation, and intellect to reach this milestone.

Table of Contents

Dedication	ii
Acknowledgements	iii
List of Tables	vii
List of Figures	x
1 Introduction	1
1.1 Growth Mixture Models and Missing Data	1
1.2 The Current Study	6
2 Literature Review	8
2.1 Missing Data	8
2.1.1 Missing Data Taxonomy	9
2.1.2 Methods for Addressing Missing Data	12
2.1.2.1 Original Methods	13
2.1.2.2 Maximum Likelihood	17
2.1.2.3 Expectation Maximization Algorithm	20
2.1.2.4 Multiple Imputation	21
2.1.2.5 Bayesian Inference Behind Multiple Imputation	23
2.1.2.6 Joint Multivariate MI	24
2.1.2.7 Fully Conditional Specification Multiple Imputation	26
2.2 Growth Models	26
2.3 Growth Models Extension to Finite Mixtures	31
2.3.1 Estimation of Growth Mixture Models	33
2.3.1.1 Maximum Likelihood Estimation with the EM Algorithm	33
2.3.1.2 Fully Bayesian Estimation	35
2.4 Handling Missing Data with Growth Mixture Models	41
2.4.1 Maximum Likelihood	41
2.4.2 Bayesian Methods	43
2.4.3 Single-Stage Multiple Imputation	44

2.4.4	Two-Stage Multiple Imputation	45
2.4.5	Summary: Comparison of Methods	51
2.5	Research Goals	52
3	Methods	54
3.1	Step 1: Data Generation	55
3.1.1	Class Separation	56
3.1.2	Sample Size	60
3.2	Step 2: Data Amputation	61
3.2.1	MNAR Missingness (CMAR ⁺)	61
3.2.2	MCAR Missingness	62
3.3	Step 3: Missing Data Handling Approaches	63
3.3.1	Label Switching	63
3.3.2	Convergence Issues	65
3.3.3	Number of Imputations	66
3.3.4	Choice of Priors	67
3.4	Summary of Manipulated Conditions	69
3.5	Step 4: Evaluation Criteria	72
3.5.1	Convergence Rates	72
3.5.2	Accuracy Rates	73
3.5.3	Relative Bias and Absolute Bias	73
3.5.4	Standard Error Bias	74
3.5.5	Factorial ANOVA	75
3.6	Expectations from the Study	76
4	Results	78
4.1	Complete Data Analysis	81
4.1.1	Convergence Rates	81
4.1.2	Classification Accuracy Rates	82
4.1.3	Relative Bias	84
4.1.4	Standard Error Bias	92
4.2	Missing Data Analysis	98
4.2.1	Convergence Rates	98
4.2.2	Classification Accuracy Rates	99
4.2.3	Relative Bias	100
4.2.3.1	Results of the Main Effects	106
4.2.3.2	Results of the Interaction Effects	110
4.2.4	Standard Error Bias	124
4.2.4.1	Results of the Main Effects	129
4.2.4.2	Results of the Interaction Effects	130
4.3	Summary of Main Findings	134
4.3.1	Full Data Analysis	134
4.3.2	Missing Data Analysis	135

5	Discussion	137
5.1	Recommendations	142
5.2	Limitations and Future Studies	143
A	Complete Data Methods Relative Bias and SE/SD Ratio Tables	149
B	Missing Data Methods SE/SD Ratios and Pairwise Comparison Tables	152
C	Code to Implement the Two-Stage Imputation Method in Mplus	162
	Bibliography	166

List of Tables

3.1	Data Generating Parameters for Varying MD Conditions	59
3.2	Coefficients for Creating Missing Data	62
3.3	Mean Relative Bias Values by Analysis Method (Complete Data Methods)	69
3.4	Overview of Simulation Conditions	70
4.1	Abbreviations and Symbols Used for Summary of Results	80
4.2	Mean Convergence Rates for Complete Data Methods Grouped by Separation and Sample Size	82
4.3	Mean Classification Accuracy Rates for Complete Data Methods Grouped by Separation and Sample Size	84
4.4	Means and Medians of the Overall Absolute Relative Bias Grouped by Method, Sample Size (top), and Class Separation (bottom)	85
4.5	Factorial ANOVA Results for Relative Bias of the Estimates of Parameter Means (Complete Data Methods)	88
4.6	Factorial ANOVA Results for Relative Bias of the Estimates of Parameter Variances (Complete Data Methods)	89
4.7	Mean Relative Bias Grouped by Method and Sample Size for Estimates of $\mu_{int_{C_2}}$, Ψ_{slope} , Θ , and π_{c1} (Complete Data Methods)	90
4.8	Mean Relative Bias Grouped by Method and Separation for Estimate of π_{c1} (Complete Data Methods)	91
4.9	Mean Relative Bias Values Grouped by Analysis Method (Complete Data Methods)	91
4.10	Means and Medians of the SE/SD Ratios Grouped by Method, Sample Size (top), and Separation (bottom) for Complete Data Methods	93
4.11	Factorial ANOVA Results for SE/SD Ratios of the Estimates of the Mean Parameters (Complete Data Methods)	94
4.12	Factorial ANOVA Results for SE/SD Ratios of the Estimates of the Variance Parameters (Complete Data Methods)	94
4.13	Mean SE/SD Ratios Grouped by Method and Sample Size (Complete Data Methods)	95
4.14	Mean Convergence Rates for Missing Data Methods Grouped by Method and Sample Size	99

4.15	Mean Classification Accuracy Rates for Missing Data Methods Grouped by Method and Separation	100
4.16	Means and Medians of the Overall Absolute Relative Bias Grouped by Method, Sample Size (Top), and Separation (Bottom) (Missing Data Methods)	101
4.17	Means and Medians of the Overall Absolute Relative Bias Grouped by Method and Missing Rate	104
4.18	Means and Medians of the Overall Absolute Relative Bias Grouped by Method and Missing Mechanism	105
4.19	Factorial ANOVA Results for Relative Bias of Estimates of Parameter Means (Missing Data Methods)	106
4.20	Factorial ANOVA Results for Relative Bias of Estimates of Parameter Variances (Missing Data Methods)	106
4.21	Mean Relative Bias Grouped by Levels of Separation (SP)	108
4.22	Mean Relative Bias Grouped by Analysis Methods (ME)	109
4.23	Mean Relative Bias Grouped by Sample Size (SS)	109
4.24	Main Effect Pairwise Comparisons Corresponding to ME \times SS Interaction of Intercept Variances for Relative Bias	113
4.25	Main Effect Pairwise Comparisons Corresponding to ME \times SS Interaction of Slope Variances for Relative Bias	116
4.26	Main Effect Pairwise Comparisons Corresponding to ME \times SS Interaction of Intercept Slope Covariances for Relative Bias	118
4.27	Main Effect Pairwise Comparisons Corresponding to ME \times SP Interaction of Proportion Parameter for Relative Bias	120
4.28	Mean Relative Bias of the Residual Variance Estimate Grouped by Method, Missing Data Rate, and Sample Size	122
4.29	Overall Mean SE/SD Ratios for Missing Data Methods Grouped by Method, Sample Size (top), and Class Separation (bottom)	125
4.30	Factorial ANOVA Results for SE/SD Ratios of Estimates of Means (Missing Data Methods)	128
4.31	Factorial ANOVA Results for SE/SD Ratios of Estimates of Variances (Missing Data Methods)	128
4.32	Mean SE/SD Ratios Grouped by Analysis Method (ME)	129
A.1	Relative Bias for All Parameters by All Conditions Using Full Data Methods	150
A.2	SE/SD Ratios for All Parameters by All Conditions Using Full Data Methods	151
B.1	SE/SD Ratios For Mean Estimates Grouped by ME and SS (Missing Data Methods)	153
B.2	SE/SD Ratios For Variance Estimates Grouped ME and SS (Missing Data Methods)	154
B.3	Main Effect Pairwise Comparisons Corresponding to ME \times SS Interaction of Intercept Variances for SE/SD Ratios	155

B.4	Main Effect Pairwise Comparisons Corresponding to ME \times SS Interaction of Slope Variances for SE/SD Ratios	156
B.5	Main Effect Pairwise Comparisons Corresponding to ME \times SS Interaction of Intercept Slope Covariances for SE/SD Ratios	157
B.6	Main Effect Pairwise Comparisons Corresponding to ME \times SS Interaction of Residual Variances for SE/SD Ratios	158
B.7	Main Effect Pairwise Comparisons Corresponding to ME \times MR Interaction of Intercept Variances for SE/SD Ratios	159
B.8	Main Effect Pairwise Comparisons Corresponding to ME \times MR Interaction of Slope Variances for SE/SD Ratios	160
B.9	Main Effect Pairwise Comparisons Corresponding to ME \times MR Interaction of Residual Variances for SE/SD Ratios	161

List of Figures

3.1	True population class 1 and class 2 growth curves	59
3.2	Running means of parameter estimates	71
4.1	Overall means and medians of the absolute relative bias by ME and SS	86
4.2	Overall means and medians of the absolute relative bias by ME and SP	87
4.3	Overall means and medians of the SE/SD ratios by ME and SS	93
4.4	SE/SD ratios for the intercept mean, slope mean, intercept variance, slope variance and intercept slope covariance by method	96
4.5	Overall means and medians of the absolute relative bias by ME and SS (Missing data methods)	102
4.6	Overall means and medians of the absolute relative bias by ME and SP (Missing data methods)	103
4.7	Relative bias of the estimates of intercept variance, slope variance, and intercept slope covariance by ME and SS, and estimates of pro- portion by ME and SP	112
4.8	Three-way ME \times SS \times MR interaction plot of the relative bias of the residual variances	123
4.9	Overall means and medians of the SE/SD ratios by ME and SS (Miss- ing data methods)	126
4.10	Overall means and medians of the SE/SD ratios by ME and MR (Missing data methods)	127
4.11	Two-way ME \times SS interaction plot of the SE/SD ratios for inter- cept variance, slope variance, intercept slope covariance, and residual variance	131
4.12	Two-way ME \times MR interaction plot of the SE/SD ratios for intercept variance, slope variance, and residual variance	132

Chapter 1: Introduction

1.1 Growth Mixture Models and Missing Data

Finite mixture models are statistical models that allow the possibility of latent heterogeneous subpopulations. For example, subpopulations can exist based on observed variables such as gender, age, or ethnicity, but finite mixture models hypothesize the existence of unknown groups beyond these observed categorizations. Finite mixture models have been applied to a wide range of statistical models typically used when the interest lies in testing the assumption of latent subgroups. Finite mixture models have also been used to account for non-normal distributions and are otherwise known to be indirect applications of the model (Bauer, 2007; Titterington, Smith, & Makovm, 1985). Of interest to applied researchers is a particular finite mixture model in the context of longitudinal studies known as a growth mixture model (GMM; B. O. Muthén & Shedden, 1999). The GMM extends finite mixture models to repeated measures or panel data. The assumption is that there exists heterogeneous latent groups that become apparent through varying patterns of change over time. For example, Colder, Campbell, Ruel, Richardson, and Flay (2002) applied a GMM to identify classes of growth trajectories of adolescent alcohol use and found five distinct groups that exhibited different patterns of growth

distinguished by levels of emotional distress and risk taking. Another more recent example is a study done by [Musu-Gillette, Wigfield, Haring, and Eccles \(2015\)](#), who applied a GMM to identify groups of students showing different trajectories of self-concept in math, and how likely these groups were to choose a math-intensive major in college (a distal outcome). The use of GMMs has also been used in other disciplines such as psychiatry ([Lin, Narayan, Drevets, Ye, & Li, 2017](#)), medicine ([Hesser, Hedman, Lindfors, Andersson, & Ljótsson, 2017](#)), and public health ([Koning et al., 2016](#)). The increased popularity of GMMs in recent years kindled a number of methodological studies that focused on issues pertaining to using the model under varying conditions. Much of the methodological research has dealt with testing indices for data-model fit ([Henson, Reise, & Kim, 2007](#); [Nylund, Asparouhov, & Muthén, 2007](#)), evaluating methods that help decide on the number of subgroups ([Bauer & Curran, 2003](#); [McNeish & Haring, 2017](#); [Nylund et al., 2007](#); [Tofighi & Enders, 2007](#)), examining the classification accuracy of models ([Enders & Tofighi, 2008](#); [Peugh & Fan, 2012](#)), comparing estimation methods for finite mixtures ([Hipp & Bauer, 2006](#); [McLachlan & Krishnan, 2007](#)), or showing how different settings other than the default settings in popular software for estimating these models can impact parameter recovery and precision ([Li, Haring, & Macready, 2014](#)).

Despite the numerous research studies devoted to refining the use of GMMs, not much attention has been devoted to testing GMMs in the presence of missing data. In the social and behavioral sciences, missing data analysis merits special attention because repeated measures data used for fitting GMMs is typically rife with missing data due to various issues like data collection error, participant non-

response to specific items, drop-out, or failure to participate in at least one wave of data collection.

According to seminal work by [Rubin \(1976\)](#), missing data can be categorized into three distinct types: Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR). MCAR missingness is established when the missingness is a result of unsystematic omission. MAR missingness is established when the missingness is related to other variables in the dataset but not the variable containing the missing itself. MNAR missingness is established when the missingness is related to the variable that has the missing values or some unobserved variable. Rubin as well as others ([Enders, 2010](#); [Schafer & Graham, 2002](#)) have shown that while MCAR missingness is benign when the percentage of missing is not substantial, MAR and MNAR missingness and incorrect handling of such missing data can lead to misleading results in certain situations. To add to the difficulties of handling missing data that is MAR or MNAR, it is impossible to establish whether missingness is MAR or MNAR. That is, there is no formal test of MAR or MNAR. Fortunately, researchers have established reliable methods to address missing data when the missingness is MAR. These methods allow for more accurate parameter estimation than some traditional ad hoc methods, such as listwise deletion, that have been naively used in the past ([Schafer & Graham, 2002](#)).

The result of studies focused on dealing with missing data, together with the advancement of computer software readily available for applying sophisticated missing data handling techniques, has allowed two methods to become very popular: full-information maximum likelihood via the Expectation-Maximization algorithm

(FIML-EM) ([Arbuckle, 1996](#)) and multiple imputation (MI) ([Rubin, 1987](#)). These two archetypal approaches have been shown to produce comparable results when used to handle missing data under the assumptions of MAR and MCAR ([Collins, Schafer, & Kam, 2001](#); [Enders, 2010](#); [Graham, 2003, 2009](#)). FIML is an estimation technique that takes into consideration any of the missingness in the data. The usefulness of FIML lies in the EM algorithm, an iterative process that, without having to discard any information, alternates between an "expectation step" and a "maximization step" to arrive at optimized parameter estimates. Alternatively, the MI approach fills in the missing data multiple times prior to fitting a model of interest to the data. The filling-in process is also an iterative process, but unlike the EM algorithm, is rooted in Bayesian methods. The complete, multiply-imputed, datasets are then estimated using the main model of interest in a separate step. The resulting parameters from the separate analyses are then aggregated using special combination rules proposed by [Rubin \(2003\)](#). More details on these techniques to handle missing data, their advantages and disadvantages, will be explained in Chapter 2.

Studies on handling missing data in the context of growth mixture models recommend using FIML over MI ([Enders & Gottschall, 2011](#); [McLachlan & Peel, 2000](#); [Sterba, 2016](#)) because MI requires the grouping information to be known a priori in order to correctly impute data ([Enders, 2010](#)) for each group. Since the grouping information for mixture models is latent, the use of MI with mixture models is problematic. Failure to specify a grouping variable when using MI has been shown to produce biased parameter estimates and incorrect identification of classes ([Enders & Gottschall, 2011](#); [Sterba, 2016](#)).

Despite the clear advantages of using FIML over MI with mixture models, MI can still be an attractive alternative over FIML, particularly for situations in which covariates are part of the model. For example, practitioners prefer using MI because of the convenience of handling the missing data in a separate step using covariates that may not pertain to the main research question (Collins et al., 2001). Furthermore, standard conditional likelihood FIML estimation methods typically used in commercial statistical software such as Mplus (L. K. Muthén & Muthén, 1998) will remove cases that have incomplete covariate information because these likelihood models assume covariates are fixed and do not contain missing data (Sterba, 2014). As a result, software programs that are capable of estimating these conditional models require that covariate information be complete or any cases containing missing covariates be completely removed¹. This type of deletion has been shown to result in biased parameter estimates and decreased power, especially if the missingness is MAR or MNAR. For situations with a considerable amount of missing covariate data, this can cause significant sample size reductions that will lead to reduced power (potential increase in Type II error rates) and erroneous statistical inferences (Schafer & Graham, 2002).

To circumvent the complexities of handling missing data in the context of GMMs, more recent studies have turned to Bayesian methods to estimate GMMs with missing data. These studies consider situations in which the missingness is related to the latent class, constituting an MNAR missingness mechanism. For ex-

¹Sterba (2014) has shown a workaround to this default method and is discussed in a subsequent section

ample, studies like [Song and Lee \(2002\)](#), [Cai, Song, and Hser \(2010\)](#), and [Lu, Zhang, and Lubke \(2011\)](#) proposed using Bayesian methods to address missing data that were MAR when using GMMs, finding favorable outcomes when using such models. However, these methods are not as readily accessible and require additional analyses and specifications special to Bayesian statistics, such as specifying proper priors, that can complicate the process. In addition, running some models using Bayesian methods can take longer periods of time to converge depending on the complexity of the model and the scales of the variables containing missing values.

As an alternative, [Harel \(2007, 2009\)](#) suggested using a two-stage imputation approach to remedy the problems of using MI with mixture models. This idea, originally proposed by [Shen \(2000\)](#) as a way to impute data of different types ([Rubin, 2003](#); [Schafer & Graham, 2002](#)), was extended to the context of mixture regression models. According to [Harel \(2009\)](#), the two-stage imputation method allows for unbiased parameter estimates that are similar to those obtained from FIML estimation. However, this method has not yet been tested in the context of GMMs and is not readily available in standard software like Mplus. In addition, it is unclear as to the types of situations in which this method will work.

1.2 The Current Study

A number of studies have used some of these missing data handling methods but using different types of models. In addition, none of these studies have investigated the different conditions where these different techniques may be useful. The

purpose of this study is to compare the different methods—those that have been suggested for handling missing data in the context of growth mixture models as well as those not yet extended in this longitudinal context—using a Monte Carlo simulation. A variety of simulation conditions that are pertinent to practitioners interested in using GMMs will be tested to provide insight on the performance of these missing data handling techniques in comparison to each other.

Chapter 2: Literature Review

It is helpful to begin with an introduction to the array of topics that will be discussed in this study and that will comprise the literature review. The first section will be an introduction to missing data terminology. This section will then be followed by a brief introduction to growth models and growth models in the context of finite mixtures. To conclude, the existing literature on handling missing data vis-à-vis growth mixture models will be discussed in more detail.

2.1 Missing Data

The existence of missing data in educational, social and behavioral science datasets is ubiquitous and presents an issue mandating action. The problems that missing data can cause and the ramifications of ignoring these problems or improperly handling them have been extensively documented since the seminal work by [Rubin \(1976\)](#), who laid the foundation in this area by establishing a classification system for the different types of missing data mechanisms and providing suggestions on how to deal with them. Since then, various methods have been proposed to address missing data based on an array of data features as well as the proposed analytic model, including

1. type of missing data,
2. type of variables that are missing,
3. types of models,
4. software availability and,
5. assumptions that should be made about the missingness.

The last 30 years alone have produced substantial literature in this area. As a result, the methodology on handling missing data now spans a diverse number of disciplines. Some of the concepts that have been carefully studied and widely adopted as a result are discussed next.

2.1.1 Missing Data Taxonomy

[Rubin \(1976\)](#) regarded missing data as random variables with a probability distribution. Missing data were classified as having one of three mechanisms:

1. Missing completely at random (MCAR)
2. Missing at random (MAR)
3. Missing not at random (MNAR).

Thorough explanations of these mechanisms can be found in [Rubin \(1976\)](#), [Schafer and Graham \(2002\)](#), and [Enders \(2010\)](#). A summary of these mechanisms and their differences are presented next.

For data that is MAR, the missingness is independent of the missing variable, but dependent on other observed variables. If \mathbf{I} represents the set of indicators of the missing data in a dataset, then MAR is formally defined as the probability of \mathbf{I} when it is related to observed variables, \mathbf{Y}_{obs} , but not the missing variable, or \mathbf{Y}_{mis} . That is,

$$\Pr(\mathbf{I}|\mathbf{Y}_{comp} = \{\mathbf{Y}_{obs}, \mathbf{Y}_{miss}\}, \phi) = \Pr(\mathbf{I}|\mathbf{Y}_{obs}, \phi), \quad (2.1)$$

where \mathbf{Y}_{comp} is the complete data consisting of observed and unobserved data (\mathbf{Y}_{obs} and \mathbf{Y}_{miss} respectively), and \Pr represents a probability distribution. The ϕ term represents a vector of parameters that captures the relationship between \mathbf{I} and the observed data. To illustrate, a simple example of missing data under a MAR mechanism is when an individual's measure of income is missing not due to an individual embarrassment to report a low income, but because of another variable related to income, such as employment or unemployment status that has been observed. In this scenario, an individual could potentially have a missing income variable if the individual did not have a job or any other means of attaining income.

If the missingness were also independent of observed variables, then this would constitute MCAR. MCAR is the mechanism in which the missingness is independent of both the missing and observed variables. This type of missing can be regarded as unsystematic missingness because the missing values present themselves in a completely random manner. More formally, this means that

$$\Pr(\mathbf{I}|\mathbf{Y}_{comp} = \{\mathbf{Y}_{obs}, \mathbf{Y}_{miss}\}, \phi) = \Pr(\mathbf{I}|\phi). \quad (2.2)$$

MCAR situations arise, for example, through accidental omission or miscoding. The literature defines MCAR and MAR missingness to be ignorable because likelihood-based maximization methods can produce unbiased estimates even if the model for missing is ignored (Little & Rubin, 2002) due to how the likelihood of the missing data can be differentiated out of the overall likelihood. This concept will be expounded on in Section 2.1.2.2.

Finally, an MNAR mechanism is thought to be operating when the missingness is dependent on the variable with missing data. In other words, the probability of missing data, \mathbf{I} , is influenced by \mathbf{Y}_{mis} , above and beyond influences through \mathbf{Y}_{obs} . That is,

$$\Pr(\mathbf{I}|\mathbf{Y}_{comp} = \{\mathbf{Y}_{obs}, \mathbf{Y}_{miss}\}, \phi) = \Pr(\mathbf{I}|\mathbf{Y}_{obs}, \mathbf{Y}_{miss}, \phi). \quad (2.3)$$

In the context of the income example, an individual's missing income variable may be suspected of being MNAR if an income measure was missing because income was low or high, or some other unmeasured variable that caused an individual to leave the income field blank. MNAR is often referred to as a non-ignorable process because a missing data model is needed in conjunction with the analytic model to produce unbiased model estimates. According to the aforementioned definitions of missingness mechanisms, missing observable data caused by latent variables constitute a MNAR mechanism because latent variables are never observed. However, as will be discussed in Section 2.4.4, missingness caused by latent variables can be constituted as "ignorable" by a concept called conditional extended ignorability (Harel, 2003; Harel & Schafer, 2002, 2009).

Unless the missingness was planned, which may be the case in planned missing designs such as cohort-sequential or accelerated longitudinal designs, MAR or MNAR are untestable assumptions and the model causing the missing data is unknown to the researcher. Even though these are untestable assumptions, in most applied situations, missing data can be considered MAR or MNAR (Glynn, Laird, & Rubin, 1986). Using deletion or single imputation methods under assumptions of MAR or MNAR can produce biased estimates (Enders, 2010; Schafer & Graham, 2002). Studies have shown that using maximum likelihood (ML) or multiple imputation (MI) is more appropriate for these situations. When the missing data are assumed to be MNAR, studies have recommended using selection models (Diggle & Kenward, 1994) or pattern mixture models (Little, 1993) as part of a sensitivity analysis, although even with these advancements there are no guarantees that the methods will work to reduce bias (Enders, 2010). The next sections will further discuss the use of MI and ML for addressing missing data under MAR.

2.1.2 Methods for Addressing Missing Data

A plethora of strategies have been proposed to address missing data, especially in cases when the missing mechanisms are MAR and MNAR. Ignoring missing data in these cases has been shown to create biased parameter estimates and inflated standard errors (de Leeuw, Hox, & Huisman, 2003) which result in lower statistical power. This section will present some of the methods that have been used in the last several decades.

2.1.2.1 Original Methods

Prior to the development of the more sophisticated methods for addressing missing data, listwise and pairwise deletion were the standard practice. Because the missingness is assumed to be haphazard under MCAR, deletion methods may be appropriate, although they will more often lead to biased estimates and reduction of statistical power when the percentage of missing is high (Enders, 2010). Despite the consequences of using these methods, many studies still continue to use them (Peugh & Enders, 2004) and deletion methods continue to be the default method in many statistical software programs (i.e., SPSS).

Listwise deletion removes cases that have at least one missing value on variables used in an analysis, essentially leaving a complete dataset but at the cost of wasted information. In addition to lowering statistical power from the loss of data, if the missingness cannot be assumed MCAR, then listwise deletion will produce biased parameter estimates (Enders, 2010).

As a way to preserve data and power, pairwise deletion only removes cases on a variable-by-variable basis. That is, the deletion method removes only the cases that contain missing data for the variables that are being used for the analysis. In this scenario, individual cases with missing data are retained as long as they have data for the variables for which an analysis is being conducted. Pairwise deletion also requires the assumption of MCAR missingness or will otherwise produce distorted estimates, which has been shown through empirical research (Arbuckle, 1996; B. O. Muthén, Kaplan, & Hollis, 1987; Wothke, 2000). For example, pairwise

deletion is often the cause of nonpositive definite matrices (Little, 1992; Wothke, 1993) and difficulty in computation of standard errors due to indefinite sample sizes which present themselves when variable information is available for some but not for others (Little, 1992). Although deletion methods are generally never recommended, they were the only available methods in the advent of statistical computing and consequently became the default method.

A popular non-parametric way to correct for the discrepancies caused by deletion methods that is often used in survey methodology is the use of an inverse probability weighting approach. In typical survey studies, sampling is done in a probabilistic manner, allowing analyses conducted on these samples to be generalized to the larger population. Sampling weights are used in order for samples to represent the population and provide unbiased estimates of population parameters (Kalton, 1983; Kish, 1965). In the presence of missing data, sampling weights can be adjusted by multiplying an inverse of a response rate to the weights and applying them to any individuals without missing data. Individuals with missing data are listwise deleted, and those remaining are given a weight that accounts for the missingness as well. These inverse weights are constructed using variables that correlate with both response propensity and the outcome variable. However, it has been shown that leaving out some of these variables from the adjustments can cause some biased variance estimates, especially if variables that correlate highly are omitted from the adjustments (D. Y. Lee, Harring, & Stapleton, in press). In addition, because the weighting approach relies on listwise deletion, this will also inevitably cause a loss of power if the rate of missing data is substantial.

Several imputation methods were introduced to address the issue of loss of power and precision that resulted from methods such as listwise or pairwise deletion. Imputation methods try to replace the missing data with plausible values to preserve power and improve precision by allowing the inclusion of variables that are related to the missingness into the analysis (Schafer & Graham, 2002). Several well-known single imputation methods have been popular mainly due to convenience and availability in software. Popular single imputation methods include arithmetic mean imputation (Wilks, 1932), hot-deck imputation (Scheuren, 2005), regression imputation (Buck, 1960), and stochastic regression imputation.

Arithmetic mean imputation replaces the missing values with the arithmetic mean of the variable with missing data using cases with values on the variable. Regression imputation replaces the missing values with a predicted value from a regression model using the complete data. Arithmetic mean and regression imputation are no longer recommended because although these imputation techniques preserve the mean structure of the data, they fail to preserve the variance and other higher moments of the data (Schafer & Graham, 2002), which leads to distorted measures of association and incorrect inferences. In addition, studies have demonstrated that the results from these two methods produce attenuated covariances and biased parameter estimates as the rate of missing data increases, even under the assumption of MCAR (Beale & Little, 1975).

Hot-deck imputation is popular in survey applications that was originally developed by the Census Bureau to deal with missing survey data. Hot-deck imputation is a technique where missing values are replaced by values from individuals

that are similar on other observed variables with the individual. Although several versions of the hot-deck imputation method exist, the general idea is to impute for any missing variables by randomly choosing from a sample of individuals with similar characteristics on other observed variables. Hot-deck imputation is known to preserve univariate distributions of the data and the variability of the filled-in data (Enders, 2010). Research has shown, however, that this imputation method does not perform well for estimation of measures of association such as correlations and regression coefficients (Brown, 1994; Schafer & Graham, 2002).

For the purpose of preserving variability, the stochastic regression imputation approach augments each predicted value from a regression imputation with a random normally distributed residual term. Simulation studies have shown that under assumptions of MAR, stochastic regression imputation can produce similar estimates as those from maximum likelihood and multiple imputation (Gold & Bentler, 2000). However, while estimates from a single imputation may produce unbiased estimates, standard errors are underestimated because the predicted values are treated as real data, ignoring the sampling error associated with missing data (Enders, 2010).

Another imputation approach that has gained popularity is predictive mean matching (PMM), which was introduced by Rubin (1986). PMM uses a regression imputation approach, but instead of using imputation values straight from the regression function, it uses the function to help match the observed cases to those with missing data. The values from these cases then serve as the imputations for the missing data. While Rubin's original suggestion was to use a single case for matching, later studies have shown that in order to obtain proper standard errors the number

of cases needs exceed five, essentially turning the method into a multiple imputation approach (Morris, White, & Royston, 2014). The method is especially convenient for non-normal data due to its semi-parametric approach (Horton & Lipsitz, 2001). However, the precise number of imputations needed to produce unbiased estimates and correctly sized standard errors is unknown and can vary based on sample size. This method can become especially problematic if sample sizes are small because the number of potential matching cases is limited (McNeish, 2017). In addition, the use of PMM has not been researched enough to compare its performance relative to other missing data methods. In addition, some software packages like SPSS implement the method using Rubin's traditional method of using a single imputation, which is not a recommended practice.

In light of all the advancement in the area of missing data methodology, researchers and practitioners in recent years have turned to methods that have been extensively tested, as studies have shown that ignoring MAR missing mechanisms can produce biased parameters if not addressed properly. These methods are full information maximum likelihood and multiple imputation, which are discussed in detail next.

2.1.2.2 Maximum Likelihood

Most statistical methods rely on maximum likelihood estimation for obtaining parameter estimates of a model. This involves identifying a likelihood probability function appropriate for the analysis and estimating the parameters that maximize

the probability given the observed data. In situations with missing data, estimation requires the use of intricate methods that strategically audition parameters in an iterative fashion until an optimal solution is derived. Several iterative methods have been developed for this purpose and will be discussed in subsequent sections. Complexities arise when a dataset contains missing data. In such situations, maximum likelihood estimation can still be utilized. This method of dealing with missing data is sometimes referred to as full-information maximum likelihood (FIML; [Anderson, 1957](#); [Dempster, Laird, & Rubin, 1977](#); [Lord, 1955](#)) estimation because it uses all available data without having to discard any cases that have missing data. FIML is currently regarded as the state-of-the-art for handling missing data, particularly when the preponderance of missingness occurs on an outcome variable, under the assumption of MAR ([Schafer & Graham, 2002](#)).

In the general maximum likelihood estimation framework, the likelihood of producing a pattern of data for p variables from N individuals can be expressed as

$$L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^N f_i(\mathbf{y}_{(i,comp)}|\boldsymbol{\theta}), \quad (2.4)$$

where

$$\mathbf{y}_{(i,comp)} = \{\mathbf{y}_{(i,obs)}, \mathbf{y}_{(i,mis)}\} = \{y_{i1}, y_{i2}, \dots, y_{ip}\}, \quad (2.5)$$

L is the overall likelihood function value which is scalar, $\boldsymbol{\theta}$ are the parameters to be estimated, and f_i is a probability density function for individual i . Any continuous variables with missing data, for example y_{ig} , can be integrated out of the probability

distribution to obtain the marginal probability of the individual's observed data given model parameters. That is,

$$f_i(\mathbf{y}_{(i,obs)}|\theta) = \int_{y_g} f_i(\mathbf{y}_{(i,comp)}|\theta)dy_g, \quad (2.6)$$

given the assumption that the missingness is MAR. Any discrete missing variables are summed over. Then, for example, for M individuals with completely observed data and $(N - M)$ individuals with missing data, the missing data likelihood can be expressed as

$$L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^N f_i(\mathbf{y}_{(i,comp)}|\boldsymbol{\theta}) \prod_{N+1}^M f_i(\mathbf{y}_{(i,obs)}|\boldsymbol{\theta}), \quad (2.7)$$

which is the product of the likelihoods for all the observations. In other words, given the varying patterns of observed and unobserved data in \mathbf{y}_i , the essence of full-information maximum likelihood estimation is to use each individual's complete data and to find the parameters across individuals that maximize the likelihood function.

Several methods have been proposed to obtain the values of $\boldsymbol{\theta}$ that maximize this likelihood function. Typically, parameters for models with closed solution forms and no missing data can be obtained using first derivatives. However, iterative procedures are required for complex models with large $\boldsymbol{\theta}$ vectors, missing data, and/or latent variables. The expectation maximization (EM) algorithm, which is discussed next, is one such iterative method that is especially useful for maximizing likelihood

functions when the dataset contains missing data.

2.1.2.3 Expectation Maximization Algorithm

The expectation maximization (EM; [Dempster et al., 1977](#)) optimization algorithm is a two-step procedure that iterates between an expectation-step (E-step) and maximization-step (M-step) to find optimal parameter estimates in the likelihood sense. In the most general sense, the E-step calculates a conditional expectation of the log-likelihood function¹. More specifically, if $\boldsymbol{\theta}^{(m)}$ is the m^{th} iteration value of $\boldsymbol{\theta}$ and $L(\boldsymbol{\theta}|\mathbf{y})$ is the complete-data log likelihood function, then the $(m + 1)^{\text{th}}$ iteration of the algorithm begins with the E-step, which requires the calculation of the expectation of $L(\boldsymbol{\theta}|\mathbf{y})$ conditional on $\boldsymbol{\theta}^{(m)}$ and \mathbf{y} . That is,

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) = \mathbb{E}\left[\log L(\boldsymbol{\theta}|\mathbf{y}); \mathbf{y}, \boldsymbol{\theta}^{(m)}\right]. \quad (2.8)$$

The M-step then maximizes $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$ by choosing $\boldsymbol{\theta}^{(m+1)}$ such that $Q(\boldsymbol{\theta}^{(m+1)}; \boldsymbol{\theta}^{(m)}) \geq Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(m)})$. The E and M steps continue until the difference between $L(\boldsymbol{\theta}^{(m+1)})$ and $L(\boldsymbol{\theta}^{(m)})$ is arbitrarily small.

Analysts rely heavily on maximum likelihood with the EM algorithm to estimate models using data that contain missing data, more so now than before due to it being the default estimation algorithm in several statistical packages. Several alternative methods for optimization are available, such as Newton-Raphson and Fisher scoring methods, yet none are as robust as the EM algorithm ([Arbuckle](#),

¹The log of the likelihood is used for the computational convenience of working with sums rather than products of likelihood functions

1996). It should be noted that these iterative optimization algorithms carry the risk of non-convergence to an optimum solution, and often times require unwieldy computations that may be difficult to carry out for more complex models. Typically, the class of models that fall under finite mixtures are maximized via the EM algorithm, because the latent classes to be estimated are essentially treated as another missing data variable. An alternative method to deal with missing data that is comparable to the ML method using EM that is rooted in Bayesian methodology is presented next.

2.1.2.4 Multiple Imputation

As a way to retain the advantages of single imputation methods and model sources of uncertainty mentioned before, Rubin (1987) recommended using multiple imputation (MI). MI fills in values for the missing data in three general steps:

1. creation of multiple complete datasets by imputation of plausible values in the cells where missing data are present,
2. analysis of the multiple complete datasets, and
3. aggregation of the results from the multiple analyses.

The underrepresented uncertainty of parameter estimates is addressed by creating several versions of complete data that produce multiple parameter estimates. These estimates are then averaged to produce a final parameter estimate. Special rules derived by Rubin (1987) are applied in the aggregation stage of standard errors to account for between imputation variance and within imputation variance.

The first step of MI consists of creating M datasets, with each dataset containing varying values for the cells with missing data. Once the data are filled in, the second stage consists of analyzing the M complete datasets and obtaining the estimates of the model parameters of interest. The resulting M point estimates are then averaged to obtain the final parameter estimates, $\bar{\theta}$. That is,

$$\bar{\theta} = \frac{1}{M} \sum_{m=1}^M \hat{\theta}_m, \quad (2.9)$$

where $\hat{\theta}_m$ is one of the M point estimates. Special rules proposed by [Rubin \(1987\)](#) are applied to obtain the final standard errors. These special rules take into account the within-imputation variability, V_w , and the between-imputation variability, V_b . Specifically, the standard errors are combined such that the resulting standard error, SE_T , is

$$SE_T = \sqrt{V_w + V_b + \frac{V_b}{M}}, \quad (2.10)$$

where

$$V_w = \frac{1}{M} \sum_{m=1}^M s_m^2 \quad (2.11)$$

$$V_b = \frac{1}{M-1} \sum_{m=1}^M (\hat{\theta}_m - \bar{\theta})^2, \quad (2.12)$$

and the s_m^2 term in Equation (2.11) represents the squared standard error from the analysis of dataset m .

Filling the missing data cells in step 1 is typically the most crucial step of the process and warrants further discussion. Several versions of the imputation step

have been introduced, but in general, these methods can be classified into two frameworks: joint-modeling (JM; [Rubin, 1987](#)) and fully conditional specification (FCS; [van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006](#)). While these two approaches are different in how they impute the missing data, they are both grounded in Bayesian methods. The Bayesian method behind multiple imputation is discussed next.

2.1.2.5 Bayesian Inference Behind Multiple Imputation

The imputation step of the MI procedure, whether by JM or FCS, requires the use of Bayesian inference. Bayesian inference considers parameters to be random variables with distributions that are inferred using the available data. The Bayesian method amounts to sampling these random variables from a series of probability distributions. Specifically, this consists of:

1. specifying probability distribution functions (or PDFs) for the parameters of interest [typically referred to as the prior distributions, or simply prior(s)],
2. specifying a likelihood function—a conditional PDF of the data given the parameters—that use the data to provide evidence about the parameters, and
3. creating a posterior distribution using the prior(s) and likelihood function to describe the distribution of the parameters in light of the data.

According to Bayes' theorem, the relationship among these three is expressed by

$$\Pr(\theta|Y) = \frac{\Pr(\theta) \Pr(Y|\theta)}{\Pr(Y)}, \quad (2.13)$$

where Y is the data, θ is a parameter of interest, $\Pr(\theta|Y)$ is the posterior distribution, $\Pr(\theta)$ is the prior distribution function for the parameter of interest and $\Pr(Y|\theta)$ is the likelihood function of the data given the parameters of interest, and $\Pr(Y)$ is the marginal distribution of Y . The marginal distribution Y on the right-hand side of the equation is simply a scaling constant that does not affect the shape of the posterior distribution, so this relation is typically expressed as a proportional relationship between the posterior distribution and the product of the prior and the likelihood distribution. That is,

$$\Pr(\theta|Y) \propto \Pr(\theta) \Pr(Y|\theta), \quad (2.14)$$

which is the fundamental relationship in Bayesian inference.

The manner in which the parameters for the missing data posteriors are obtained is what sets the JM approach and FCS approach apart. These two approaches are discussed next.

2.1.2.6 Joint Multivariate MI

The joint multivariate modeling imputation approach consists of a data augmentation process. During the I-step, for each missing data pattern (or unique combinations of observed and missing variables), missing values are drawn from a posterior distribution of the missing data that is conditioned on the observed data, \mathbf{Y}_{obs} , and mean vector and covariance matrix (contained in $\boldsymbol{\theta}$) of all the observed and missing variables (assuming multivariate normality). In other words, the filled-

in data at the t^{th} step in the particular I-step are random draws from a posterior predictive distribution that is conditional on the observed data and the mean vector and covariance matrix of the data drawn from the previous P-step of the missing data model. That is,

$$\mathbf{Y}_t^* \sim \Pr(\mathbf{Y}_{mis} | \mathbf{Y}_{obs}, \boldsymbol{\theta}_{t-1}^*). \quad (2.15)$$

At the P-step, using a Monte Carlo simulation, new values for $\boldsymbol{\theta}$ are drawn from respective posterior distributions of the means and covariances that are formulated using appropriate priors and likelihoods based on the filled-in data from the preceding I-step. That is,

$$\boldsymbol{\theta}_t^* \sim \Pr(\boldsymbol{\theta} | \mathbf{Y}_{obs}, \mathbf{Y}_t^*). \quad (2.16)$$

These newly drawn mean and covariance parameters contained in $\boldsymbol{\theta}$ are then used to construct the equations used to obtain the new sets of imputations in the subsequent I-step. This back and forth computation continues until the distributions for all the parameters in $\boldsymbol{\theta}$ are stationary, or have converged². The imputations created at specified steps of the entire data augmentation process are saved to make up the final set of imputations. A major assumption of using the joint multivariate approach is that the data are normally distributed, which may not always be the case.

²The meaning of convergence in Bayesian methods is discussed in detail in [Gelman and Rubin \(1992\)](#)

2.1.2.7 Fully Conditional Specification Multiple Imputation

The fully conditional specification (FCS) method is often referred to as sequential regression or chained equations. Like JM, FCS is grounded in Bayesian inference, yet the manner in which missing data is imputed is different. Mainly, while the JM-MI approach uses a joint multivariate distribution to impute the missing data, FCS uses separate univariate distributions of the variables with missing data, essentially eliminating the need for all variables to be normally distributed. In this regard, FCS has the flexibility of allowing the multiple imputation of a combination of normally distributed and non-normally distributed missing data. Variables with missing data are imputed one at a time, starting with the variable with the lowest rate of missingness. Subsequently imputed variables condition on all previously imputed variables until all the missing data has been imputed. Like in the JM approach, MCMC simulation is used to draw new sets of parameters for the conditional posterior distributions that were used to create the imputations, using the filled in values as part of the likelihood function making up the posterior from where the new parameters for the imputations are drawn from. This process continues until a specified number of times. For multivariate normal data, the FCS has been shown to be equivalent to JM ([Hughes et al., 2014](#); [Mistler & Enders, 2017](#)).

2.2 Growth Models

Several options are now available for analyzing data from longitudinal studies. In the social and behavior sciences, these studies typically use repeated measures or

panel designs to collect data from subjects at various time points, although intensive longitudinal designs are becoming more prevalent. The use of latent growth models (LGM), also known as latent curve models in the SEM literature (McArdle, 1986; Meredith & Tisak, 1990), have represent a class of models commonly used to analyze such data. Although the use of the name latent growth models and latent curve models are in some ways synonymous, the more general term latent growth modeling represent a more generalized approach and will be used from here on out. The idea behind LGMs is that subjects share the same functional form (e.g., linear, quadratic, etc.), but have distinct shapes stemming from idiosyncratic parameterizations. For example, with linear LGMs, inference will involve the estimation of individual starting points (intercepts) and growth rates (slopes). Generally, the functional form of the model is determined on theoretical grounds or derived from an exploration of the data and will dictate the number of parameters per individual.

LGMs can be regarded as two-level regression models where observations at each time within individuals are the first level, and the individuals are the second level (see, e.g. Singer & Willett, 2003). The observations at each timepoint are regarded as indicators of factors representing patterns of change. For example, to establish a linear LGM, the factors represent the intercept and slope, and the loadings are specified constants. Something to note is that in the latent curve modeling literature, loadings can also be estimated to represent other patterns (Meredith & Tisak, 1990).

For LGMs, a model for a single data point for individual i at time t , or y_{it} , is

typically specified as follows:

$$y_{it} = \eta_{0i} + \eta_{1i}\text{TIME}_t + \epsilon_{it}, \quad (2.17)$$

where η_{0i} and η_{1i} are the individual's intercept and slope growth factors, respectively, and ϵ_{it} represents the residual at time t allowing an individual's fitted function to deviate from their data. Also, the TIME_t term represents the t^{th} timepoint for individual i . The intercept and slope growth factors, η_{0i} and η_{1i} respectively, can be further expanded as

$$\eta_{0i} = \alpha_{\eta_0} + \zeta_{0i}, \quad (2.18)$$

and

$$\eta_{1i} = \alpha_{\eta_1} + \zeta_{1i}, \quad (2.19)$$

respectively. Although unspecified in Equations (2.18) and (2.19), these growth factors can easily incorporate time-invariant covariates (i.e., individual attributes). Individuals' random effects, or deviations that exist from the mean for each growth factor, are represented by ζ_{0i} for the intercept, and ζ_{1i} for the slope. These random effect terms are assumed to be multivariate normally distributed with a mean vector of zero. Combined, using matrix notation commonly used in SEM, Equations 2.17, 2.18, and 2.19 become

$$\mathbf{y}_i = \mathbf{\Lambda}_i \boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \quad (2.20)$$

where

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{\zeta}_i. \quad (2.21)$$

For T repeated measures or time points per individual where $t = 1, 2, \dots, T$, \mathbf{y}_i is a $T \times 1$ vector of measures for individual i and $\boldsymbol{\eta}_i$ is a $p \times 1$ vector of the growth factors. This notation of T implies that each individual is assumed to have the same number of timepoints and no missing data. For a linear model with two growth factors, p is equal to 2. The $\boldsymbol{\Lambda}_i$ term is a $T \times p$ matrix of factor loadings that is specified to model the desired functional form for the data for each individual. It should be mentioned that the $\boldsymbol{\Lambda}_i$ term is specific to each individual, for example, when applying this model to time-unstructured data (McNeish & Matta, 2018). By way of example, assuming that every individual follows the same functional form of growth, $\boldsymbol{\Lambda}_i$ for a linear growth model with measurements at three equally-spaced time-points would be

$$\boldsymbol{\Lambda}_i = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}. \quad (2.22)$$

The $\boldsymbol{\alpha}$ term represents a $p \times 1$ vector containing the intercepts for the growth factor (e.g., α_{η_0} and α_{η_1} in Equations 2.18 and 2.19), and $\boldsymbol{\zeta}_i$ is a $p \times 1$ vector of random effects for each growth factor. The $\boldsymbol{\epsilon}_i$ term is a $T \times 1$ individual vector of residuals across the T timepoints. Both $\boldsymbol{\zeta}_i$ and $\boldsymbol{\epsilon}_i$ are assumed to be mutually independent and multivariate normally distributed with means equal to zero and covariance matrices represented by $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}_i$, respectively. That is,

$$\boldsymbol{\zeta}_i \sim MVN(\mathbf{0}, \boldsymbol{\Psi}), \quad (2.23)$$

$$\boldsymbol{\epsilon}_i \sim MVN(\mathbf{0}, \boldsymbol{\Theta}_i), \quad (2.24)$$

where Ψ is a $p \times p$ matrix of the variances and covariances of the growth factors and Θ_i is the $T \times T$ covariance matrix of the time-specific residuals. In addition, the residuals and random effects are assumed to be uncorrelated.

The \mathbf{y}_i terms are a linear combination of normally distributed growth factors and residuals, making \mathbf{y}_i also multivariate normally distributed. That is,

$$\mathbf{y}_i \sim MVN(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \quad (2.25)$$

where the model implied mean $\boldsymbol{\mu}_i$ and covariance $\boldsymbol{\Sigma}_i$ are

$$\boldsymbol{\mu}_i = \Lambda_i \boldsymbol{\alpha}, \quad (2.26)$$

and

$$\boldsymbol{\Sigma}_i = \Lambda_i \Psi \Lambda_i' + \Theta_i. \quad (2.27)$$

In standard latent growth modeling, it is assumed that samples come from a homogeneous population. In some instances, however, multiple heterogeneous groups may exist within a population, in which case the LGM can be modified as a multiple-group model that allows groups to have unique growth patterns. Groups are typically identified by discrete covariates that are manifested in the population (e.g. gender, age group). Groups may also be theorized latent groups. Models that allow the possibility of such latent groups are called growth mixture models and will be discussed next.

2.3 Growth Models Extension to Finite Mixtures

The growth mixture model (GMM) allows for the assumption that there exists an unobserved heterogeneous population underlying the data (B. O. Muthén, 2001, 2004; B. O. Muthén & Muthén, 2000; B. O. Muthén & Shedden, 1999; Nagin, 1999). In other words, the GMM is a multiple group model but with latent heterogeneous classes in the population. This model falls under a more general category of models called finite mixture models (McLachlan & Peel, 2000; B. O. Muthén & Shedden, 1999).

Assuming multivariate normality, the combined probability density of a vector of continuous outcome variables for each individual can be expressed as

$$f(\mathbf{y}_i|\boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{k=1}^K \pi_k f_k(\mathbf{y}_i|\boldsymbol{\theta}_k), \quad (2.28)$$

where K represents the number of latent classes, $\boldsymbol{\theta}_k$ is a parameter vector (e.g., means and covariances) for each latent class, f_k represents the marginal density for class k , and the π_k term represents the proportion of the sample belonging to each class k , or otherwise known as the mixing proportion for class k , where $0 < \pi_k < 1$, and $\pi_k = 1 - \sum_{j=1}^{K-1} \pi_j$. In other words, for each class there exists a specific density function that is represented by a statistical model with unique class parameters (McLachlan & Peel, 2000).

For each class k , a linear GMM can be expressed by Equations 2.20 and 2.21

with a subscript k . That is,

$$\mathbf{y}_i = \mathbf{\Lambda}_i \boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i \quad (2.29)$$

and

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha}_k + \boldsymbol{\zeta}_i. \quad (2.30)$$

where the model-implied mean vector and covariance matrix for class k , analogous to the previously presented Equations 2.26 and 2.27 are expressed as

$$\boldsymbol{\mu}_i|k = \mathbf{\Lambda}_i \boldsymbol{\alpha}_k, \quad (2.31)$$

and

$$\boldsymbol{\Sigma}_i|k = \mathbf{\Lambda}_i \boldsymbol{\Psi}_k \mathbf{\Lambda}_i' + \boldsymbol{\Theta}_{ik}, \quad (2.32)$$

respectively. The residual terms are also class specific such that

$$\boldsymbol{\zeta}_i|k \sim N(\mathbf{0}, \boldsymbol{\Psi}_k) \quad (2.33)$$

$$\boldsymbol{\epsilon}_i|k \sim N(\mathbf{0}, \boldsymbol{\Theta}_{ik}), \quad (2.34)$$

although it is common to constrain the $\boldsymbol{\Psi}_k$ and $\boldsymbol{\Theta}_{ik}$ terms to be equal across classes, as is the case with Mplus. Class-specific covariates can also explain variation in \mathbf{y} and can also serve as predictors of the mixing proportion parameter through a multinomial logistic link function (see, e.g., [Maddala, 1983](#); [Nagin, 1999](#)).

2.3.1 Estimation of Growth Mixture Models

Although there are several approaches to estimate the parameters for the GMM, this study will focus on two approaches that are currently very popular in the GMM literature and which are also well regarded for handling missing data: Maximum Likelihood using the Expectation Maximization algorithm and Bayesian estimation methods, specifically using a Gibbs sampler. This section is devoted to introducing these two approaches.

2.3.1.1 Maximum Likelihood Estimation with the EM Algorithm

Estimation of the parameters in Equations 2.29 and 2.30 requires maximization of the log-likelihood in Equation 2.28. This can be accomplished using an iterative procedure like Newton-Raphson (Liu, Hancock, & Haring, 2011) for direct maximization, quasi-Newton or the EM algorithm (Dempster et al., 1977; McLachlan & Krishnan, 2007; McLachlan & Peel, 2000). The focus of this section will be on the EM algorithm, specifically for finite mixtures.

Using the ideas introduced in Sections 2.1.2.2 and 2.1.2.3, the log-likelihood of the mixture model to be maximized can be expressed as follows

$$\log L(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^N f(\mathbf{y}_i|\boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{i=1}^N \log \left[\sum_{k=1}^K \pi_k f_k(\mathbf{y}_i|\boldsymbol{\theta}_k) \right], \quad (2.35)$$

for N individuals with parameters defined the same as Equation 2.28. The challenge of working with this likelihood is that, in addition to model parameters, the

classification parameter, π_k , is latent and therefore completely missing. [Dempster et al. \(1977\)](#) demonstrated how this likelihood could be assumed to be similar to other missing data problems. Essentially, Equation 2.35 can be re-written with the assumption that the class parameter is observed. That is, for every individual in a sample of size N , let \mathbf{z}_i be a vector containing K zero-one binary variables, z_{ik} , indicating individual i 's disassociation/association to class k , or $\mathbf{z}_i = \{c_{i1}, c_{i2}, \dots, c_{ik}\}$ where

$$c_{ik} = \begin{cases} 1, & \text{if subject } i \text{ belongs to class } k \\ 0, & \text{otherwise} \end{cases}$$

Then, the log-likelihood with the ‘‘observed’’ class variable is:

$$\log L(\boldsymbol{\theta} | \mathbf{y}, \mathbf{z}) = \sum_{i=1}^N \sum_{k=1}^K z_{ik} \{ \log(\pi_k) + \log[f_k(\mathbf{y}_i | \boldsymbol{\theta}_k)] \}. \quad (2.36)$$

When z_{ik} is assumed to be observed, then $\pi_k = \sum_{i=1}^N z_{ik} / N$, and the maximization of the mixture model becomes less cumbersome. Missing data are also handled as explained in Section 2.1.2.2.

Maximization of the log-likelihood then becomes conditional only on the observed \mathbf{y} and candidates of $\boldsymbol{\theta}$, and the same ideas presented in Section 2.1.2.3. The E-step involves updating the values of $\boldsymbol{\theta}$ by using the conditional, ‘‘complete-data’’ expectation of the likelihood function presented in Equation 2.36. That is,

$$\begin{aligned} Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)}) &= \mathbb{E} \left[\log L(\boldsymbol{\theta}, \mathbf{z}) \mid \mathbf{y}, \boldsymbol{\theta}^{(m)} \right] \\ &= \sum_{i=1}^N \sum_{k=1}^K \mathbb{E}(c_{ik} | \mathbf{y}, \boldsymbol{\theta}^{(m)}) \{ \log(\pi_k) + \log[f_k(\mathbf{y}_i | \boldsymbol{\theta}_k)] \} \end{aligned}$$

where

$$\begin{aligned}\mathbb{E}(c_{ik}|\mathbf{y}, \boldsymbol{\theta}^{(m)}) &= \frac{\pi_k f_k(\mathbf{y}_i|\boldsymbol{\theta}^{(m)})}{\sum_{k=1}^K \pi_k f_k(\mathbf{y}_i|\boldsymbol{\theta}^{(m)})} \Bigg|_{\boldsymbol{\theta}^m} \\ &= \Pr(\boldsymbol{\theta}^{(m)}).\end{aligned}\tag{2.37}$$

Essentially, the E-step amounts to computing a posterior probability for each individual evaluated at the current parameter values at the m^{th} iteration.

The posterior probabilities are then carried over to the M-step, where they are used for maximizing the conditional expectation by replacing the unknown class indicators c_{ik} with π_k and computing other model parameters in $\boldsymbol{\theta}$. These new parameters are then carried over to iteration $(m + 1)$ starting with the E-step. The E-step and M-steps are iterated until the parameters and class proportions that maximize the likelihood satisfy the convergence criterion ([Harring, 2012](#)).

2.3.1.2 Fully Bayesian Estimation

Bayesian estimation using a MCMC simulation method is another possible approach for obtaining parameter estimates for the GMM ([Asparouhov & Muthén, 2010](#); [Gelman, Carlin, Stern, & Rubin, 1995](#)), which can also be implemented to handle missing data. Although there are a number of nuanced MCMC methods that have been developed, the Gibbs sampler is one of the more widely used methods ([Diebolt & Robert, 1994](#); [Geman & Geman, 1984](#); [B. O. Muthén, Muthén, & Asparouhov, 2010](#)). This section will discuss the specifics of the Gibbs sampler.

During the Gibbs sampler, a sequence of values for unknown parameters, la-

tent variables, and missing observations are iteratively obtained in order to create posterior distributions for the unknown parameters and missing data variables. As was discussed in Section 2.1.2.5 (presented by Equation 2.13), this is done by conditioning on the observed data and prior information and iteratively updating one parameter and missing value at a time. Parameters are continuously updated by conditioning on the priors and observed data, but also on the values from each previous iteration.

In other words, if $\boldsymbol{\theta}_i$ is a vector of all unknown parameters and missing data at iteration i , where $i = 1, 2, \dots, I$, then the parameters to be estimated, including the missing data, are separated into D groups ($d = 1, 2, \dots, D$ such that $\boldsymbol{\theta}_i = \{\boldsymbol{\theta}_{1i}, \dots, \boldsymbol{\theta}_{Di}\}$) of unknown variables that share the same prior distributions (Asparouhov & Muthén, 2010; S.-Y. Lee, 2007). At each iteration, the terms within $\boldsymbol{\theta}_i$ are updated by conditioning on the previous instance of all other elements of $\boldsymbol{\theta}_i$, the observed data (\mathbf{Y}_{obs}) and the specified prior distribution. For example, for iteration i , each element in $\boldsymbol{\theta}_i$ is updated following the sequence:

$$\begin{aligned}
& [\boldsymbol{\theta}_{1i} | \boldsymbol{\theta}_{2(i-1)}, \dots, \boldsymbol{\theta}_{(D)(i-1)}, \mathbf{Y}_{obs}, \text{Pr}(\boldsymbol{\theta})] \\
& [\boldsymbol{\theta}_{2i} | \boldsymbol{\theta}_{1i}, \boldsymbol{\theta}_{3(i-1)}, \dots, \boldsymbol{\theta}_{(D)(i-1)}, \mathbf{Y}_{obs}, \text{Pr}(\boldsymbol{\theta})] \\
& \dots \\
& [\boldsymbol{\theta}_{Di} | \boldsymbol{\theta}_{1i}, \dots, \boldsymbol{\theta}_{(D-1)(i-1)}, \mathbf{Y}_{obs}, \text{Pr}(\boldsymbol{\theta})],
\end{aligned} \tag{2.38}$$

where $\text{Pr}(\boldsymbol{\theta})$ is the prior distribution.

Specifically when addressing for any missing data, missing data values are

updated one at a time. Given the joint posterior distribution:

$$\Pr(\mathbf{Y}_{mis}, \mathbf{\Omega}, \mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\theta}^* | \mathbf{Y}_{obs})$$

where $\boldsymbol{\theta}^*$ contains mean and variance parameters, \mathbf{Y}_{mis} is the collection of missing data, $\mathbf{\Omega}$ is the matrix of latent variables, $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_i)$ is the matrix of association/disassociation variables, $\boldsymbol{\pi}$ is the vector of proportion variables, at iteration $(j + 1)$ with current observations of $\mathbf{\Omega}^{(j)}$, $\mathbf{Z}^{(j)}$, $\boldsymbol{\pi}^{(j)}$, $\boldsymbol{\theta}^{*(j)}$, $\mathbf{Y}_{mis}^{(j)}$, the sampler iterates the following steps:

1. Step a: Generate $(\mathbf{Y}_{mis}^{(j+1)}, \mathbf{\Omega}^{(j+1)}, \mathbf{Z}^{(j+1)})$ from $\Pr(\mathbf{Y}_{mis}, \mathbf{\Omega}, \mathbf{Z} | \mathbf{Y}_{obs}, \boldsymbol{\pi}, \boldsymbol{\theta}^*)$
2. Step b: Generate $(\boldsymbol{\theta}^{*(j+1)}, \boldsymbol{\pi}^{(j+1)})$ from $\Pr(\boldsymbol{\theta}^*, \boldsymbol{\pi} | \mathbf{Y}_{obs}, \mathbf{Y}_{mis}^{(j+1)}, \mathbf{\Omega}^{(j+1)}, \mathbf{Z}^{(j+1)})$

Step (a) can be further broken down. Specifically, $\mathbf{Y}_{miss}^{(j+1)}$ is generated from

$$\Pr(\mathbf{Y}_{miss} | \boldsymbol{\theta}^{*(j)}, \boldsymbol{\pi}^{(j)}, \mathbf{\Omega}^{(j+1)}, \mathbf{Z}^{(j+1)}, \mathbf{Y}_{obs}),$$

where:

$$\Pr(\mathbf{Y}_{miss} | \boldsymbol{\theta}^{*(j)}, \boldsymbol{\pi}^{(j)}, \mathbf{\Omega}^{(j+1)}, \mathbf{Z}^{(j+1)}, \mathbf{Y}_{obs}) = \prod_{i=1}^n \Pr(y_{missi} | \boldsymbol{\theta}^*, \boldsymbol{\omega}, w_i)$$

and

$$\Pr(y_{missi} | \boldsymbol{\theta}^*, \boldsymbol{\omega}, w_i) \stackrel{D}{=} N(\boldsymbol{\mu}_{miss,(ik)} + \boldsymbol{\Lambda}_{miss,(ik)} \boldsymbol{\eta}_{ik}, \boldsymbol{\Psi}_{miss,(ik)}).$$

Here, $\boldsymbol{\mu}_{miss,(ik)}$ is a $p \times 1$ mean sub-vector of $\boldsymbol{\mu}_k$ with any elements corresponding to observed components deleted, and $\boldsymbol{\Lambda}_{miss,(ik)}$ and $\boldsymbol{\Psi}_{miss,(ik)}$ are $p \times q$ sub-matrices of

Λ_k and Ψ_k , respectively, with observed components deleted (S.-Y. Lee, 2007, page 373). Similar to ML estimation, if the missingness is under the assumption of MAR, then the missingness mechanism is ignorable and should not be responsible for the bias of the parameter estimates (Little & Rubin, 2002).

Typically, several replications, or chains of the sampling are created in order to assess convergence. This is easily done by varying starting values and seeds. The separate chains and iterations continue until convergence can be determined which is assessed using the potential scale reduction factor (PSR; Gelman et al., 1995), sometimes together with graphical analysis of the chains. For multiple chains, C , where C is an integer such that $C > 1$, the PSR is a function of the within-chain variation, or

$$W = \frac{1}{C} \sum_{j=1}^C \frac{1}{I} \sum_{i=1}^I (\theta_{ij} - \bar{\theta}_{.j})^2, \quad (2.39)$$

relative to between-chain variation, or

$$B = \frac{1}{C-1} \sum_{j=1}^C (\bar{\theta}_{.j} - \bar{\theta}_{..})^2, \quad (2.40)$$

where θ_{ij} is the value of θ at iteration i in chain j , $\bar{\theta}_{.j} = \frac{1}{I} \sum_{i=1}^I \theta_{ij}$ and $\bar{\theta}_{..} = \frac{1}{C} \sum_{j=1}^C \bar{\theta}_{.j}$. Then the *PSR* is obtained by the equation

$$PSR = \sqrt{\frac{W+B}{W}}, \quad (2.41)$$

where convergence is determined by *PSR* values close to 1 or below 1.1 (Gelman et al., 1995). When this condition of convergence has been satisfied, the final posterior

distributions for all unknown variables are created from a segment of the chains, typically towards the end of all the generated θ_t s. The characteristics of the posterior distributions for each parameter, such as the mean, are then used as estimates.

The prior distribution for each parameter is usually specified by the researcher in order to add a level of certainty about the parameter. Priors can be specified to be noninformative when there is great amount of uncertainty and little information about a certain parameter. Alternatively, they can be specified to be informative when there is a high degree of certainty and information. For example, a researcher might use information from previous studies in order to incorporate a certain amount of information into the prior.

The choice of priors is an important aspect of estimating GMM parameters, especially when sample sizes per class are small or when classes are not well separated (Depaoli, 2013; Tueller & Lubke, 2010). The use of conjugate priors, or priors that mathematically lead to the same distributional family as the posterior distribution, is common (Asparouhov & Muthén, 2010; S.-Y. Lee, 2007). The Gibbs sampler in the context of GMMs involves sampling values for the class proportions, growth factor means, growth factor variances, and residual variances. In the presence of missing data, additional sampling is done for each missing data value (Distribution ??). The appropriate priors for these unknown parameters are described next.

For the class proportion parameters, a Dirichlet distribution (Dir) is an appropriate conjugate prior distribution. That is,

$$\boldsymbol{\pi}_k \sim Dir(\delta_1, \dots, \delta_K) \tag{2.42}$$

where the $\delta_1 \dots \delta_K$ are the hyperparameters, or the parameters that describe the probability distribution. For this particular distribution, the level of confidence about the class proportion parameters is specified through the δ terms, which convey the level of information and confidence about the class proportions.

A normal distribution is an appropriate conjugate prior for the growth factor means. That is,

$$\boldsymbol{\alpha}_k \sim N(\boldsymbol{\mu}_{\alpha_k}, \boldsymbol{\Psi}) \quad (2.43)$$

where the hyperparameters $\boldsymbol{\mu}_{\alpha_k}$ and $\boldsymbol{\Psi}$ represent the expectation for the factor means and the matrix of the factor variance/covariances, respectively. An appropriate prior for the $\boldsymbol{\Psi}$ term here is the inverse Wishart (IW) distribution such that

$$\boldsymbol{\Psi} \sim IW(\boldsymbol{\Omega}, \nu), \quad (2.44)$$

where hyperparameter $\boldsymbol{\Omega}$ is a $p \times p$ matrix of values greater than zero and hyperparameter ν is an integer value such that $\nu > p - 1$.

For the residual variances, the appropriate conjugate prior is the inverse gamma (IG) distribution such that if ω_{tt} represents a single cell in the $T \times T$ square residual variance matrix, $\boldsymbol{\Theta}$, then

$$\omega_{tt} \sim IG(a_{\omega_{tt}}, b_{\omega_{tt}}), \quad (2.45)$$

with hyperparameters $a_{\omega_{tt}}$ and $b_{\omega_{tt}}$ controlling the shape and scale of the distribution to manipulate the level of informativeness about the each element of the residual variance matrix. In the context of GMMs, Bayesian methods with varying prior

specifications have been shown to recover parameter estimates quite well under specific conditions such as certain sample sizes, class proportions, and specifications of priors.

Next, a brief discussion on how all these different missing data handling methods that were presented have fared with GMMs is presented.

2.4 Handling Missing Data with Growth Mixture Models

2.4.1 Maximum Likelihood

Missing data in the growth mixture modeling context can be problematic if not handled correctly, even when the missing data are assumed to be ignorable. Using a multiple group growth model, [Enders and Gottschall \(2011\)](#) found that parameter estimates attenuated group differences when the grouping variable was not considered as part of the imputation, even when the missingness was MCAR because any group differences in the means or variances were ignored during the data imputation. Their suggestion when it came to finite mixture models was to use FIML-EM instead, especially since the grouping variable is always unobserved for such models.

As [Sterba \(2016\)](#) pointed out, the use of FIML-EM can be problematic when covariate information is missing. Models that incorporate covariate information require the use of conditional models. These conditional models are typically fit using a conditional likelihood that require complete covariate information. Otherwise, standard statistical software such as Mplus that have the capability to run such models

will listwise delete data with missing covariates by default. The purpose of the study by [Sterba \(2016\)](#) was precisely to show how to use Mplus to prevent listwise deletion of data by using a joint likelihood approach, whereby in addition to estimating the parameters of the conditional model, the parameters of the covariates' marginal distribution are also estimated. Without this workaround, however, if the missingness is based on any of the covariates (MAR mechanism), then the listwise deletion of these cases could potentially bias parameter estimates.

Although studies seem to indicate that FIML-EM is a viable solution for the missing data problem with GMMs, some still consider MI as a possible solution for addressing missing data in the context of finite mixtures—particularly when missingness occurs on covariate information. While the literature on this topic is very limited, a considerable amount of research has been put forth in the past 20 years by a select few who have suggested using a two-stage multiple imputation approach ([Harel, 2007](#); [Harel & Schafer, 2009](#); [Rubin, 2003](#); [Shen, 2000](#)). The two-stage multiple imputation approach, which will be discussed in [Section 2.4.4](#), considers the class variable to be a second stage of the imputation process. The studies that consider the two-stage imputation vouch for the method under conditions of MAR. However, not much research has been conducted in this particular area and none of the simulation studies on this approach considered different conditions that are typically investigated in the context of finite mixture models.

In addition to using FIML-EM and two-stage imputation, Bayesian methods have been explored in the area of growth mixture modeling, and some studies have also considered methods for addressing missing data in these contexts. These studies

and their findings are presented next.

2.4.2 Bayesian Methods

In the GMM context, missing data can be handled using a Gibbs sampler, as was explained in Section 2.3.1.2. Using a Bayesian approach to growth mixture models has been shown to produce favorable results. In a simulation study by [Depaoli \(2013\)](#), parameter recovery for a three class growth mixture model was investigated under varying degrees of class separation, sample size, class proportion, and method of analysis (FIML vs. fully Bayesian approach using different priors). The study showed that a Bayesian framework with informative priors produced less biased estimates of the means (slopes and intercepts) than when using maximum likelihood estimation, particularly when class separations were larger (Mahalanobis distances greater than 1.0) and samples consisted of as little as 150 individuals. Estimates of the variance parameters, however, were not well recovered under all conditions. One limitation of the study, however, was that missing data was not considered.

Investigations of GMMs under varying missing data conditions have also been conducted. In a simulation study, [Lu et al. \(2011\)](#) investigated the use of a fully Bayesian approach for estimating GMMs when different types of missingness were introduced in the analysis. Considering sample size, class probability, rate of missing, and missing data mechanisms MNAR and MCAR, the study concluded that Bayesian estimation generally performed well but only under sample sizes conditions of 1000 or more. In addition, their missing rates were in the range of 5% to 20%,

and sample sizes below 500 were not considered, although they made the claim that a Bayesian method is useful as long as missing data rates stay below 5% for small sample sizes. This claim however, was not substantiated by any past research nor as part of their simulation study.

Other studies have also investigated the use of Bayesian methods with missing data, but none have considered different missing data mechanisms in the context of GMMs with various conditions like the two aforementioned studies. For example, [S.-Y. Lee and Tang \(2006\)](#) investigated the use Gibbs sampling for parameter recovery of a factor analytic model with non-ignorable missing data, which was essentially a replication of work by [S.-Y. Lee and Song \(2003\)](#), with the exception that MNAR missingness was considered. In particular, their analysis focused on the choice of priors for these models, concluding that the choice of priors did not make much difference. The conditions they manipulated, however, were very narrow in scope, and more importantly did not consider the existence of multiple latent classes.

In general, while there exist studies which have considered the use of Bayesian methods for GMMs, none have conducted a systematic investigation on the use of Bayesian methods with missing data under varying conditions that are pertinent to the study of both GMMs and missing data.

2.4.3 Single-Stage Multiple Imputation

Aside from the study by [Enders and Gottschall \(2011\)](#), which cautioned against using MI for general mixture models, the use of single-stage MI has not been thor-

oughly investigated. A few studies have, however, investigated the use of MI on longitudinal models. For example, [Huque, Carlin, Simpson, and Lee \(2018\)](#) investigated the use of different JM and FCS methods specifically designed to accommodate for missing time-dependent covariates measured at irregular time intervals. Their study found that the standard JM and FCS methods performed adequately for longitudinal studies, and that they were only really useful when data were irregularly spaced. [Schafer and Yucel \(2002\)](#) and [Asparouhov and Muthén \(2010\)](#) also introduced nuanced versions of the JM imputation method, specifically designed for mixed effects multilevel models. [van Buuren \(2011\)](#) later introduced an analogous method using FCS and [Mistler and Enders \(2017\)](#) also compared nuanced versions of the JM and FCS methods specifically for multilevel modeling. Although such studies have considered these missing data handling methods in the context of growth models, none conducted a systematic investigation of the performance of MI in the context of growth mixture models.

2.4.4 Two-Stage Multiple Imputation

Another suggested method for addressing the missing data problem in GMMs has been brought to attention by [Harel \(2003\)](#) and [Harel and Schafer \(2009\)](#), who proposed using a two-stage imputation method. The two-stage imputation method is an extension of the nested MI approach proposed by [Shen \(2000\)](#) and then applied in survey data by [Rubin \(2003\)](#). This method essentially groups the missing data, Y_{mis} , into qualitatively different types, allowing for a separate assessment of contribution

of each type of missing data to the overall uncertainty. The crux of the two-stage imputation is in the assumption of extended ignorability as described in detail by [Harel \(2003\)](#) and [Harel and Schafer \(2009\)](#). Recalling from the ignorability conditions presented by Equation 2.1, if the missing data, Y_{mis} , is regarded as the decomposition of two types of missing data, or $Y_{mis} = (Y_{mis}^A, Y_{mis}^B)$, and the missing data indicator \mathbf{I}^+ represents the missingness indicator matrix for the two types of missing data, then the missingness is regarded to be MAR⁺ if

$$\Pr(\mathbf{I}^+ | \mathbf{Y}_{obs}, \mathbf{Y}_{miss}^A, \mathbf{Y}_{miss}^B, \boldsymbol{\phi}^+) = \Pr(\mathbf{I}^+ | \mathbf{Y}_{obs}, \boldsymbol{\phi}^+). \quad (2.46)$$

According to [Rubin \(1987\)](#), under assumptions of traditional MAR, imputations for any missing data can be drawn from $\Pr(\mathbf{Y}_{miss} | \mathbf{Y}_{obs})$, meaning that \mathbf{I} can be ignored. Analogously, [Harel \(2003\)](#) shows that \mathbf{I}^+ can be ignored and the following relationships become true:

$$\Pr(\mathbf{Y}_{miss}^A | \mathbf{Y}_{obs}, \mathbf{I}^+) = \Pr(\mathbf{Y}_{miss}^A | \mathbf{Y}_{obs}), \quad (2.47)$$

and

$$\Pr(\mathbf{Y}_{miss}^B | \mathbf{Y}_{obs}, \mathbf{I}^+) = \Pr(\mathbf{Y}_{miss}^B | \mathbf{Y}_{obs}, \mathbf{Y}_{miss}^A). \quad (2.48)$$

Proofs of these relationships are available in ([Harel, 2003](#), p. 37). A weaker condition based on these relationships that are pertinent to mixture models, where the missingness in the manifest variables might be conditioned on the completely unobserved latent class variable (e.g., missing manifest variable, \mathbf{Y}_{miss}^B , is dependent on

completely unobserved latent class variable, \mathbf{Y}_{miss}^A), is the conditional extended ignorability condition. Harel defines a conditional extended ignorability assumption, or CMAR⁺, where \mathbf{I}^+ is conditional on the observed data, the second group of missing data, and the parameter relating these variables. That is,

$$\Pr(\mathbf{I}^+ | \mathbf{Y}_{obs}, \mathbf{Y}_{miss}^A, \mathbf{Y}_{miss}^B, \phi^+) = \Pr(\mathbf{I}^+ | \mathbf{Y}_{obs}, \mathbf{Y}_{miss}^A, \phi^+). \quad (2.49)$$

If the CMAR⁺ condition can be assumed, then the following relationship is also true:

$$\Pr(\mathbf{Y}_{miss}^B | \mathbf{Y}_{obs}, \mathbf{Y}_{miss}^A, \mathbf{I}^+) = \Pr(\mathbf{Y}_{miss}^B | \mathbf{Y}_{obs}, \mathbf{Y}_{miss}^A), \quad (2.50)$$

which, under Rubin's original assumptions of MAR, by which imputations can be drawn from $\Pr(\mathbf{Y}_{miss} | \mathbf{Y}_{obs})$, imputations for a second set of missing data can be drawn from $\Pr(\mathbf{Y}_{miss}^B | \mathbf{Y}_{obs}, \mathbf{Y}_{miss}^A)$ as long as the mechanism that made \mathbf{Y}_{miss}^B is not related to \mathbf{Y}_{miss}^A . For example, if \mathbf{Y}_{miss}^A represents the missing class indicators (completely unobserved) and \mathbf{Y}_{miss}^B represents the missing values in the manifest data, if the mechanism that created \mathbf{Y}_{miss}^B is not dependant on \mathbf{Y}_{miss}^A or \mathbf{Y}_{miss}^B , then it is considered MAR⁺. If the mechanism that created \mathbf{Y}_{miss}^B depends only on \mathbf{Y}_{miss}^A and possibly other observed manifest variables, a possibility in latent class models where the missingness can depend on the latent class, then this situation constitutes CMAR⁺ according to [Harel \(2003, p. 51\)](#).

In the context of a model with latent classes, [Harel and Schafer \(2009\)](#) regard the latent indicators of class as MCAR, and require that missing data in the manifest

variables be at worst MAR. To reiterate, however, although the manifest variables may be conditional on the completely unobserved latent class, which by traditional definition would constitute MNAR, it is considered to be CMAR⁺, and therefore ignorable. According to this, MAR can be assumed and MI should work at both stages, first when latent classes are imputed, and then after when the manifest variables are imputed for each imputed latent class. If the manifest variables for some reason were considered to be MNAR, then the MNAR mechanism would need to be addressed as well.

In the two-stage MI method, m sets of Y_{mis}^A , or class indicators, are first generated and retained from an MCMC chain in the I-step. That is,

$$Y_{mis}^{A(j)} \sim \Pr(Y_{mis}^A | Y_{obs}), \quad (2.51)$$

where $j = 1, 2, \dots, m$. Then, for each m sets of the latent classes, assuming that the latent classes are fixed as known, $n - 1$ sets of Y_{mis}^B are generated. That is,

$$Y_{mis}^{B(j,k)} \sim \Pr(Y_{mis}^B | Y_{obs}, Y_{mis}^{A(j)}), \quad (2.52)$$

where $k = 1, 2, \dots, n$. Here, there $n - 1$ sets because the first set is generated from first stage of imputations. This will result in a total of mn complete datasets that are to be analyzed.

After analysis, the resulting mn parameter estimates are averaged to obtain

the final estimates. That is,

$$\bar{\theta}_{..} = \frac{1}{mn} \sum_{j=1}^m \sum_{k=1}^n \hat{\theta}^{(j,k)}, \quad (2.53)$$

where $\hat{\theta}^{(j,k)}$ is the estimated parameter for stage 1 imputed data set j , for stage 2 imputed data set k .

Standard errors are combined using a nuanced version of Rubin's rules derived by Shen (2000). Analogous to Rubin's rules, Shen's rules formulate the resulting standard errors, SE_T^{2stg} , to be a function of the within-imputation variability, V_w^{2stg} , the between-imputation variability, V_b^{2stg} , and a third component: the complete-data variance, $\hat{U}_{..}$. Specifically, if $U^{(j,k)}$ is the estimated squared standard error of the parameter estimate, then

$$SE_T^{2stg} = \sqrt{\hat{U}_{..} + \left(1 - \frac{1}{n}\right)V_w^{2stg} + \left(1 + \frac{1}{m}\right)V_b^{2stg}}, \quad (2.54)$$

where the complete-data variance is

$$\hat{U}_{..} = \frac{1}{mn} \sum_{j=1}^m \sum_{k=1}^n U^{(j,k)}, \quad (2.55)$$

the between-block imputation variance is

$$V_b^{2stg} = \frac{1}{m-1} \sum_{j=1}^m (\bar{\theta}_j - \bar{\theta}_{..})^2, \quad (2.56)$$

and the within-block imputation variance is

$$V_w^{2stg} = \frac{1}{m} \sum_{j=1}^m \frac{1}{n-1} \sum_{k=1}^n (\hat{\theta}_{jk} - \bar{\theta}_j)^2. \quad (2.57)$$

The two-stage imputation method has been mainly applied to empirical data in situations involving missing data. For example, it has been applied when the missing data can be thought of as being of distinct types of missingness, such as surveys containing planned and unplanned missing data ([Graham, Taylor, & Cum-sille, 2001](#)), longitudinal models with ignorable intermittent missing data points ([Hedeker & Gibbons, 1997](#)) and non-ignorable dropout missing ([Harel, 2003](#)). The method has also been applied with latent variable models that involve analysis with missing data and unobserved latent variables such as latent class regression models ([Harel, Chung, & Miglioretti, 2013](#)) and latent class contingency tables [Winship, Mare, and Warren \(2002\)](#).

Despite finding studies that used the two-stage imputation method to address missing data, there were no studies that tested the method through simulation on different conditions of missingness or GMMs. Thus, while the method has been theoretically justified and applied in several different scenarios, the extent to which the method will work and how well it will perform relative to other missing data handling methods such as FIML, single imputation, or fully Bayesian methods is unclear.

2.4.5 Summary: Comparison of Methods

As was mentioned previously, [Enders \(2010\)](#) and [Enders and Gottschall \(2011\)](#) have strongly cautioned against using MI and have suggested using FIML-EM for finite mixtures in the presence of missing data. While this is the general recommendation, aside from the study by [Enders and Gottschall \(2011\)](#) and [Depaoli \(2013\)](#), there were no other studies that systematically investigated the performance for FIML-EM on GMMs. [Lu et al. \(2011\)](#) demonstrated evidence through simulation that fully Bayesian methods are another viable option for handling missing data in the context of GMMs. The research was limited however, because the conditions tested were limited. [Depaoli \(2013\)](#) also investigated the estimation of GMMs under a Bayesian framework with varying conditions, but did not investigate how missing data would effect the results. The two-stage imputation method seems to be a viable solution to the single multiple imputation problem and has been used for several missing data problems involving latent classes, but none of the studies investigated the method on GMMs with missing data, whether by simulation or empirical data analysis.

In light of of these studies, there is an evident gap in the literature regarding GMMs and missing data. Therefore, the purpose of the current study is to fill this gap by testing how the different methods for addressing missing data, as suggested by the current available research studies, perform on a growth mixture model.

2.5 Research Goals

The main purpose of this study is to extend the use of these proposed methods: FIML-EM, fully Bayesian using Gibbs sampling, and a two-stage MI approach, to handle missing data in view of modeling conditions that are common in the growth mixture modeling and missing data literature. In many GMM simulation studies, these conditions include sample size, number of classes, and latent class separation. In many missing data studies, these conditions include rate of missing and missingness mechanism. Other important considerations that have been the focus of studies on GMMs such as class enumeration and fit evaluation criteria (see, e.g., [Bauer, 2007](#); [Bauer & Curran, 2003](#)) will not be considered, which are considered one of many limitation of the study. These factors will not be explored because the intention of the current study is to focus on the effects of missingness on parameter estimation. Specifically, the current research will compare the performance of list-wise deletion, FIML, single-stage MI, two-stage MI, and a fully Bayesian approach for estimating growth mixture models when there is missing data at various occasions under varying conditions commonly tested in the GMM literature as well as different missing data conditions that may have an influence on how these methods perform.

The specific research questions that this study would like to address are as follows:

1. For what rate of missing data and class separations will the GMM produce biased parameter estimates when utilizing the different missing data handling methods under conditions of MAR (or CMAR⁺) missingness and MCAR miss-

ingness?

2. When the goal is to classify individuals into groups, for which method will the growth model provide the highest percentage of classification accuracy for varying rates of missing data, sample size, class separation and missingness mechanism?
3. Are there any advantages to using any of the missing data handling methods in the context of GMMs? Particularly, is using a more complicated method worth the extra effort for producing unbiased parameter estimates?
4. In addition to bias, which missing data handling methods produce standard error estimates that are most accurate?

In order to derive at answers to these questions, a Monte Carlo simulation study will be conducted, details of which will be laid out in Chapter 3. Chapter 4 will then focus on the results from the simulation study outlined in Chapter 3, and Chapter 5 will conclude the study with a discussion on the results, limitations of the study, and possible future extensions to the current study.

Chapter 3: Methods

A Monte Carlo simulation method will be used to compare the differences between the varying methods available for handling missing data in the context of growth mixture models. Monte Carlo simulation involves the comparison of different methods of data analysis using known population parameters as the basis of comparison. Conditions that are known to impact the parameter recovery of the population parameters include: missingness mechanism, missing rate, sample size, class proportion, and class separation. The simulation will involving the following steps:

Step 1: A series of complete datasets will be generated using a set of known parameters. The desired sample size, class proportion, and class separation conditions will be achieved at this step.

Step 2: Each dataset will be made incomplete, a process called data amputation, at various percentages of missingness, following rules of MNAR (or CMAR+) and MCAR missing data mechanisms.

Step 3: The incomplete datasets will be analyzed using six missing data handling approaches: Listwise deletion, FIML-EM, MI, two-stage MI, and Bayesian

via Gibbs sampling.

Step 4: Parameter estimates produced from each approach for all conditions will be compared with the parameters from a complete data analysis (analysis on the data prior to incorporating any missing data conditions) and evaluated on the relative bias of parameter estimates, accuracy of the standard errors of the estimates, classification accuracy, and convergence rates.

Each section in this chapter will provide details on each of these four steps. The chapter will conclude with a summary of the expected outcomes from the simulation.

3.1 Step 1: Data Generation

The first part of the data simulation will involve generating data from a two-class growth model with random intercepts and slopes. This will be done by parameterizing a two-group factor analytic model such that the loadings will be fixed to model linear growth, as was described in Section 2.2, and varying the means of the latent growth factors for the two classes to achieve varying degrees of class separation. Latent growth factor variances and covariances will be invariant for the two classes. The degree of separation between the two classes will be further discussed in Section 3.1.1. Although simulation studies in the past have investigated other issues with more than two classes, this number was chosen in order to keep the study focused on the missing data handling issue.

The number of timepoints or repeated measures will also be based on previous research that have investigated growth modeling and GMMs through simulation. A

general recommendation by [Willett, Singer, and Martin \(1998\)](#) and [Vickers \(2003\)](#) is to use at least three timepoints for linear growth models, although [B. O. Muthén and Curran \(1997\)](#) showed how the required sample size to achieve the same amount of power sharply dropped when going from three to four timepoints, with a minimal drop observed after adding more than four timepoints. Other simulation studies exploring the use of GMMs by Monte Carlo simulation also used at least four timepoints (see e.g., [Depaoli, 2013](#); [Lu et al., 2011](#); [McNeish & Harring, 2017](#); [Nylund et al., 2007](#); [Tofighi & Enders, 2007](#)). For this reason, the current study will use a population model with four timepoints.

Sample sizes for each class and separation between classes will be manipulated during the data generation. Details on how these conditions will be chosen and manipulated are presented next.

3.1.1 Class Separation

Latent class separation has been shown to have a large effect on the estimation accuracy of GMMs. For GMMs, class separation is often regarded as the amount of overlap between latent class growth trajectories, where the greater the overlap between classes, the closer the distance between classes, and the more difficult it is to differentiate between latent classes. The specific conditions to manipulate class separation will be based on previous studies that have investigated realistic measures of separation typically found in empirical data.

Class separation can be defined in several different ways. For example, [Tofighi](#)

and Enders (2007) achieved varying levels of separation by specifying different within-class variance parameters across latent classes. Other studies have defined class separation through differences in the latent growth factors (Depaoli, 2013; Li, Chen, Cui, & Liu, 2017; Lubke & Neale, 2006; Tueller & Lubke, 2010), although Depaoli (2013) showed that this is equivalent to differences in the manifest repeated measure variables. Following the suggestion of these studies, the current study will generate data from subpopulations with different latent growth factors to achieve varying degrees of class separation as defined by the Mahalanobis distance.

Assuming that the class variance/covariance matrices are equal, the Mahalanobis distance (MD; Mahalanobis, 1936) is a measure used to quantify class separation [although measures of distance when covariance matrices are not assumed equal are also available (Anderson & Bahadur, 1962)]. For the current study, the MD will be used as a measure of class separation, and as a result, it will be assumed that the covariances of class growth parameters between classes are equal. The MD is defined as follows:

$$\text{MD} = [(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})]^{1/2}, \quad (3.1)$$

where the $\boldsymbol{\mu}$ terms represent the means of the growth factors (intercept means and slope means of the two classes) and the $\boldsymbol{\Sigma}^{-1}$ represents the inverse of the common variance/covariance matrix of the growth parameters for all classes. The MD distance can therefore be manipulated by varying the values of the population latent growth intercept and slope means for each class. Previous studies have used MDs

as low as 0.5 to represent very poor separation, MDs as high as 2.0 to represent high separation, and MDs of 1.0 and 1.5 to represent poor and moderate separation (see, e.g. [Depaoli, 2013](#); [Lubke & Muthén, 2007](#); [Lubke & Neale, 2006](#); [Tueller & Lubke, 2010](#)). These same distances will be used for the current study.

In order to achieve the four MD conditions, the class 1 parameters will remain fixed while the class 2 parameters will be changed to achieve the desired MD. The class 1 parameters will be fixed based on parameters that were used in [Kaplan \(2002\)](#), which investigated trajectories of reading development among kindergartners and first graders. These parameters were also used in a simulation study by [Depaoli \(2013\)](#). Growth parameter variances and covariances, as well as time-specific residuals will remain fixed. Population values that will be used for the data generation for each distance condition are presented in [Table 3.1](#). In addition to population parameter values, the population class 1 and class 2 growth curves can be seen in [Figure 3.1](#).

Table 3.1

Data Generating Parameters for Varying MD Conditions

Parameter Means	High separation (MD \approx 2.0)	Moderate separation (MD \approx 1.5)	Poor separation (MD \approx 1.0)	Very poor separation (MD \approx 0.5)
Class 1 Intercept	48.00	48.00	48.00	48.00
Class 1 Slope	3.00	3.00	3.00	3.00
Class 2 Intercept	40.82	43.10	45.66	47.34
Class 2 Slope	4.00	4.00	4.00	3.60

$$\text{Growth parameter variances } (\Psi) = \begin{bmatrix} 18.00 & \\ & 2.00 \end{bmatrix}$$

$$\text{Time-specific residuals } (\Theta) = 15\mathbf{I}_4$$

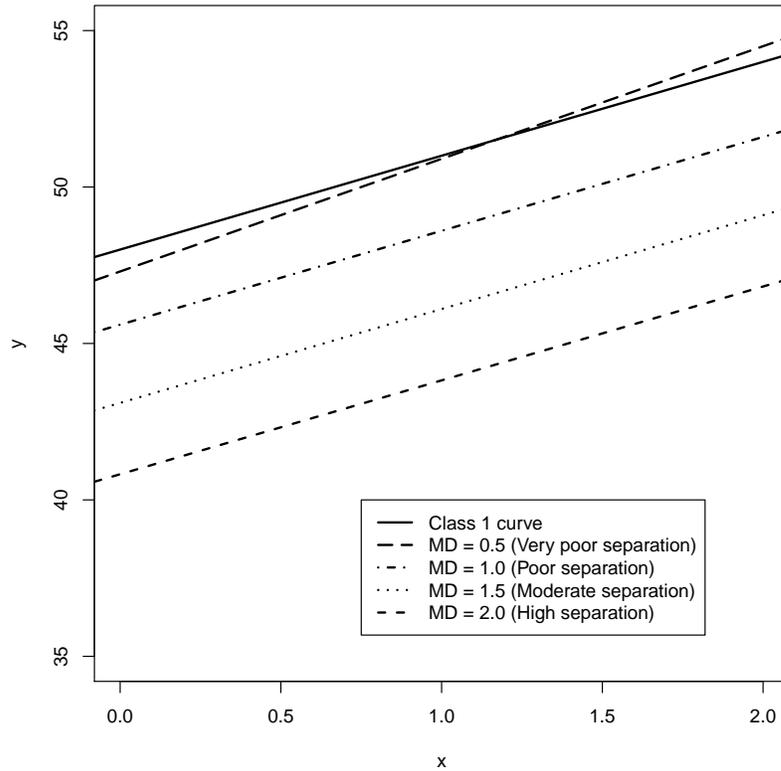


Figure 3.1. True population class 1 and class 2 growth curves for varying degrees of separation.

3.1.2 Sample Size

The sample sizes that will be investigated for the current study will come from previous simulation studies. Simulation studies involving growth modeling have demonstrated that sample size interacts closely with class separation and class proportion to achieve varying degrees of successful parameter recovery (see, e.g. [Asparouhov & Muthén, 2010](#); [Depaoli, 2013](#); [Lubke & Neale, 2006](#); [Tueller & Lubke, 2010](#)). Generally, the findings from these studies suggest that lower sample sizes require larger distances and closer to equal class proportions in order to properly identify classes and achieve acceptable levels of bias. Most of these studies, however, have not considered the impact of their results when there is missing data.

[Depaoli \(2013\)](#) used sample sizes of 150 and 800 with four levels of class separation for linear growth mixture models and generally found better estimates with sample sizes of 800 when separation was poor. The study by [Tueller and Lubke \(2010\)](#) found that convergence and parameter estimates were consistently acceptable when sample sizes were greater than 300 together with class separation greater than $MD = 1.5$ and generally performed well with sample sizes of 100 but separation greater than $MD = 2.0$ or sample sizes of 1000 with separation at $MD = 0.5$. Another similar study by [Lubke and Neale \(2006\)](#) investigated small sample sizes, finding that sample sizes as small as 75 per class were needed even when the distance between class was small ($MD = 1.5$) and as small as 25 per class were needed for correct model detection when separation was large ($MD = 3$). In light of these studies, given that similar MD distances will be used to generate data for the

different classes, total sample sizes of 100, 200, 500, and 1000 will be used for the current study, and class proportions will be kept equal.

3.2 Step 2: Data Amputation

The second step following data generation will be to systematically delete data, a process otherwise known as data amputation. The amputation process will be carried out so that the desired missing rate and strength of correlation with variables that cause the missingness can be achieved all whilst adhering to the desired missingness mechanism.

3.2.1 MNAR Missingness (CMAR⁺)

To simulate a MNAR missingness mechanism, or the CMAR⁺ according to [Harel \(2003\)](#), the probability that individual i will have data point y_{it} at time t missing will be calculated based on the following logistic function:

$$\Pr\{R_{it} = 1\} = \frac{\exp(m)}{1 + \exp(m)}. \tag{3.2}$$

R_{it} is the missingness indicator, where if $R_{it} = 1$ then the individual i 's observation at time t is missing. Also, $m = \beta Z_i + c$, where Z_i represents the unobserved class variable, β is the coefficient used to control the relation between class and the propensity to be missing at timepoint 1 and c will be used to control the rate of missing. Presented in [Table 3.2](#) are the coefficients that will be used to create the desired rates of missing and correlation from the class variable.

Table 3.2

Coefficients for Creating Missing Data

miss. rate	ρ	β	c
5%	0.25	4.0	2.0
10%	0.34	4.0	1.3
20%	0.36	2.0	0.5
30%	0.37	1.7	-0.1
40%	0.36	1.5	-0.6

Note. The correlation term ρ here is an approximation of the point-biserial correlation typically used to quantify the relationship between a continuous variable (missingness at timepoint 1) and a dichotomous variable (class).

3.2.2 MCAR Missingness

For each outcome variable (outcome at each timepoint), a randomly drawn value from 0 to 1 from a uniform distribution will be compared against a specified threshold that will create an overall rate of missing. For example, to achieve a 20% rate of missing, the threshold will be 0.20, where if the randomly generated value is below 0.20, the value will be amputated and otherwise left as is (not missing). In this way, the missingness is completely haphazard and adheres to the rules of MCAR missingness.

Finally, the missing rate for each sample needs to be specified. Since the purpose of this study is to find the tolerable amount of missing data before it is no longer possible to use GMMs, low percentages of missing data from 5% up to 40% will be evaluated. Specifically, percentages of 5, 10, 20, 30, and up to 40% will be investigated.

3.3 Step 3: Missing Data Handling Approaches

After the data has been amputated, the six different missing data handling approaches described in Chapter 2 will be implemented to analyze the data. To summarize, the methods that will be tested are:

1. Listwise deletion (LD)
2. Full-Information Maximum likelihood (FI)
3. Multiple imputation (MI)
4. Two-stage multiple imputation (2M)
5. Fully Bayesian approach (FB)

There are several potential issues that may arise during the simulation study that must be addressed. For example, label switching is a common problem when conducting MCMC simulations dealing with GMMs. This issue is especially pertinent to any methods using Bayesian methods. Convergence is another issue that must be addressed when dealing with GMMs. These issues will be the focus of the following sub-sections.

3.3.1 Label Switching

Previous research indicates that label-switching can be problematic for finite mixture models (Tueller, Drotar, & Lubke, 2011). Label switching occurs when

the order of the latent classes arbitrarily "switch" during estimation. When using Bayesian methods, including multiple imputation, there are two different places where the label switching can occur. First, label switching can occur arbitrarily during the MCMC chain (i.e., within-chain label switching). Second, it can occur across chains (i.e., between-chain label switching). In both these situations, label switching can distort the final estimates and complicate the assessment of convergence.

The issue of label switching has been studied by [Celeux, Hurn, and Robert \(2000\)](#), and several ways to prevent it have been suggested. A common approach that is used to avoid within-chain label switching, which will be used in this study, is to constrain the prior for the intercept of one class to be greater than the intercept of the other. This will then prevent the Gibbs sampler from switching mid-chain. Between-chain label switching will be avoided by using a single chain. It should be noted that when a single chain is used, convergence can be assessed by taking the PSR of the values towards the end of the chain ([Asparouhov & Muthén, 2010](#)).

For non-Bayesian methods such as FIML-EM, label switching is possible between replications. To avoid this issue, means of the growth factors will be constrained so that the mean for class 1 is always greater than the mean for class 2. This method is commonly used in simulation studies involving GMMs to avoid between replication label switching.

3.3.2 Convergence Issues

Simulation studies involving finite mixture models are often susceptible to convergence issues, whether by FIML-EM or MCMC methods. For example, [Tueller and Lubke \(2010\)](#) explained that nonconvergence can be a cause of several factors such as model misspecification, model nonidentification, insufficient starting values, or characteristics of the data such as multivariate outliers. Local maxima solutions also pose a threat to solutions produced by FIML-EM methods. For Bayesian MCMC methods, label switching within and between chains will also cause problems in the assessment of convergence using the PSR as well as problems with the estimation of parameters.

For FIML-EM methods, to avoid local maxima solutions, [Tueller and Lubke \(2010\)](#) recommend replicating the final likelihood by varying the starting values and setting a maximum number of replicated likelihoods. In order to implement this suggestion, 100 replications with perturbed starting values will be used, and a minimum criteria of 10 optimized likelihoods will be used as a way to show convergence to a global maximum.

For the MCMC methods, convergence will be assessed using the PSR factor, where a PSR of 1 with a tolerance of ± 0.05 will be considered converged. As was discussed in the previous section, label switching will be accounted for by specifying certain constraints within a chain and creating one single chain to avoid any between chain label switching.

Low convergence rates can compromise the results of the simulation study

because a low number of simulation replications will not represent a true sampling distribution of the population parameters. It is expected that some conditions, such as ones with low sample sizes, high percentages of missing data, and low separation will experience higher rates of non-convergence. Such issues will be identified so that any replications with convergence problems will be skipped until enough replications are collected to satisfy the desired number of replications. Also, the total number of replications needed to satisfy the minimum number of replications will be recorded in order to report on an overall convergence rate. These convergence rates will provide evidence of the findings from the simulation study.

3.3.3 Number of Imputations

For the multiple imputation approach, there are several specifications that must be made. One of these is the choice of number of imputations. It has been suggested that the number imputations for single-stage MI be set at $M = 40$ in order to avoid any power issues ([Graham, 2009](#)). [McGinniss and Harel \(2016\)](#) showed that under MAR conditions and smaller rates of missing information (less than 40%), increasing the number of imputations for multiple-stage MI models did not produce significantly different bias or coverage rates among varying combinations of number of imputations at each stage, especially for problems involving estimation of point estimates and variances. Therefore, they recommended using as few as $N = 10$ imputations in the first stage and $M = 2$ imputations in the second stage. To be conservative, for the two-stage MI approach, $N = 20$ imputations will be used for

the first stage and $M = 4$ imputations will be used for the second stage for a total of 80 imputations per dataset.

3.3.4 Choice of Priors

Using a Bayesian approach in the context of GMMs requires a careful consideration of the types of priors that will be used, especially when sample sizes are small since the influence of priors becomes greater as sample size decreases. In certain situations, variances of growth factors are especially susceptible to prior misspecification (Depaoli, 2013; McNeish, 2016). To investigate the kind of priors that should be used for the growth model in the current simulation study, a preliminary simulation will be conducted on the full data (without any of the missing data conditions) to see how different priors will affect variance parameter estimation. The simulation will involve comparing maximum likelihood estimation using the EM algorithm, and a Bayesian MCMC method using prior specifications for the variance parameters as suggested by McNeish (2016) for latent growth models. The prior specifications that will be compared are as follows:

1. The default Mplus improper inverse Wishart distribution prior for the variance parameters, or $IW(\mathbf{0}, d = -p - 1)$, where the first argument represents a scale matrix, the second argument represents the degrees of freedom, and p represents the number of elements in the diagonal of the factor covariance matrix. For the current study, the priors for the elements of the variance covariance matrix will be specified as $IW\left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, -3\right)$.

2. An inverse Wishart distribution for the variance covariance matrix, where the first argument is the identity matrix, \mathbf{I} , and the second argument is p . This is called the proper inverse prior, which was previously recommended by [Chen \(2011\)](#), [Lunn, Jackson, Best, Thomas, and Spiegelhalter \(2013\)](#), and [Asparouhov and Muthén \(2010\)](#). In other words, a possible proper prior for the variance covariance matrix is $IW(p\mathbf{I}, p)$ ¹. For this particular study, the prior for the variance covariance structure will be specified as $IW\left(\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, 2\right)$.
3. A data driven prior that substitutes the variance covariance elements of the IW distribution with those estimated from a standard FIML-EM run. In other words, the elements of the IW distribution will depend on the estimates of the variance covariance matrix, or $IW\left(2 \times \begin{bmatrix} a & b \\ b & c \end{bmatrix}, 2\right)$, where a, b, c are the estimates of the intercept variance, intercept slope covariance, and slope variance, respectively.

The results from the complete data study (no missing data conditions) are summarized in [Table 3.3](#), where the aggregated relative bias values of all the parameters for each method are presented. The Table shows that the noninformative, inverse Wishart priors (IW) for the variances showed the lowest values of relative bias. None of the estimates of the variance parameters were extremely biased. Although these findings were inconsistent with suggestions from past studies, for this particular situation, the improper inverse priors seemed to be the best choice to use for the missing data handling methods that would be compared in the main

¹Per [Lunn et al. \(2013\)](#), who advise scaling \mathbf{I} by p

simulation study.

Table 3.3

Mean Relative Bias Values by Analysis Method (Complete Data Methods)

Variable	MLEM	IW	PW	DD
Class 1 Intercept	0.04	0.00	0.02	0.00
Class 2 Intercept	-0.01	0.00	-0.01	-0.11
Class 1 Slope	-0.25	-0.04	-0.07	-0.06
Class 2 Slope	0.14	0.03	0.06	0.05
Intercept Variance	-0.05	0.08	-0.19	0.01
Slope Variance	-0.19	-0.04	-0.19	-0.14
Int. Slope Covariance	-0.20	-0.06	0.64	0.06
Residual Variance	-0.01	0.00	0.01	0.00
Proportion	-0.11	-0.02	-0.02	0.09

Note. MLEM = full data ML, IW = full data Bayesian with improper IW priors, PW = full data Bayesian with proper IW priors, DD = full data Bayesian with data-driven priors. Values in bold indicate problematic values according to several studies (Curran, West, & Finch, 1996; Flora & Curran, 2004; Kaplan, 1989).

3.4 Summary of Manipulated Conditions

Table 3.4 shows a summary of the conditions that will be fixed and manipulated. Fully crossed, a total of 800 combinations of manipulated conditions will be evaluated. For each sample size and degree of separation condition, an additional 32 conditions will be evaluated using the complete data using the ML-EM method or a fully Bayesian approach with improper IW priors. The results from the complete data analysis will serve as a basis for comparison for the missing data methods.

Table 3.4

Overview of Simulation Conditions

Factors	Levels
Repeated measures	4 timepoints
Latent classes	2 classes
Sample size	100, 200, 500, 1000
Degree of separation (MD)	0.5, 1.0, 1.5, 2.0
Missing rate	5, 10, 20, 30, 40
Missing mechanism	MCAR, CMAR ⁺
Missing data handling method	LD, FI, MI, FB, 2M

Note. LD = listwise deletion, FI = full-information maximum likelihood (EM), MI = single-stage multiple imputation, FB = fully Bayesian, 2M = two-stage multiple imputation, MCAR = missing completely at random, MNAR = missing not at random, CMAR⁺ = conditional extended missing at random.

For each sample size and class separation condition, datasets will be created in order to achieve a sufficient number of replications that will achieve stable average bias values. During a preliminary simulation, a running mean of the most egregious condition (40% missing, listwise deletion, with a sample size of 100 and very poor separation) was examined, with the idea that the parameter estimates from this condition would perhaps take the longest to stabilize. Both the mean and variance estimates were monitored. As can be seen in Figure 3.2, 100 replications was sufficient for the means and variance estimates to become stable. Therefore, it was determined that 100 replications would be used for every cell in the simulation.

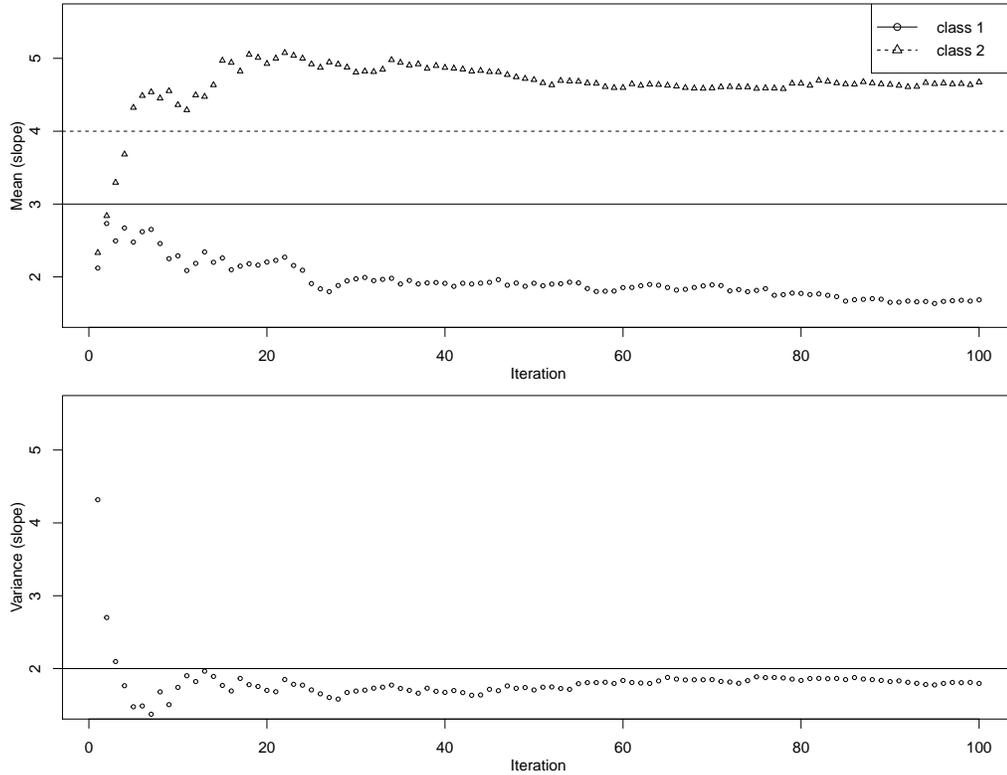


Figure 3.2. Running means of the slope mean (top) and slope variance (bottom) parameter estimates for the listwise deletion method. Horizontal lines are provided to show the true parameter values.

Class proportions, or the number of cases in each latent class, is another important manipulable factor that can have an effect on the estimation of GMMs. For example, [Tueller and Lubke \(2010\)](#) and [Tofghi and Enders \(2007\)](#) found that classes with relatively small sample sizes decreased the accuracy of identifying the classes. Although unrealistic, the proportion of each latent class will be kept the same in order to keep the study focused on the missing data handling techniques and other conditions.

All data simulation and manipulation will be conducted using Mplus version

8.1 (L. K. Muthén & Muthén, 1998) and R (Team, 2008). To facilitate the manipulation of all the conditions, R will be used to call Mplus, manipulate the codes, and collect the necessary output that will be used to evaluate the methods. The criteria that will be used to make the comparisons are discussed next.

3.5 Step 4: Evaluation Criteria

Several criteria will be considered when evaluating the results of the following simulation study. Convergence rates, accuracy rates, parameter estimate relative bias, and parameter estimate standard errors will be considered. Because the main interest of the study is to investigate how the different missing data handling methods perform under certain missing data conditions, the overall performance of the methods will be evaluated first. This will be done by comparing the means and medians of the absolute relative bias and standard error bias rates of all the parameter estimates combined, grouped by method of analysis. Then, to pinpoint specific differences across parameters, a factorial ANOVA will be conducted to find any manipulated conditions or interactions between conditions that show any meaningful differences across specific parameter estimates that account for differences in relative bias and/or standard error ratios.

3.5.1 Convergence Rates

It is expected that some conditions, such as ones with low sample sizes, high percentages of missing data, and low separation will experience higher rates of non-

convergence. Some replications may converge but produce nonsensical negative variance estimates. Solutions from such replications can taint the overall results of a particular cell, so they will not be included as a valid replication. The final convergence rate will be calculated as the total number of fully-converged datasets divided by the total number of replications that were needed to get to the desired number of 100 replications. The susceptibility of non-converge will be higher for the two-stage imputation because any of the imputed datasets at each stage may experience convergence issues. To address this, extra imputations will be created at each stage and filtered through to obtain imputed sets that have converged and have produced sensible solutions (i.e., positive estimates of the variance).

3.5.2 Accuracy Rates

Classification accuracy will be evaluated based on the percentage of correct and incorrect classifications. High classification accuracy will mean that a high percentage of individuals were classified into the true class. Most likely class memberships for each individual can be obtained from the posterior probability after estimation. These class assignments will then be compared to the individual's true class.

3.5.3 Relative Bias and Absolute Bias

For each manipulated condition, parameter recovery of the fixed and random effects for the set number of replications will be assessed using relative bias. Relative bias is computed as the difference between the mean of the parameter estimates and

the true value of the parameter divided by the true value of the parameter. That is,

$$\hat{\theta}_{RB} = \frac{\left(\frac{\sum_{b=1}^{reps} \hat{\theta}_b}{reps} - \theta_0 \right)}{\theta_0},$$

where $\hat{\theta}_b$ is the parameter estimate, θ_0 is the population value for θ and $reps$ is the number of replications. Relative bias is a percentage, and ideally will be close to zero. In many studies, a cut-off of 10% relative bias is considered substantial (see, e.g. [Curran et al., 1996](#); [Flora & Curran, 2004](#); [Kaplan, 1989](#)). The same criteria will be used to flag unacceptable levels of bias.

In addition to relative bias, absolute bias will be considered. Absolute bias will be calculated as the absolute value of the relative bias, which will be used to assess the overall bias observed across parameter estimates for each data analysis method. These absolute bias measures will then be aggregated using means in order to observe the overall bias for each analysis method. Medians will also be computed in order to assess the bias without any outliers.

3.5.4 Standard Error Bias

Bias of the standard errors of the estimates, or the analogous posterior standard deviations of the estimates for the Bayesian method, will be evaluated by the ratio between the square root of the mean variance of $\hat{\theta}$ and the standard deviation of the 100 parameter estimates. That is,

$$\text{SE Bias} = \frac{SE(\hat{\theta})}{SD_1(\hat{\theta})},$$

where

$$SE(\hat{\theta}) = \sqrt{100^{-1} \sum_{b=1}^{100} [SE_b(\hat{\theta})]^2},$$

and $SD_1(\hat{\theta}) = SD(\hat{\theta}) \cdot \sqrt{99/100}$, which is the corrected sample standard deviation of the 100 parameter estimates. Standard errors are considered to be accurate when this ratio is close to 1. Standard error bias values greater than 1 indicate overestimated standard errors, which lead to Type II errors, and values less than 1 indicate underestimated errors, which lead to Type I errors.

3.5.5 Factorial ANOVA

In order to pinpoint any effects of the manipulated conditions on the parameter estimates, a split-plot design factorial ANOVA will be conducted. In this particular design, the missing data handling method will be regarded as a within-subject factor, and all other conditions will be considered between-subject factors. All main effects (sample size, class separation, missing mechanism, missing rate, and missing data handling method) and up to three-way interactions will be considered on the relative bias and standard error bias of the estimates of the means, variances and class proportion parameters (9 parameters in total: two intercept means, two slope means, intercept and slope variances, intercept and slope covariances, residual variances, and two class proportions). Only factors and interactions that are identified as statistically significant (p -value ≤ 0.05) and with effect sizes of $\eta^2 \geq 0.06$ (Cohen, 1988) will be presented and discussed with additional detail.

3.6 Expectations from the Study

Given the past studies and the review of literature that was presented in Chapter 2, the expectations of the simulation study are as follows:

1. Previous studies have suggested using FIML-EM in finite mixture contexts, so the expectation is that it will perform well in terms of classification accuracy, parameter recovery and standard error estimation across both missingness mechanisms, large sample sizes, and moderate to large class separation. Its performance with smaller sample sizes and poorer class separation conditions should be not as optimal as using a Bayesian estimation approach.
2. For samples less than 500, the fully Bayesian method is expected to perform better than FIML-EM-based methods in terms of classification rates and parameter recovery. Differences in performance will not be as apparent for larger samples.
3. The two-stage MI approach should outperform single-stage MI, regardless of missingness mechanism, because previous studies indicated that MI will bias parameter estimates when the grouping structure is not accounted for ([Enders & Gottschall, 2011](#)).
4. Larger rates of missing will severely impact the performance of most methods and an interaction with sample size and class separation with missing data handling method should be significant. In other words, the extent of severity of convergence, classification accuracy, parameter estimate bias, and parameter

estimate standard error bias will mainly depend on the sample size and class separation.

Chapter 4: Results

This chapter will present the results from the simulation study that was described in Chapter 3. A few aspects of the simulation that were not planned a-priori will be discussed and considered in the organization of this chapter. The original analysis plan did not include running a separate analysis on the full dataset (dataset without any missing data) using different methods. Initially, the plan was to run the full data analysis using only maximum likelihood via EM (MLEM), and to compare the results from this analysis with those from the missing data methods when missing data conditions (missing rate and missingness mechanism) were factored into the simulation design. This would not have been a proper way to compare the different methods because the Bayesian-based methods for dealing with missing data would not have been comparable. An initial simulation comparing results from using ML-EM and Bayesian inference using the default Mplus noninformative priors revealed contrasting convergence rates, accuracy rates, and model parameter estimates. As a result of this preliminary investigation, the simulation design was slightly modified so that the full, unaltered data (without missing data), was analyzed using both ML-EM and the Bayesian approach using different priors first. The details of the different priors that were tested is discussed in Section 3.3.4 of Chapter 2. A thor-

ough review of the results from this comparison is presented in Section 4.1 of the current chapter.

Based on the results from the full data analysis, it made more sense to compare parameter estimates with estimates from the full data rather than the true estimates because the estimates from the full data analysis represented the best-case scenario. In other words, estimates from any missing data methods would be expected to produce results no better than the full data analysis, so it would only make sense that the estimates from the missing data methods be compared to the estimates from the full data methods.

With these changes in mind, Section 4.2 will discuss the results from the analysis conducted on the data with the missing data conditions. First, the convergence rates from each missing data handling method for each level of the manipulated conditions will be presented. Then, a summary of the accuracy rates will be provided. Finally, the relative bias and standard error bias of the parameter estimates will be presented accordingly based from a factorial ANOVA of the manipulated conditions. To conclude the chapter, the final section, Section 4.3, will be a summary of the main findings. To aid in the discussion of the chapter, the commonly used abbreviations, symbols, and terms used in Chapter 3 and beyond are provided in Table 4.1.

Table 4.1

Abbreviations and Symbols Used for Summary of Results

Condition/Variable	Term	Levels
Sample Size	SS	100, 200, 500, 1000
Separation	SP	0.5 (vpoor), 1.0 (poor), 1.5 (mod), 2.0 (high)
Missing Rate (%)	MR	5, 10, 20, 30, 40
Missing Mechanism	MM	MCAR, CMAR ⁺
Method	ME	MLEM, IW, PW, DD, LD, FI, MI, FB, 2M
Intercept and Slope Means	μ_{int}	
Slope Mean	μ_{slope}	
Intercept Variance	Ψ_{int}	
Slope Variance	Ψ_{slope}	
Intercept Slope Covariance	Ψ_{12}	
Residual Variance	Θ	
Class Proportion	π	

Note. MLEM = full data ML, IW = full data Bayesian with improper IW priors, PW = full data Bayesian with proper IW priors, DD = full data Bayesian with data-driven priors, LD = listwise deletion, FI = full-information maximum likelihood (EM), MI = single-stage multiple imputation, FB = fully Bayesian, 2M = two-stage multiple imputation.

4.1 Complete Data Analysis

The full data analysis involved comparing convergence rates, classification accuracy rates, relative bias of the parameter estimates and accuracy of the standard errors from GMM modeling on data without any missing data using two methods: ML estimation via EM (MLEM) and Bayesian estimation via Gibbs sampling. As was discussed in Section 3.3.4, the Bayesian method involved using three different sets of priors on the estimates of the variance parameters in order to identify a good set of priors for the missing data analysis. The results from this preliminary simulation study are presented next.

4.1.1 Convergence Rates

As was discussed in Section 3.3.2, convergence was identified when the replication satisfied all convergence criteria for the method that was being tested. The convergence rates for the methods used to analyze the complete data are presented in Table 4.2. The table shows that the Bayesian methods produced higher convergence rates than MLEM across different sample size and class separation conditions, with the exception of when the data-driven (DD) priors were used. This was an expected result because in order to obtain the priors for the DD condition, the MLEM estimation from where these priors came from would need to have converged. In other words, if the initial ML estimation did not converge, then the DD prior method would also have counted as non-converged.

Table 4.2

Mean Convergence Rates for Complete Data Methods Grouped by Separation and Sample Size

Method	SP	N = 100	N = 200	N = 500	N = 1000
MLEM	vpoor	0.42	0.57	0.78	0.78
IW		0.85	0.89	0.88	0.88
PW		0.93	0.91	0.97	0.93
DD		0.51	0.60	0.68	0.76
MLEM	poor	0.43	0.65	0.79	0.83
IW		0.86	0.87	0.89	0.93
PW		0.91	0.94	0.98	0.97
DD		0.53	0.61	0.68	0.74
MLEM	mod	0.50	0.65	0.78	0.78
IW		0.88	0.90	0.88	0.92
PW		0.94	0.94	0.94	0.96
DD		0.56	0.61	0.77	0.76
MLEM	high	0.55	0.68	0.79	0.69
IW		0.70	0.95	0.94	0.91
PW		0.94	0.93	0.96	0.96
DD		0.59	0.70	0.85	0.86

The convergence rates for the Bayesian approach using inverse-Wishart (IW) and proper inverse-Wishart (PW) priors remained relatively stable at 85% or more, with the exception of one condition at N= 100 and high separation for the IW priors. The PW priors showed the highest rates of convergence across all conditions near 90%. The IW priors also produced relatively high convergence rates and matched the convergence rates from using the PW priors. As sample size increased, the rate of convergence increased for the MLEM method, increasing from 55% to 70% in the high separation condition, although never reaching the same rate of above 90% that the Bayesian method reached when using PW priors.

4.1.2 Classification Accuracy Rates

The accuracy rates for the full data analyses are presented in Table 4.3. The accuracy rates for the MLEM and Bayesian approach using different prior specifica-

tions were quite poor when the separation was very poor regardless of sample size, dwindling at around 50% to 55%, although the Bayesian inference in general was producing classification accuracy rates that were slightly higher (no more than 2% greater). As separation increased, however, accuracy rates increased for all methods. Accuracy rates also increased as the sample size increased for the high separation condition, although the magnitude of difference was different for different methods. For example, for the MLEM method, the classification accuracy rates reached up to 11% higher when the separation was high. With Bayesian inference, the accuracy rates went from as low as 53% to as high as 70% (using PW priors) when the separation was high.

Another notable observation is that the Bayesian approach across different prior specifications produced similar classification accuracy rates across sample size and separation conditions. The DD priors always produced slightly lower (around 1% to 2% less) classification accuracy rates, which may have been due to the fact that the priors were constructed based on the results from the MLEM method. These results indicate that class separation contributed to how accurately the models were able to classify individuals, which is consistent with the literature. Similar to the observations made with the convergence rates, the Bayesian method consistently produced higher accuracy rates than MLEM across all cells, albeit not as noticeably as the convergence rates.

Table 4.3

Mean Classification Accuracy Rates for Complete Data Methods Grouped by Separation and Sample Size

Method	SP	N = 100	N = 200	N = 500	N = 1000
MLEM	vpoor	0.51	0.51	0.51	0.52
IW		0.53	0.54	0.54	0.54
PW		0.53	0.53	0.53	0.54
DD		0.53	0.53	0.53	0.53
MLEM	poor	0.53	0.54	0.53	0.54
IW		0.59	0.59	0.58	0.58
PW		0.59	0.59	0.58	0.59
DD		0.59	0.59	0.57	0.56
MLEM	mod	0.57	0.56	0.57	0.57
IW		0.64	0.64	0.64	0.63
PW		0.64	0.65	0.64	0.64
DD		0.65	0.63	0.62	0.60
MLEM	high	0.59	0.61	0.63	0.63
IW		0.69	0.69	0.68	0.68
PW		0.70	0.70	0.70	0.70
DD		0.68	0.68	0.68	0.68

4.1.3 Relative Bias

The means and medians of the overall absolute relative bias measures across ME and SS, and ME and SP are provided in Table 4.4. The absolute bias values were calculated by taking the absolute bias of the relative bias of each parameter and aggregating the values across all the parameters. This was done in order to get a general overview of bias for each method at varying conditions of sample size and class separation. Medians are provided alongside the means in order to more accurately assess the effects of any parameters that had extreme outlying relative bias measures. Values in bold indicate relative bias values exceeding the 10% acceptable threshold according to [Curran et al. \(1996\)](#) and others. A summary of this information is also provided by plots in Figure 4.1 for each sample size condition, and the plots in Figure 4.2 for each separation condition.

Focusing on medians across sample sizes, the tables show that the Bayesian methods using the IW priors and DD priors produced less biased parameter estimates than those from MLEM. Table 4.4 shows that overall, parameters estimated using MLEM were biased for all sample size conditions and the bias increased as sample size decreased. The MLEM estimates for smaller samples sizes, on average, produced relative bias measures of approximately 6.5% higher than the acceptable threshold of 10%. The mean was consistently higher for MLEM method, a reflection of a few parameters being extremely biased. The IW method was consistently below the 10% threshold, but the mean at high separation produced a relative bias value of 11.2% although the median bias remained at 2.7%, which indicates that a few parameters were extremely biased. The PW priors produced medians of relative bias below the acceptable threshold but means consistently above, again indicating that certain parameters were quite biased well beyond the 10% acceptable cut-off.

Table 4.4

Means and Medians of the Overall Absolute Relative Bias Grouped by Method, Sample Size (top), and Class Separation (bottom)

Method	N = 100		N = 200		N = 500		N = 1000	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
MLEM	0.167	0.145	0.161	0.166	0.125	0.092	0.125	0.100
IW	0.079	0.023	0.079	0.029	0.067	0.021	0.076	0.028
PW	0.151	0.059	0.148	0.040	0.127	0.051	0.114	0.033
DD	0.090	0.053	0.085	0.070	0.098	0.093	0.123	0.120
Method	vpoor		poor		mod		high	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
MLEM	0.161	0.166	0.131	0.121	0.137	0.103	0.149	0.108
IW	0.082	0.034	0.037	0.014	0.069	0.025	0.112	0.027
PW	0.165	0.059	0.152	0.036	0.119	0.035	0.105	0.031
DD	0.129	0.108	0.094	0.085	0.087	0.070	0.086	0.057

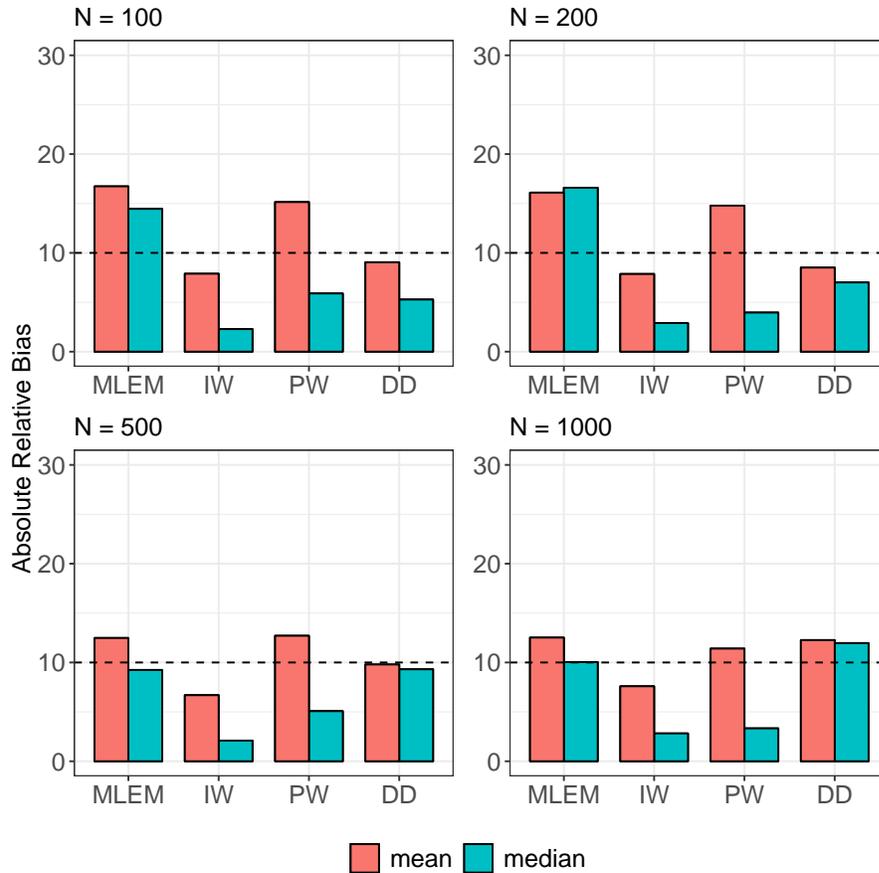


Figure 4.1. Overall means and medians of the absolute relative bias by ME and SS.

The bar plots across class separation (SP) and method (ME) in Figure 4.2 show that the means and medians of the relative bias measures decreased with larger separation. For example, the relative bias measures decreased when the separation went from very poor to high. The MLEM medians of the relative bias values went from 16% to 10%, the IW method medians of the relative bias went from 3.4% to 2.7%, the PW method medians of the relative bias went from 3.4% to 2.7% and the PW method medians of the relative bias went from 5.9% to 3.1%. Interestingly, the mean relative bias increased when the class separation became high, for example,

for the IW method, although the medians generally decreased. Again, this was an indication that certain parameters did not change much even with the higher sample size or when class separation was higher.

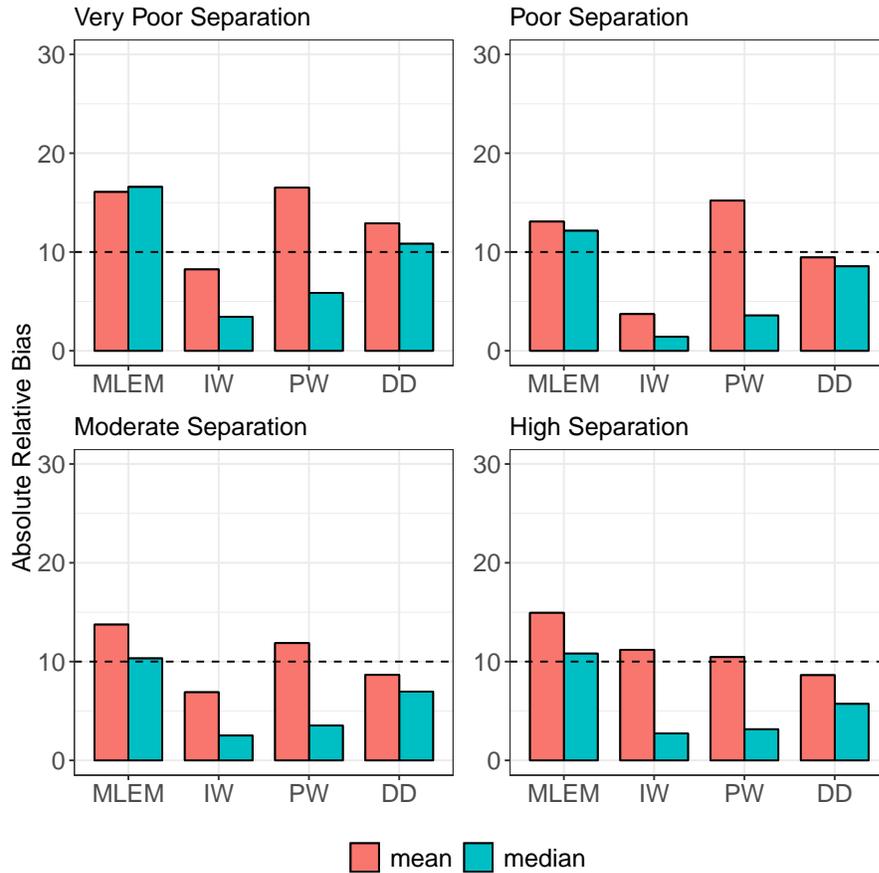


Figure 4.2. Overall means and medians of the absolute relative bias by ME and SP.

To pinpoint any meaningful differences in relative bias of each parameter estimate across conditions, a split-plot design factorial ANOVA was used. Method (ME) was regarded as a within subject factor while class separation (SP) and sample size (SS) were used as between subject factors. Presented in Table 4.5 and Table 4.6 are the percentages of variance explained (η^2 values) by the manipulated conditions

for the relative bias of each parameter estimate. Only effects that were significant according to the criteria specified in Section 3.5.5 were investigated. The table shows that there was a significant SP×ME interaction for relative bias of the estimates of the class proportion parameter ($\eta^2 = 13$). In addition, a significant SS×ME interaction was flagged for the relative bias of the the estimates of the class 2 intercept ($\eta^2 = 15.9$), proportion ($\eta^2 = 11.8$), slope variance ($\eta^2 = 10.4$), and residual variance ($\eta^2 = 31.3$). Among parameters with relative bias not flagged for any interaction effects, significant main effects for SP was observed for the relative bias of the estimates of the class 1 intercept mean ($\eta^2 = 50$), class 2 slope mean ($\eta^2 = 20.4$), intercept variance ($\eta^2 = 66.7$), intercept slope covariance ($\eta^2 = 34.5$), and residual variance ($\eta^2 = 6.5$) parameter estimates. Significant main effects for ME were also observed for the relative bias of all the estimates.

Table 4.5

Factorial ANOVA Results for Relative Bias of the Estimates of Parameter Means (Complete Data Methods)

Factor	$\mu_{int_{C1}}$			$\mu_{int_{C2}}$			$\mu_{slop_{C1}}$			$\mu_{slop_{C2}}$			π_{c1}		
	F	p-val.	η^2	F	p-val.	η^2	F	p-val.	η^2	F	p-val.	η^2	F	p-val.	η^2
SP×ME	-	-	-	-	-	-	-	-	-	-	-	-	13.9	< 0.01	13
SS×ME	-	-	-	10.4	< 0.01	15.9	-	-	-	-	-	-	12.7	< 0.01	11.8
SP	373.7	< 0.01	50	47.3	< 0.01	20.4	-	-	-	35.2	< 0.01	10.6	-	-	-
SS	47.5	< 0.01	6.3	-	-	-	108	< 0.01	8	-	-	-	18.3	< 0.01	7.6
ME	822.4	< 0.01	35.5	99.4	< 0.01	50.5	551.5	< 0.01	77.9	125.3	< 0.01	74.4	207.2	< 0.01	64.4

Note. Cells containing a “-” indicate non-significance (p -value > 0.05) and/or effect sizes less than 6%. SP: class separation; SS: sample size; ME: analysis method.

Table 4.7 is provided to examine the ME×SS interaction of the relative bias for the estimates of the class 2 intercept, slope variance, residual variance and class proportion parameter. First, while the ANOVA pointed to differences in the rela-

Table 4.6

Factorial ANOVA Results for Relative Bias of the Estimates of Parameter Variances (Complete Data Methods)

Factor	Ψ_{int}			Ψ_{slop}			Ψ_{12}			Θ		
	F	p -val.	η^2	F	p -val.	η^2	F	p -val.	η^2	F	p -val.	η^2
SS×ME	-	-	-	8.6	< 0.01	10.4	-	-	-	35.2	< 0.01	31.3
SP	445	< 0.01	66.7	-	-	-	113.5	< 0.01	34.5	4.8	0.028	6.5
ME	602.8	< 0.01	24.5	192.5	< 0.01	77.5	444.1	< 0.01	58.1	187.2	< 0.01	55.4

Note. Cells containing a “-” indicate non-significance (p -value > 0.05) and/or effect sizes less than 6%.

tive bias measures produced among the methods for each level of sample size, the majority of these relative bias measures, with the exception of the slope variance parameter, were below the 10% threshold, meaning that the methods produced acceptable levels of bias. Focusing on the slope variance parameter, all the methods, with the exception of the IW method, produced relative bias measures above 10%. The relative bias for the MLEM method and PW methods both increased in magnitude from -17% to -20% as sample sizes increased, while the DD method decreased from -18% to -12% in magnitude as sample sizes increased. The IW method followed a similar trend as the DD method, increasing in relative bias from 2% to 7% as the sample size increased. However, the IW method was the only method that was able to keep the relative bias of the slope variance parameter under the acceptable threshold.

The DD method and MLEM method also produced severely biased parameter estimates of the intercept means and proportion parameters at sample sizes of 500 and 1000 (-12% and -19% relative bias, respectively). The MLEM method also produced severely biased parameter estimates of the proportion parameter at sample

sizes of 100 and 200 (-12% and -18%, respectively).

Table 4.7

Mean Relative Bias Grouped by Method and Sample Size for Estimates of $\mu_{int_{C_2}}$, Ψ_{stop} , Θ , and π_{c1} (Complete Data Methods)

Method	Variable	N = 100	N = 200	N = 500	N = 1000
MLEM	$\mu_{int_{C_2}}$	-0.02	-0.01	-0.01	-0.01
IW		0.00	0.00	0.01	0.00
PW		-0.02	-0.02	-0.01	-0.01
DD		-0.04	-0.08	-0.12	-0.19
MLEM	Ψ_{stop}	-0.17	-0.17	-0.20	-0.20
IW		0.02	-0.03	-0.05	-0.07
PW		-0.17	-0.19	-0.19	-0.20
DD		-0.18	-0.13	-0.12	-0.12
MLEM	Θ	-0.02	-0.01	0.00	0.00
IW		0.00	0.00	0.00	0.00
PW		0.02	0.01	0.01	0.01
DD		0.00	0.00	0.00	0.00
MLEM	π_{c1}	-0.12	-0.18	-0.06	-0.08
IW		-0.01	-0.02	-0.02	-0.02
PW		-0.02	-0.03	-0.03	-0.02
DD		0.03	0.05	0.10	0.19

Note. Values in bold indicate severely biased estimates (greater than 10% relative bias in either direction).

Table 4.8 shows the significant ME×SP interaction for relative bias values of the proportion parameter. In addition to confirming that the MLEM method produced biased estimates of the proportion parameter, it confirms that the class separation condition was an important factor of how well the proportion parameter was recovered when using the MLEM and DD methods. Specifically, very poor separation conditions produced severely negatively biased estimates of the proportion parameter. For the MLEM method, the relative bias was -18% when separation was very poor, 12% when separation was poor, and -11% when separation was moderate, all which are considered severely biased. The DD method produced positively biased estimates of 13% for very poor separation and 11% for poor separation. The IW and PW method were unaffected by separation and produced unbiased estimates for all

separation conditions (between 2% and 1% negative bias).

Table 4.8

Mean Relative Bias Grouped by Method and Separation for Estimate of π_{c1} (Complete Data Methods)

Method	Variable	vpoor	poor	mod	high
MLEM	π_{c1}	-0.18	-0.12	-0.11	-0.03
IW		-0.02	-0.01	-0.02	-0.01
PW		-0.02	-0.02	-0.02	-0.02
DD		0.13	0.11	0.09	0.03

Table 4.5 shows that the majority of the variance for the differences in relative bias for certain parameters was explained by the main effects of the method. For example, for the estimates of the class 1 and class 2 slope parameters, 78% and 74% variance, respectively, was explained by the differences in the relative bias observed by method of analysis. The aggregated relative bias measures for any parameters that were not involved in any significant two-way interactions are provided in Table 4.9. The table shows that the MLEM produced severely biased estimates of the class 1 slope means (-25%) and class 2 slope means (14%), as well as the intercept slope covariance (-20%). The PW method produced severely biased estimates of the intercept variance (-19%) and the intercept slope covariance (64%).

Table 4.9

Mean Relative Bias Values Grouped by Analysis Method (Complete Data Methods)

Parameter	MLEM	IW	PW	DD
$\mu_{int_{c1}}$	0.04	0.00	0.02	0.00
$\mu_{slop_{c1}}$	-0.25	-0.04	-0.07	-0.06
$\mu_{slop_{c2}}$	0.14	0.03	0.06	0.05
Ψ_{int}	-0.05	0.08	-0.19	0.01
Ψ_{12}	-0.20	-0.06	0.64	0.06

4.1.4 Standard Error Bias

To get a general idea of how well each of the methods estimated the standard errors, the standard error ratios for all the parameters were aggregated using means and medians. These means and medians were grouped by sample size, and the resulting values are presented in Table 4.10. The values in this table are also graphically provided by Figure 4.3. The values of the medians are compared against the means in order to minimize the influence of any extreme SE/SD ratios. The Figure shows that the MLEM method produced underestimated standard errors regardless of sample size, and the standard errors decreased even further as sample size increased, from a median SE/SD ratio of 0.68 to 0.62. The IW priors produced overestimated standard errors for sample sizes less than 1000, but the ratios drawing closer to 1 as sample sizes increased was indicative that larger sample sizes corrected the standard error closer to the sampling variability. The PW priors produced the most accurate standard errors at sample size 500 (SE/SD ratio of 0.97 median values), but overestimated standard errors for sample size 100 (SE/SD ratio of 1.06) and 200 (SE/SD ratio of 1.10). The data-driven (DD) priors consistently produced underestimated standard errors and became more underestimated as sample size increased, from a mean of 0.91 ratio for $N = 100$ to 0.81 ratio for $N = 1000$. The overall mean and median of the standard error bias rates for each method across separation showed similar patterns as the plots across sample sizes so they are not presented, but the bottom of Table 4.10 shows similar patterns as those that were presented across sample sizes.

Table 4.10

Means and Medians of the SE/SD Ratios Grouped by Method, Sample Size (top), and Separation (bottom) for Complete Data Methods

Method	N = 100		N = 200		N = 500		N = 1000	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
MLEM	0.73	0.68	0.74	0.66	0.70	0.66	0.67	0.62
IW	1.30	1.26	1.22	1.20	1.13	1.11	1.03	1.03
PW	1.06	1.06	1.09	1.10	0.98	0.97	0.91	0.91
DD	0.91	0.99	0.87	0.97	0.87	0.96	0.81	0.81

me	vpoor		poor		mod		high	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
MLEM	0.75	0.71	0.71	0.64	0.69	0.63	0.70	0.61
IW	1.15	1.11	1.20	1.18	1.19	1.14	1.14	1.11
PW	1.04	1.05	1.03	1.00	1.02	1.01	0.95	0.94
DD	0.90	0.98	0.86	0.95	0.85	0.91	0.85	0.94

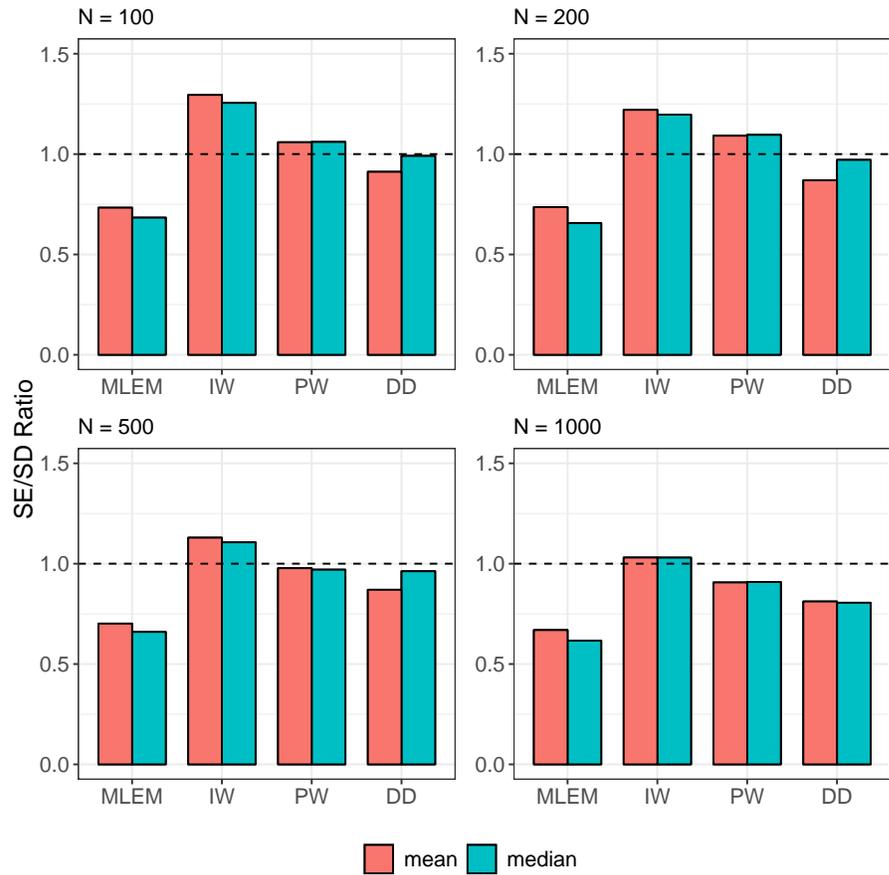


Figure 4.3. Overall means and medians of the SE/SD ratios by ME and SS.

The results from the ANOVA for the differences in SE/SD ratios presented in Table 4.11 indicate that there was a significant ME by SS interaction for the differences in the SE/SD ratios of the intercept means and residual variances. All other parameters were flagged as having significant differences in the SE/SD ratios for varying conditions of the main effects for sample sizes (SS), class separations (SP), and/or methods (ME).

Table 4.11

Factorial ANOVA Results for SE/SD Ratios of the Estimates of the Mean Parameters (Complete Data Methods)

Factor	μ_{intC1}			μ_{intC2}			μ_{slopC1}			μ_{slopC2}		
	<i>F</i>	<i>p-val.</i>	η^2	<i>F</i>	<i>p-val.</i>	η^2	<i>F</i>	<i>p-val.</i>	η^2	<i>F</i>	<i>p-val.</i>	η^2
SS×ME	10.1	< 0.01	12.6	-	-	-	-	-	-	-	-	-
SP	-	-	-	-	-	-	-	-	-	-	-	-
SS	-	-	-	-	-	-	-	-	-	-	-	-
ME	188.1	< 0.01	78.3	575	< 0.01	91.7	233.6	< 0.01	88.9	134.2	< 0.01	82.6

Note. Cells containing a “-” indicate non-significance (*p*-value > 0.05) and/or effect sizes less than 6%. SP: class separation; SS: sample size; ME: analysis method.

Table 4.12

Factorial ANOVA Results for SE/SD Ratios of the Estimates of the Variance Parameters (Complete Data Methods)

Factor	Ψ_{int}			Ψ_{slop}			Ψ_{12}			Θ		
	<i>F</i>	<i>p-val.</i>	η^2	<i>F</i>	<i>p-val.</i>	η^2	<i>F</i>	<i>p-val.</i>	η^2	<i>F</i>	<i>p-val.</i>	η^2
SS×ME	-	-	-	-	-	-	-	-	-	3.2	< 0.01	21.8
SP	-	-	-	-	-	-	14.1	< 0.01	15.8	-	-	-
SS	6.7	0.012	13.4	69.9	< 0.01	22.6	42.2	< 0.01	47.2	19.7	< 0.01	46.7
ME	123	< 0.01	73.4	75.9	< 0.01	63.1	15.1	< 0.01	17.6	-	-	-

Note. Cells containing a “-” indicate non-significance (*p*-value > 0.05) and/or effect sizes less than 6%.

The SE/SD ratios produced by each method by sample size for the intercept means and residual variances are presented in Table 4.13. Focusing on the intercept means, the SE/SD ratios ranged from 0.47 to 1.26. The standard errors were more

accurate (closest to 1) for the Bayesian approaches (IW, PW, DD) than the standard errors from MLEM, given that the ratios produced by the MLEM method were at around 50%, which indicates that the MLEM method underestimated the standard errors. Among the Bayesian methods, the PW prior produced the closest SE ratios to 1 for the intercept means when the sample size was 100 and 200 (0.98 and 0.93, respectively), while the IW prior produced the most accurate, albeit overestimated, standard errors at sample sizes of 500 and 1000 (1.07 and 0.97, respectively). The DD prior produced overestimated standard errors and the magnitude of them became greater as sample size increased, from a SE ratio of 1.12 at sample size 100 to 1.47 at sample size 1000. The variability of these ratios were not as great for the residual variances, although the differences by method and sample were flagged as significant by the ANOVA.

Table 4.13

Mean SE/SD Ratios Grouped by Method and Sample Size (Complete Data Methods)

Method	Variable	N = 100	N = 200	N = 500	N = 1000
MLEM	μ_{intC_1}	0.49	0.49	0.52	0.47
IW		1.43	1.18	1.07	0.97
PW		0.98	0.93	0.85	0.85
DD		1.12	1.05	1.26	1.47
MLEM	Θ	1.05	1.05	1.01	1.01
IW		0.97	1.07	0.97	1.02
PW		0.96	1.08	0.97	1.02
DD		0.99	1.03	0.99	0.99

Finally, for the rest of the parameters that were only flagged for SE/SD ratio differences for across each method, plots were produced and can be seen in Figure 4.4. The standard errors from MLEM were consistently underestimated while the standard errors for the Bayesian methods varied, but we generally closer to 1.

The intercept and slope variance measures for the data driven priors method were severely underestimated for the class 2 intercept, while the IW prior method was consistently producing overestimated standard errors. According to the plot, the proper inverse prior method produced the most accurate standard errors, with most of the SE/SD ratio of the parameters around 1.

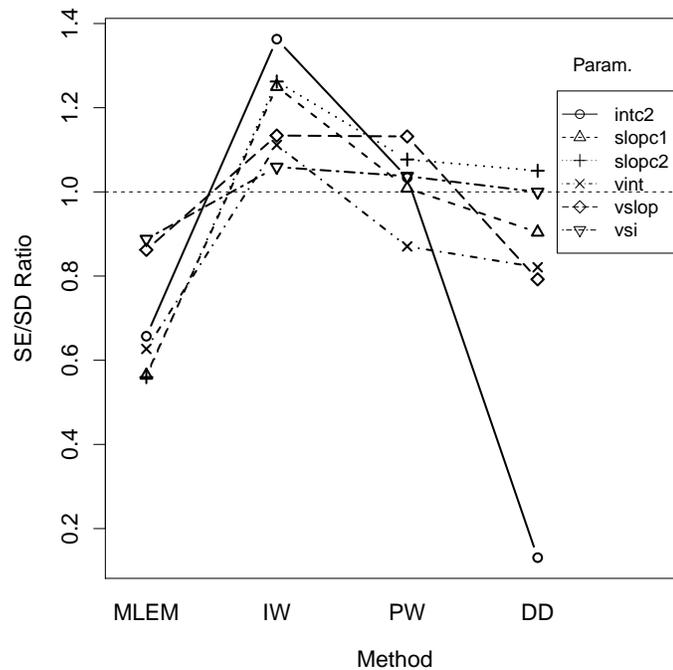


Figure 4.4. SE/SD ratios for class 2 intercept mean (intc2), class 1 and class 2 slope mean (slopc1, slopc2), intercept variance (vint), slope variance (vslop) and intercept slope covariance (vsi) for each method.

Based on the results observed from this section, all the Bayesian-based methods for handling missing data (FB and 2M) used the improper inverse Wishart (IW) prior because the convergence rates, accuracy rates, relative bias and standard error bias results indicated that the the Bayesian approach with the IW priors performed relatively better than the other approaches. Moving forward, it became important to

compare results from the ML-based methods (LD, FI, and MI) with the results from the full-data analysis using MLEM estimation, and the results from the Bayesian-based methods (FB and 2M) with the results from the full-data analysis using Bayesian inference with IW priors.

4.2 Missing Data Analysis

The results from the missing data analysis will be presented next in a similar format as the previous section, which discussed the results from the full data analysis using ML and the Bayesian method using three different prior specifications. Like in the previous section, a series of factorial ANOVAs will be used to narrow down the discussion to some key findings. First, the convergence rates across the missing data handling methods will be compared, focusing on the methods. Next, the accuracy rates will be compared. In the last two sections, the results from the comparison of relative bias and standard error bias will be presented.

4.2.1 Convergence Rates

Table 4.14 shows clear differences in the convergence rates across methods and sample sizes. As expected, the LD method produced the lowest convergence rates (28%). The FI method produced relatively low convergence rates compared to using MI, FB, and 2M, with a convergence rate as low as 38% when the sample size was 100. The Bayesian approach produced the highest rates of convergence, even when sample sizes were as small as 100, never going below 80%. The two-stage MI also produced relatively high rates of convergence, although not as high as the fully Bayesian approach, which maintained convergence rates between 81% and 89% .

Compared to the complete methods, the missing data methods produced lower rates of convergence when the missing data were introduced. One unexpected result occurred at sample sizes of 100, where the FB and 2M methods produced higher rates

of convergence (90% and 89%, respectively) than when the full data was analyzed using Bayesian inference with IW priors (82%).

Table 4.14

Mean Convergence Rates for Missing Data Methods Grouped by Method and Sample Size

Method	N = 100	N = 200	N = 500	N = 1000
MLEM	0.48	0.63	0.79	0.77
IW	0.82	0.90	0.90	0.91
LD	0.28	0.40	0.57	0.67
FI	0.39	0.52	0.66	0.75
MI	0.67	0.79	0.85	0.89
FB	0.90	0.90	0.84	0.82
2M	0.89	0.89	0.84	0.81

4.2.2 Classification Accuracy Rates

Presented in Table 4.15 are the accuracy rates for all the methods by separation condition. It is clear from this table that these rates improved as the separation increased from very poor to high for every method, with the largest improvement for the FB method, which went from 53% accuracy when the separation was very poor, to 66% when the separation was high. At the very poor classification condition, all the methods produced the same poor results of around 51% to 53% accuracy rates. The classification accuracy rates were also slightly better for the Bayesian method and the imputation methods across all separations, with the largest differences observed between FB and FI when the class separation was moderate. When the separation was high, FI, MI, and 2M produced similar classification accuracy rates, but the FB method was around 6% higher than all other methods. As ex-

pected, the missing data conditions dropped the classification accuracy rates across all methods.

Table 4.15

Mean Classification Accuracy Rates for Missing Data Methods Grouped by Method and Separation

Method	vpoor	poor	mod	high
MLEM	0.51	0.54	0.57	0.62
IW	0.54	0.58	0.64	0.68
LD	0.51	0.53	0.57	0.60
FI	0.51	0.53	0.56	0.60
MI	0.52	0.54	0.57	0.61
FB	0.53	0.57	0.62	0.66
2M	0.52	0.54	0.56	0.60

4.2.3 Relative Bias

Tables of the absolute relative bias grouped by sample sizes and analysis methods are presented in Table 4.16. The absolute relative bias measures for all the parameters were aggregated by method and by sample size and class separation to observe the general bias that was produced by each method. The large discrepancies between the means and medians of the absolute relative bias values indicates that the overall bias may be a result of a few parameter estimates that were extremely biased. For example, for sample sizes of 100, the medians for all the methods, with the exception of the MLEM and LD methods, were below the 10% threshold while many of the means were above the threshold. In general, the LD, FI, and MI methods showed the most problematic relative bias values for all samples sizes (values in bold). The FB and 2M methods produced the closest relative bias measures to the

full data method (IW), while the LD, FI, and MI methods were generally the most biased and close to values around the MLEM method.

Table 4.16

Means and Medians of the Overall Absolute Relative Bias Grouped by Method, Sample Size (Top), and Separation (Bottom) (Missing Data Methods)

Method	N = 100		N = 200		N = 500		N = 1000	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
MLEM	0.17	0.14	0.16	0.17	0.12	0.09	0.13	0.10
IW	0.08	0.02	0.08	0.03	0.07	0.02	0.08	0.03
LD	0.19	0.10	0.16	0.10	0.13	0.09	0.13	0.10
FI	0.15	0.09	0.13	0.09	0.12	0.08	0.13	0.10
MI	0.13	0.08	0.13	0.08	0.12	0.07	0.12	0.09
FB	0.10	0.03	0.08	0.02	0.07	0.02	0.07	0.03
2M	0.08	0.03	0.07	0.03	0.08	0.03	0.08	0.03
Method	vpoor		poor		mod		high	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
MLEM	0.16	0.17	0.13	0.12	0.14	0.10	0.15	0.11
IW	0.08	0.03	0.04	0.01	0.07	0.03	0.11	0.03
LD	0.15	0.10	0.15	0.10	0.15	0.10	0.16	0.10
FI	0.15	0.10	0.12	0.08	0.12	0.07	0.14	0.11
MI	0.17	0.09	0.13	0.10	0.10	0.06	0.11	0.08
FB	0.07	0.04	0.05	0.02	0.08	0.02	0.12	0.03
2M	0.08	0.04	0.05	0.02	0.06	0.02	0.10	0.04

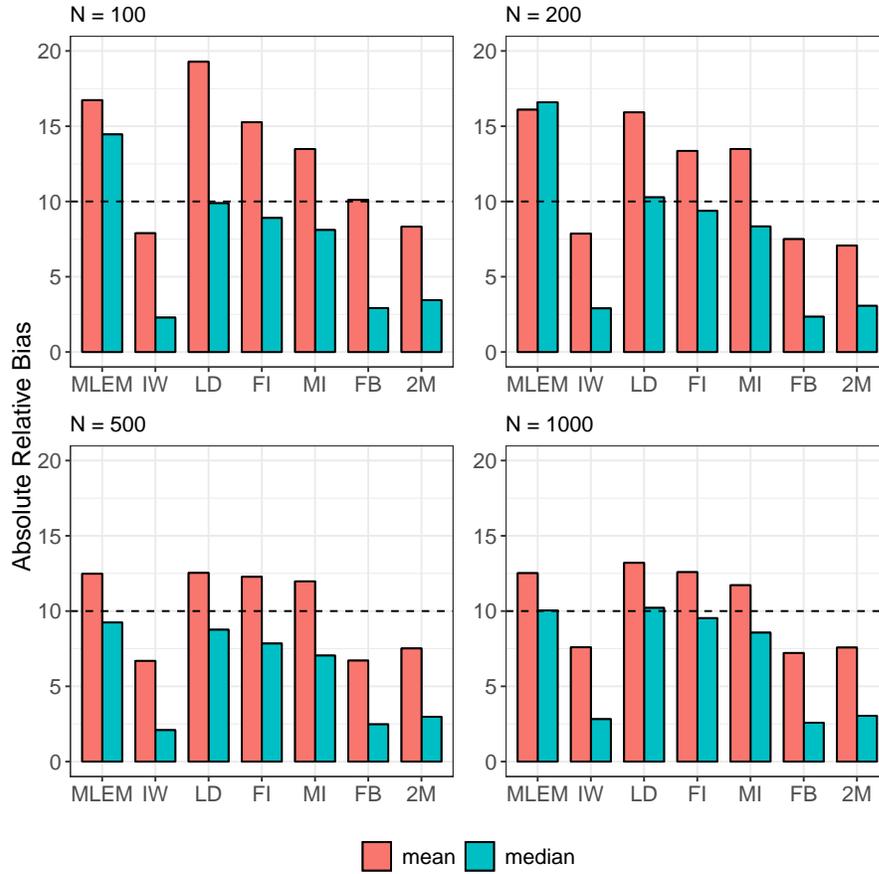


Figure 4.5. Overall means and medians of the absolute relative bias by ME and SS. The results from the full data analysis methods are included in the plots as a basis for comparison.

The bottom portion of Table 4.16 presents similar information grouped by the four class separation conditions. Overall, the very poor separation condition produced the estimates that were very biased, but the high separation condition also produced large values of relative bias. The mean of the relative bias values were nearly all above 10%. Again, the large discrepancies between the means and medians is an indication that the bias may have been attributed to a few parameters estimates that produced very large values of relative bias.

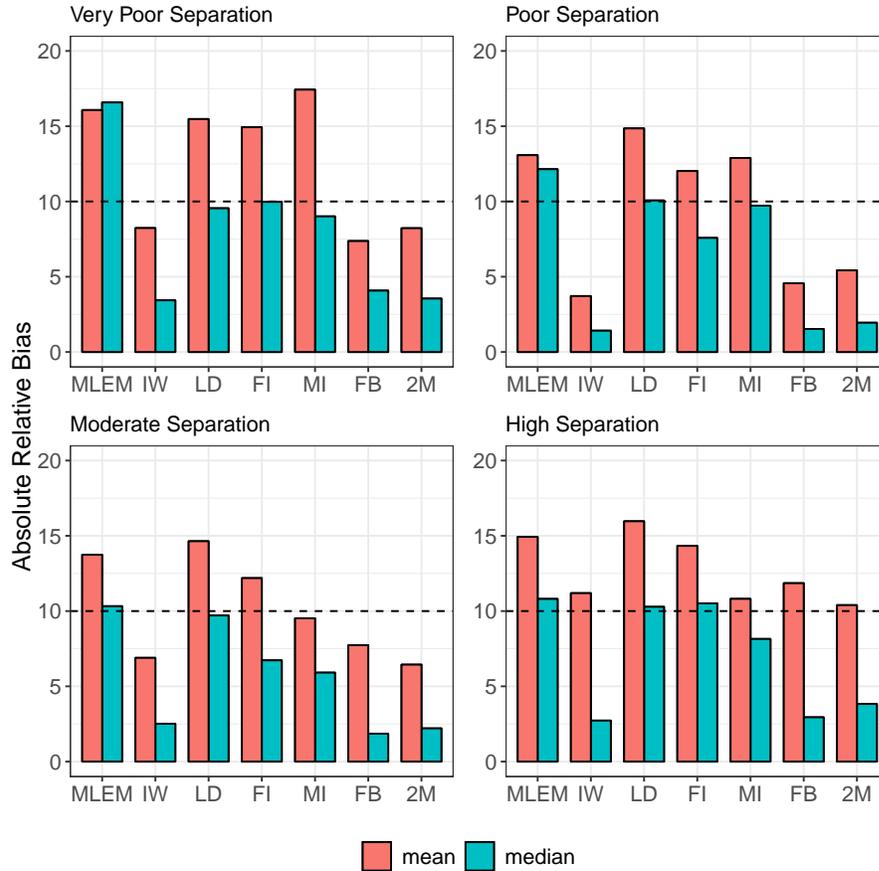


Figure 4.6. Overall means and medians of the absolute relative bias by ME and SP. The results from the full data analysis methods are included in the plots as a basis for comparison.

Table 4.17 shows the absolute relative bias for each method aggregated by each missing data rate. In general, larger percentages of missing data produced larger overall measures of relative bias, and these may be attributed to a few select variables because the medians were largely unaffected. For example, using the LD method, the median at 5% missing rate was at 9% relative bias and at 40% missing rate it was at 10% relative bias, but the mean relative bias went from 12.5% at 5% missing to almost 20% at 40% missing. The same observation can be made for FI,

MI, FB, and 2M methods, where the rate of change between the mean and median are different as missing data rate increased. The table also shows that the FB and 2M methods seem to have mitigated the effects of the missing data much better than the LD, FI, and MI, given that although missing rate increased, median bias also remained at around the same rates as the IW method (between 2.2 and 3.4% absolute relative bias). The discrepancies between the mean and median for these methods was also an indication that there may have been a few parameters that increased the values of relative bias as the rate of missing increased, as it can be seen that the mean relative bias at 40% was near 10% for both FB and 2M (9% and 8%, respectively).

Table 4.17

Means and Medians of the Overall Absolute Relative Bias Grouped by Method and Missing Rate

Method	5% miss.		10% miss.		20% miss.		30% miss.		40% miss.	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
MLEM†	0.145	0.136	-	-	-	-	-	-	-	-
IW†	0.075	0.026	-	-	-	-	-	-	-	-
LD	0.126	0.099	0.130	0.101	0.142	0.095	0.166	0.099	0.199	0.100
FI	0.124	0.082	0.125	0.092	0.130	0.093	0.138	0.086	0.153	0.097
MI	0.136	0.091	0.129	0.084	0.125	0.086	0.124	0.079	0.119	0.071
FB	0.067	0.022	0.069	0.025	0.078	0.026	0.084	0.025	0.097	0.028
2M	0.073	0.030	0.074	0.030	0.077	0.032	0.077	0.034	0.080	0.034

†These methods were not applied on any missing data conditions. Results reported here are from 0% missing rate condition.

Finally, Table 4.18 shows side-by-side tabulations of the overall absolute relative bias measures aggregated by missingness mechanisms and methods. The differences in the overall absolute relative bias measures were almost identical.

Table 4.18

Means and Medians of the Overall Absolute Relative Bias Grouped by Method and Missing Mechanism

me	MCAR		CMAR+	
	Mean	Median	Mean	Median
MLEM†	0.14	0.14	-	-
IW†	0.08	0.03	0.08	0.03
LD	0.16	0.10	0.14	0.10
FI	0.14	0.09	0.13	0.09
MI	0.13	0.08	0.13	0.08
FB	0.08	0.02	0.08	0.03
2M	0.08	0.03	0.08	0.03

†These methods were not applied on any missing data conditions. Results reported here are from 0% missing rate condition.

To pinpoint meaningful differences in relative bias of each parameter estimate between the conditions of the various manipulated conditions, a split-plot design factorial ANOVA was conducted, where method (ME) was considered a within subject factor and other manipulated factors (class separation (SP), sample size (SS), missing data rate (MR), missingness mechanism (MM)) were considered between subject factors. Presented in Table 4.19 and Table 4.20 are the percentages of variance explained (η^2 values) by the differences in the manipulated conditions for the relative bias of each parameter estimate. Only main effects and interaction effects that were significant according to the criteria presented in Section 3.5.5 were tabulated. A three-way interaction among missing rate (MR), sample size (SS), and method of analysis (ME) for the relative bias of the residual variance was flagged as significant. In addition, two-way interactions were flagged between SP×ME (class proportion parameter), SS×ME (intercept variance, slope variance, and intercept-slope covariance parameters), and ME×SS (intercept variance, slope variance, and

residual variance). All three-way interactions, which subsume any two-way interactions and main effects, as well as any two-way interactions, which subsume any significant main effects, will be discussed in more detail next.

Table 4.19

Factorial ANOVA Results for Relative Bias of Estimates of Parameter Means (Missing Data Methods)

Factor	μ_{intC1}		μ_{intC2}		μ_{slopC1}		μ_{slopC2}		π_{c1}	
	<i>F</i>	η^2	<i>F</i>	η^2	<i>F</i>	η^2	<i>F</i>	η^2	<i>F</i>	η^2
SP×ME	-	-	-	-	-	-	-	-	29.1	12.5
SP	4633.9	39.8	4187.6	57.8	268.4	10.3	203.7	10.7	156.5	17.4
SS	759	6.5	-	-	-	-	-	-	69.7	7.7
MM	-	-	-	-	-	-	-	-	177	6.6
ME	3309.2	42.4	891.6	24	1311.3	61.9	1097.5	64	-	-

Note. Cells containing a “-” indicate non-significance (p -value > 0.05) and/or effect sizes less than 6%. Values in bold are discussed in further detail. MM: missing mechanism; MR: missing rate; SP: class separation; SS: sample size; ME: analysis method.

Table 4.20

Factorial ANOVA Results for Relative Bias of Estimates of Parameter Variances (Missing Data Methods)

Factor	Ψ_{int}		Ψ_{slop}		Ψ_{12}		Θ	
	<i>F</i>	η^2	<i>F</i>	η^2	<i>F</i>	η^2	<i>F</i>	η^2
MR×SS×ME	-	-	-	-	-	-	85.8	15.2
SS×ME	282.4	6.4	104.8	9.5	176.3	8.3	552.6	24.4
MR×ME	-	-	-	-	-	-	343.1	20.2
SP	6676.8	56	-	-	2214.8	35.7	-	-
SS	-	-	752.2	20.7	1016.2	16.4	-	-
ME	3119.3	23.7	1559.4	47.1	1106.7	17.3	1960.1	28.9

Note. Cells containing a “-” indicate non-significance (p -value > 0.05) and/or effect sizes less than 6%. Values in bold are discussed in further detail. MM: missing mechanism; MR: missing rate; SP: class separation; SS: sample size; ME: analysis method.

4.2.3.1 Results of the Main Effects

To compare pairs of means of the relative bias values of the main effects, the Tukey’s HSD procedure was used. Tables 4.21 through 4.23 show the means of the

relative bias measures for any parameters for which the manipulated conditions were not flagged for interactions. Table 4.21 shows that while increasing class separation decreased the relative bias of the estimates of the intercept and slope mean parameters, they increased the relative bias of the estimates of the intercept variance and intercept slope covariance parameter. For example, the relative bias of the class 1 intercept mean when the separation was very poor was positively biased at 6%, but it gradually decreased to 0% as the class separation increased to high. Similarly, the relative bias of the class 2 slope means decreased from -22% when the separation was very poor to -13% when the separation was high. This similar pattern was not observed for the estimates of the variance, however. For the intercept variance, the very poor and high separation conditions produced severely biased parameter estimates in opposite directions, where the mean of the relative bias for the very poor condition was negatively biased at -21% and the mean of the relative bias for the high condition was positively biased at 21%. This similar pattern can be observed for the intercept slope covariance (Ψ_{12}), where the relative bias went from 21% when separation was very poor to -48% when the separation was high. Another observation to point out is that the most problematic parameter estimates were observed for the slope mean, intercept variance, and intercept slope covariance, as the majority of the aggregated relative bias values were outside of the $\pm 10\%$ acceptable range.

Table 4.21

Mean Relative Bias Grouped by Levels of Separation (SP)

SP	μ_{intC1}	μ_{intC2}	μ_{slopC1}	μ_{slopC2}	Ψ_{int}	Ψ_{12}
vpoor	0.06	-0.04	-0.22	0.16	-0.21	0.21
poor	0.04	-0.02	-0.18*	0.10*	-0.16	-0.04
mod	0.02	0.00	-0.16*	0.10*†	0.02	-0.36
high	0.00	0.02	-0.13	0.10†	0.21	-0.48

Note. Means are averaged over levels of ME, SS, MR, and MM. All pairwise comparisons were significant at the $\alpha = 0.05$ level unless otherwise noted. *Differences between poor and moderate separation were not significant. †Differences between moderate and high separation were not significant. All p -values were adjusted using the Tukey method. Values in bold indicate severe relative bias.

In regards to the ME main effect, Table 4.22 shows significant differences between all the relative bias values produced by the methods for most of the parameter estimates except for the intercept and slope means. The LD, FI and MI methods showed severely biased estimates of the slope mean (greater than 20%), while the FB and 2M methods were within the acceptable threshold of $\pm 10\%$. The differences were also significantly different, with the exception of the relative bias of the intercept and slope estimate using the FB and 2M methods. The means of the relative bias from the full data analysis are also provided as a way to examine how the missing data conditions affected the estimation of some of the parameters for the different methods. This comparison indicates that the relative bias was not affected when missing data were introduced because the relative bias for the MLEM-based methods remained at around -25% and the Bayesian methods remained at around 3-4% for the bias of these particular parameter estimates.

The SS main effect for the intercept mean also showed significant differences for different sample sizes, and Table 4.23 showed a decrease in the relative bias of

the parameter estimates as sample size increased, from a relative bias of 3.9% at sample size 100 to a relative bias of 1.7% at sample size 1000. While these pairwise differences were flagged as significant, they were within the 10% acceptable bias threshold.

Table 4.22

Mean Relative Bias Grouped by Analysis Methods (ME)

ME	μ_{intC1}	μ_{intC2}	μ_{slopC1}	μ_{slopC2}
MLEM	0.04	-0.01	-0.25	0.14
IW	0.00	0.00	-0.04	0.03
LD	0.06	-0.02	-0.25	0.14 †
FI	0.04	-0.03	-0.26	0.21
MI	0.03	-0.02	-0.21	0.14 †
FB	0.00*	0.00*	-0.06	0.03
2M	0.00*	0.00*	-0.09	0.05

Note. Means are averaged over levels of SP, SS, MR, and MM. All pairwise comparisons were significant at the $\alpha = 0.05$ level unless otherwise noted. *Differences between FB and 2M were not significant. †Differences between LD and MI were not significant. All p -values were adjusted using the Tukey method. Values in bold indicate severe relative bias.

Table 4.23

Mean Relative Bias Grouped by Sample Size (SS)

SS	μ_{intC1}
100	0.039
200	0.035
500	0.023
1000	0.017

Note. Means are averaged over levels of SP, ME, MR, and MM. All pairwise comparisons were significant at the $\alpha = 0.05$ level unless otherwise noted. All p -values were adjusted using the Tukey method.

4.2.3.2 Results of the Interaction Effects

To investigate the two-way interaction effects ($ME \times SS$ and $ME \times SP$), main effect pairwise comparisons between the methods were conducted for the different levels of the second factor. These comparisons are provided in Tables 4.24 through 4.27. The flagged three-way interaction ($MR \times SS \times ME$) was further investigated using the relative bias measures presented in Table 4.28 and a plot of these values provided by Figure 4.8.

Focusing on the estimates of the intercept variance, Table 4.24 shows significant differences in the relative bias for almost all comparisons between the methods except for FI and MI at sample size 500 and 1000, and FB and 2M at sample size 1000. The differences between methods decreased as the sample size increased. For example, at sample size 100, a near 50% difference in the relative bias value was observed between LD and FB, whereas when the sample size is 200, this difference decreased to 34% and continued to decrease at sample size 500 to 21%, and to 14% at sample size 1000. For sample size 100, the largest differences were observed between LD and FB or 2M (49% and 45%, respectively), FI and FB and 2M (29% and 25%, respectively), and MI and FB and 2M (36% and 32%, respectively). The smallest differences were observed between FI and MI (7%), and FB and 2M (4%). For sample size 200 to 1000, similar patterns emerged, where LD and FB or 2M, FI and FB or 2M, and MI and FB or 2M produced the largest differences, whereas FI and MI, and FB and 2M produced the smallest differences in measures of relative bias. The top left panel of Figure 4.7 shows these differences graphically. Another observation

to note that is made more evident by observing the Figure is that for sample size 100, all the conditions produced relative bias values outside of the acceptable 10% threshold. After sample size 100, only the FB and 2M method produced intercept variance estimates within the 10% threshold whereas for the other methods, it took a sample size of 1000 to produce intercept variance estimates that were acceptable.

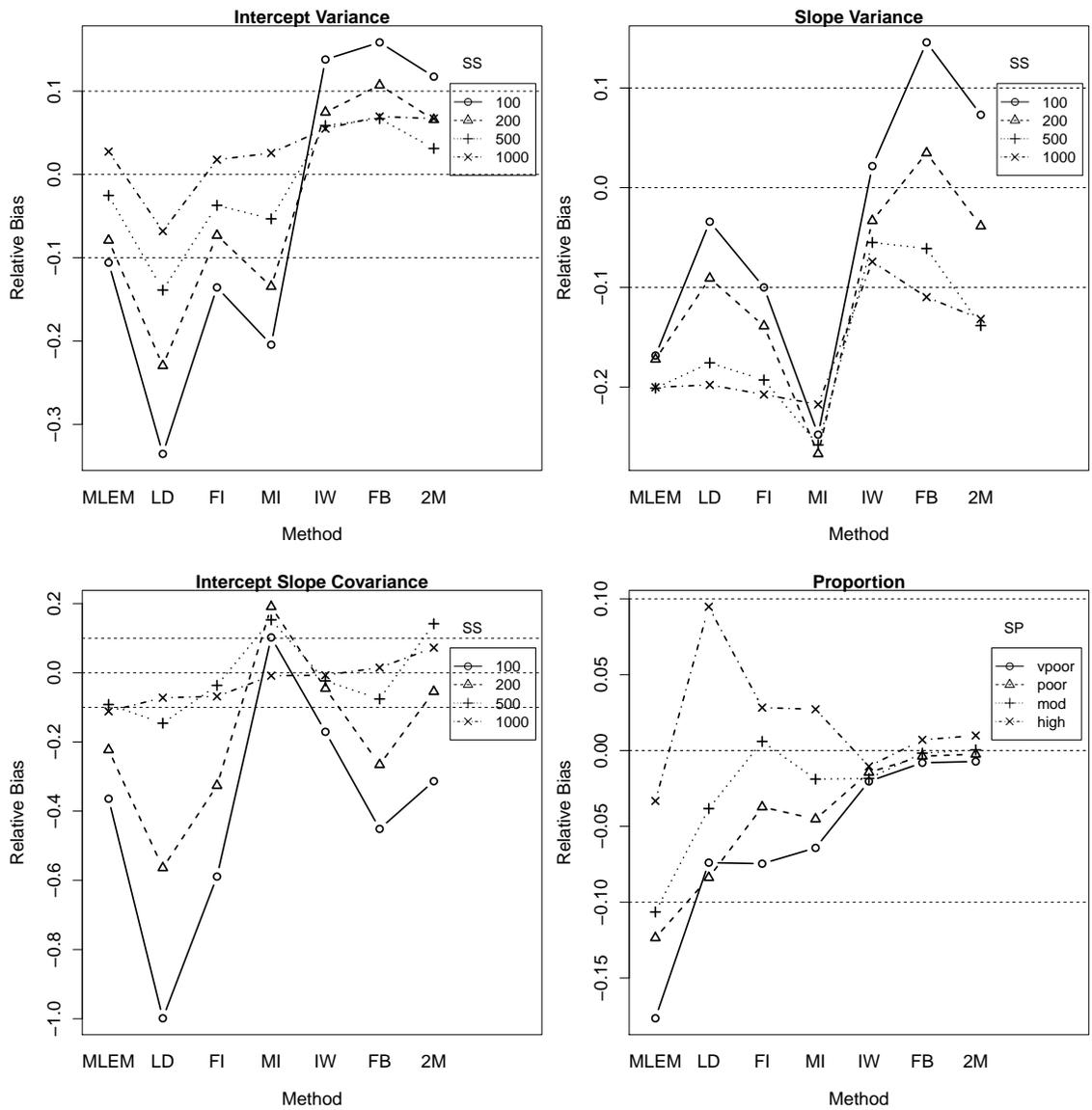


Figure 4.7. Interaction plots for $ME \times SS$ of the relative bias measure for the intercept variance (top left), slope variance (top right), intercept slope covariance (bottom left) and interaction plot for the $ME \times SP$ of relative bias of the estimate of the class proportion parameter (bottom right).

Table 4.24

Main Effect Pairwise Comparisons Corresponding to $ME \times SS$ Interaction of Intercept Variances for Relative Bias

Contrast (ME)	SS	$\bar{X}_{ME=j,SS=k}$	$\bar{X}_{ME=j',SS=k'}$	Difference
LD - FI	100	-0.34	-0.14	-0.20
LD - MI		-0.34	-0.20	-0.13
LD - FB		-0.34	0.16	-0.49
LD - 2M		-0.34	0.12	-0.45
FI - MI		-0.14	-0.20	0.07
FI - FB		-0.14	0.16	-0.29
FI - 2M		-0.14	0.12	-0.25
MI - FB		-0.20	0.16	-0.36
MI - 2M		-0.20	0.12	-0.32
FB - 2M		0.16	0.12	0.04
LD - FI	200	-0.23	-0.07	-0.16
LD - MI		-0.23	-0.13	-0.10
LD - FB		-0.23	0.11	-0.34
LD - 2M		-0.23	0.07	-0.30
FI - MI		-0.07	-0.13	0.06
FI - FB		-0.07	0.11	-0.18
FI - 2M		-0.07	0.07	-0.14
MI - FB		-0.13	0.11	-0.24
MI - 2M		-0.13	0.07	-0.20
FB - 2M		0.11	0.07	0.04
LD - FI	500	-0.14	-0.04	-0.10
LD - MI		-0.14	-0.05	-0.09
LD - FB		-0.14	0.07	-0.21
LD - 2M		-0.14	0.03	-0.17
FI - MI		-0.04	-0.05	0.02*
FI - FB		-0.04	0.07	-0.10
FI - 2M		-0.04	0.03	-0.07
MI - FB		-0.05	0.07	-0.12
MI - 2M		-0.05	0.03	-0.08
FB - 2M		0.07	0.03	0.04
LD - FI	1000	-0.07	0.02	-0.09
LD - MI		-0.07	0.03	-0.09
LD - FB		-0.07	0.07	-0.14
LD - 2M		-0.07	0.07	-0.14
FI - MI		0.02	0.03	-0.01*
FI - FB		0.02	0.07	-0.05
FI - 2M		0.02	0.07	-0.05
MI - FB		0.03	0.07	-0.04
MI - 2M		0.03	0.07	-0.04
FB - 2M		0.07	0.07	0.00*

Note. Means are averaged over levels of SP, MR, and MM. All pairwise comparisons were significant at the $\alpha = 0.05$ level unless otherwise noted by a *. All p -values were adjusted using the Tukey method.

Focusing on the estimates of the slope variance, Table 4.25 shows some counter intuitive results. For example, for some of the methods, smaller sample sizes actually produced relative bias measures closer to 0. This was the case for the values produced by the LD, FI and MI methods, where for sample size 100, the relative bias for these methods was 3%, 10%, and 25%, respectively, for sample size 200, the relative bias increased to 9%, 14%, and 27%, for sample size 500, it increased further to 18% and 19% for the LD and MI methods, and for sample size 1000, these values increased even further to 20% and 21% for the LD and MI methods. The Bayesian-based methods (FB and 2M) showed a different pattern where the sample sizes of 100 and 1000 produced the greatest degrees of relative bias (i.e., FB at sample size 100 produced 15% relative bias and FB at sample size 1000 produced 11% relative bias) and sample sizes of 200 and 500 produced relative bias below 10% (i.e., FB at sample size 200 produced 3% relative bias and FB at sample size 500 produced 6% relative bias).

The top right panel of Figure 4.7 shows the relative bias values of the slope variance together with the relative bias values produced by the complete data analysis methods (MLEM and IW). The figure shows that the MLEM method produced extremely negatively biased estimates of the slope variance (around -20% relative bias), which makes the results observed from the sample size of 100 for the LD and MI methods seem questionable. The results observed for sample sizes of 500 and 1000 for the ML-based methods were actually consistent with the complete data MLEM method. The IW method was actually negatively biased, and the Figure shows that when missing data were added, the variability of the relative bias values

increased for both the FB and 2M method, with relative bias measures deviating the most away from the IW method for sample sizes of 100 and 200.

Table 4.25

Main Effect Pairwise Comparisons Corresponding to ME × SS Interaction of Slope Variances for Relative Bias

Contrast (ME)	SS	$\bar{X}_{ME=j,SS=k}$	$\bar{X}_{ME=j',SS=k'}$	Difference
LD - FI	100	-0.03	-0.10	0.07
LD - MI		-0.03	-0.25	0.21
LD - FB		-0.03	0.15	-0.18
LD - 2M		-0.03	0.07	-0.11
FI - MI		-0.10	-0.25	0.15
FI - FB		-0.10	0.15	-0.25
FI - 2M		-0.10	0.07	-0.17
MI - FB		-0.25	0.15	-0.39
MI - 2M		-0.25	0.07	-0.32
FB - 2M		0.15	0.07	0.07
LD - FI	200	-0.09	-0.14	0.05
LD - MI		-0.09	-0.27	0.18
LD - FB		-0.09	0.03	-0.13
LD - 2M		-0.09	-0.04	-0.05
FI - MI		-0.14	-0.27	0.13
FI - FB		-0.14	0.03	-0.17
FI - 2M		-0.14	-0.04	-0.10
MI - FB		-0.27	0.03	-0.30
MI - 2M		-0.27	-0.04	-0.23
FB - 2M		0.03	-0.04	0.07
LD - FI	500	-0.18	-0.19	0.02*
LD - MI		-0.18	-0.26	0.08
LD - FB		-0.18	-0.06	-0.11
LD - 2M		-0.18	-0.14	-0.04
FI - MI		-0.19	-0.26	0.07
FI - FB		-0.19	-0.06	-0.13
FI - 2M		-0.19	-0.14	-0.05
MI - FB		-0.26	-0.06	-0.20
MI - 2M		-0.26	-0.14	-0.12
FB - 2M		-0.06	-0.14	0.08
LD - FI	1000	-0.20	-0.21	0.01*
LD - MI		-0.20	-0.22	0.02
LD - FB		-0.20	-0.11	-0.09
LD - 2M		-0.20	-0.13	-0.07
FI - MI		-0.21	-0.22	0.01*
FI - FB		-0.21	-0.11	-0.10
FI - 2M		-0.21	-0.13	-0.08
MI - FB		-0.22	-0.11	-0.11
MI - 2M		-0.22	-0.13	-0.09
FB - 2M		-0.11	-0.13	0.02

Note. Means are averaged over levels of SP, MR, and MM. All pairwise comparisons were significant at the $\alpha = 0.05$ level unless otherwise noted by a *. All p -values were adjusted using the Tukey method.

Focusing on the estimates of the intercept slope covariance, Table 4.26 shows both that the extreme relative bias values are corrected by larger sample sizes and that the differences between the method create smaller differences. The sample size effect can be seen by the extreme change in relative bias when using FI, for example, which goes from -59% for sample size 100, to -7% for sample size 1000. In addition, the greatest differences were between LD and FB or 2M, FI and 2M, and MI and FI and 2M. For sample size 100, the different between LD and FB is -55%, the difference between FI and 2M is -28%, and the difference between MI and FB is 55%. For sample size 1000, these differences become -9%, -14%, and -14%. Examining the bottom left plot of Figure 4.7, it is clear that sample size 100 created the most biased estimates of intercept slope covariance. The estimates using the MI method produced covariances with relative bias values that deviated from all other methods (positively biased). The FB and 2M methods produced relative bias values closest to the IW method, even when sample sizes were at 100 and 200, albeit extremely downwards biased.

Table 4.26

Main Effect Pairwise Comparisons Corresponding to ME × SS Interaction of Intercept Slope Covariances for Relative Bias

Contrast (ME)	SS	$\bar{X}_{ME=j,SS=k}$	$\bar{X}_{ME=j',SS=k'}$	Difference
LD - FI	100	-1.00	-0.59	-0.41
LD - MI		-1.00	0.10	-1.10
LD - FB		-1.00	-0.45	-0.55
LD - 2M		-1.00	-0.31	-0.69
FI - MI		-0.59	0.10	-0.69
FI - FB		-0.59	-0.45	-0.14
FI - 2M		-0.59	-0.31	-0.28
MI - FB		0.10	-0.45	0.55
MI - 2M		0.10	-0.31	0.42
FB - 2M		-0.45	-0.31	-0.14
LD - FI	200	-0.56	-0.33	-0.24
LD - MI		-0.56	0.19	-0.75
LD - FB		-0.56	-0.27	-0.30
LD - 2M		-0.56	-0.05	-0.51
FI - MI		-0.33	0.19	-0.52
FI - FB		-0.33	-0.27	-0.06
FI - 2M		-0.33	-0.05	-0.27
MI - FB		0.19	-0.27	0.46
MI - 2M		0.19	-0.05	0.25
FB - 2M		-0.27	-0.05	-0.21
LD - FI	500	-0.15	-0.04	-0.11
LD - MI		-0.15	0.15	-0.30
LD - FB		-0.15	-0.08	-0.07
LD - 2M		-0.15	0.14	-0.29
FI - MI		-0.04	0.15	-0.19
FI - FB		-0.04	-0.08	0.04*
FI - 2M		-0.04	0.14	-0.18
MI - FB		0.15	-0.08	0.23
MI - 2M		0.15	0.14	0.01*
FB - 2M		-0.08	0.14	-0.22
LD - FI	1000	-0.07	-0.07	0.00*
LD - MI		-0.07	-0.01	-0.06
LD - FB		-0.07	0.01	-0.09
LD - 2M		-0.07	0.07	-0.14
FI - MI		-0.07	-0.01	-0.06
FI - FB		-0.07	0.01	-0.08
FI - 2M		-0.07	0.07	-0.14
MI - FB		-0.01	0.01	-0.02*
MI - 2M		-0.01	0.07	-0.08
FB - 2M		0.01	0.07	-0.06

Note. Means are averaged over levels of SP, MR, and MM. All pairwise comparisons were significant at the $\alpha = 0.05$ level unless otherwise noted by a *. All p -values were adjusted using the Tukey method.

Focusing on the relative bias of the proportion parameter, Table 4.27 shows smaller values of relative bias overall and several comparisons among the methods showing non-significant differences between the values produced among the different methods as the class separation increases. At moderate and high class separation conditions, these differences become only significant between LD and all other methods. The ME×SP interaction plot for the proportion parameter presented in the bottom right panel of Figure 4.7 shows clear differences for vpoor and poor conditions. Although almost all estimates were within the $\pm 10\%$ acceptable bias range and many of the differences between MLEM-based methods and Bayesian-based methods, the Bayesian-based methods (IW, FB, and 2M) still produced the smallest values of relative bias regardless of sample size. For MLEM-based methods, the pattern was mostly unclear, and surprisingly, the complete data MLEM method produced some severely underestimated parameters when the separation was less than high.

Table 4.27

Main Effect Pairwise Comparisons Corresponding to ME × SP Interaction of Proportion Parameter for Relative Bias

Contrast (ME)	SP	$\bar{X}_{ME=j,SS=k}$	$\bar{X}_{ME=j',SS=k'}$	Difference
LD - FI	vpoor	0.07	0.07	0.00*
LD - MI		0.07	0.06	0.01*
LD - FB		0.07	0.01	0.07
LD - 2M		0.07	0.01	0.07
FI - MI		0.07	0.06	0.01*
FI - FB		0.07	0.01	0.07
FI - 2M		0.07	0.01	0.07
MI - FB		0.06	0.01	0.06
MI - 2M		0.06	0.01	0.06
FB - 2M		0.01	0.01	0.00*
LD - FI		poor	0.08	0.04
LD - MI	0.08		0.05	0.04
LD - FB	0.08		0.00	0.08
LD - 2M	0.08		0.00	0.08
FI - MI	0.04		0.05	-0.01*
FI - FB	0.04		0.00	0.03
FI - 2M	0.04		0.00	0.03
MI - FB	0.05		0.00	0.04
MI - 2M	0.05		0.00	0.04
FB - 2M	0.00		0.00	0.00*
LD - FI	mod	0.04	-0.01	0.04
LD - MI		0.04	0.02	0.02*
LD - FB		0.04	0.00	0.04
LD - 2M		0.04	0.00	0.04
FI - MI		-0.01	0.02	-0.02
FI - FB		-0.01	0.00	-0.01*
FI - 2M		-0.01	0.00	-0.01*
MI - FB		0.02	0.00	0.02*
MI - 2M		0.02	0.00	0.02*
FB - 2M		0.00	0.00	0.00*
LD - FI	high	-0.09	-0.03	-0.07
LD - MI		-0.09	-0.03	-0.07
LD - FB		-0.09	-0.01	-0.09
LD - 2M		-0.09	-0.01	-0.08
FI - MI		-0.03	-0.03	0.00*
FI - FB		-0.03	-0.01	-0.02*
FI - 2M		-0.03	-0.01	-0.02*
MI - FB		-0.03	-0.01	-0.02*
MI - 2M		-0.03	-0.01	-0.02*
FB - 2M		-0.01	-0.01	0.00*

Note. Means are averaged over levels of SS, MR, and MM. All pairwise comparisons were significant at the $\alpha = 0.05$ level unless otherwise noted by a *. All p -values were adjusted using the Tukey method.

The three-way interaction among ME, SS, and MR conditions for the estimates of the residual variance is presented in Figure 4.8 with values that were used to produce the plots provided by Table 4.28. To make fair comparisons, the values and plots of the relative bias for the full data analysis using ML and IW priors are also provided. The first glaring observation is that larger missing rates produced more biased estimates across all sample sizes and methods. As expected, the greatest bias was observed for the estimates from the LD method, where the bias increased to unacceptable ranges for sample size of 100 at missing data rate of 30% (-10% relative bias) and 40% (-19% relative bias). Another notable observation is that bias of the residual variances generally decreased as sample size increased. For example, for the FI method, at a 40% missing rate, the residual variance went from -5.7% for sample size 100 to .2% for sample size 1000. This is made clear by the fact that the scale of the plots decrease as the sample sizes increase. Overall, once the sample size reached 200, none of the methods produced estimates outside of the 10% relative bias criteria except for when the missing data rate was at 30 and 40% and the LD method was used.

Table 4.28

Mean Relative Bias of the Residual Variance Estimate Grouped by Method, Missing Data Rate, and Sample Size

Method	MR	N = 100	N = 200	N = 500	N = 1000
IW	0	-0.001	-0.004	-0.001	0.002
MLEM	0	-0.025	-0.011	-0.003	0.003
LD	5	-0.028	-0.013	-0.002	0.001
FI	5	-0.027	-0.014	-0.002	0.001
MI	5	-0.008	-0.008	0.001	0.003
FB	5	-0.003	-0.005	-0.001	0.003
2M	5	0.003	-0.002	0.000	-0.003
LD	10	-0.031	-0.013	-0.005	0.001
FI	10	-0.029	-0.014	-0.003	0.002
MI	10	-0.006	-0.006	0.003	0.004
FB	10	-0.004	-0.006	-0.003	0.002
2M	10	0.007	-0.002	-0.002	-0.001
LD	20	-0.060	-0.023	-0.006	-0.001
FI	20	-0.031	-0.019	-0.005	0.000
MI	20	0.004	0.003	0.004	0.004
FB	20	-0.012	-0.009	-0.005	0.000
2M	20	0.030	0.007	0.004	-0.001
LD	30	-0.100	-0.042	-0.015	-0.007
FI	30	-0.038	-0.025	-0.009	0.000
MI	30	0.014	0.011	0.008	0.005
FB	30	-0.016	-0.015	-0.010	-0.003
2M	30	0.056	0.020	0.007	0.000
LD	40	-0.193	-0.092	-0.035	-0.010
FI	40	-0.057	-0.041	-0.012	-0.002
MI	40	0.035	0.017	0.011	0.008
FB	40	-0.026	-0.024	-0.016	-0.003
2M	40	0.096	0.034	0.009	0.004

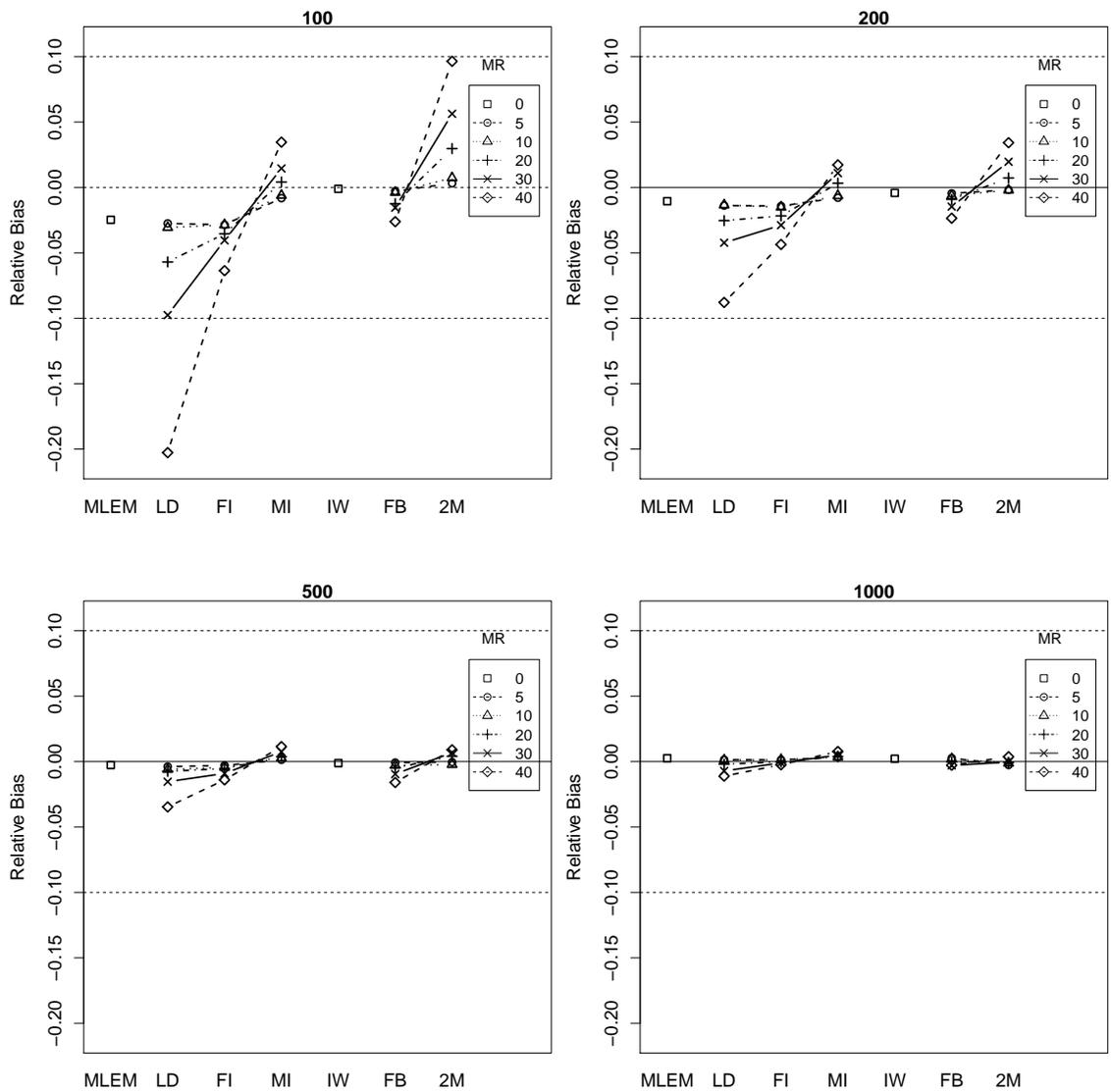


Figure 4.8. Three-way interaction $ME \times SS \times MR$ for the relative bias of the residual variances.

4.2.4 Standard Error Bias

The overall standard errors across method and varying sample sizes can be seen in the top portion of Table 4.29. These values are also graphically presented by Figure 4.9. Graphically, the trends look very clear. The MLEM method and all other methods based on ML were consistently underestimated. The Bayesian-based methods (IW, FB, and 2M) were consistently overestimated. Although difficult to tell from the bar plots, the LD method produced even more underestimated standard errors as sample size increased, producing a mean SE/SD ratio of 0.78 for sample size 100 and a SE/SD ratio of 0.72 for sample size 1000. This ratio also decreased when using the FI method, from 0.75 to 0.69. These results were actually consistent with the complete data MLEM method, which produced a ratio of 0.73 for sample size 100 and a ratio of 0.67 for sample size 1000. The MI and FB methods actually approached 1 as sample size increased (from 0.89 to 0.96 and from 1.30 to 1.04, respectively). The 2M method produced the most overestimated standard errors of all, and the larger sample size actually increased the standard errors (from 1.50 to 2.1). The overall standard errors across class separation conditions are also provided on the bottom portion of Table 4.29 and showed a similar trend as those observed from different sample size conditions.

For each missing data rate, the MI method produced accurate standard errors as the rate of missing increased, producing SE/SD ratios from 0.72 to 1.123 when missing rate went from 5% to 40%. The FB method kept standard errors consistent with SE/SD ratios near 1.15, which were closest to SE/SD ratios produced by the full

data analysis using IW, while the LD and FI methods maintained underestimated standard errors, similar to those produced by the full data MLEM method (at around 0.7 for the means).

Table 4.29

Overall Mean SE/SD Ratios for Missing Data Methods Grouped by Method, Sample Size (top), and Class Separation (bottom)

Method	N = 100		N = 200		N = 500		N = 1000	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
MLEM	0.733	0.684	0.736	0.656	0.701	0.660	0.669	0.617
IW	1.295	1.255	1.220	1.197	1.130	1.107	1.031	1.030
LD	0.789	0.668	0.772	0.677	0.734	0.665	0.726	0.674
FI	0.751	0.672	0.747	0.639	0.714	0.657	0.694	0.650
MI	0.896	0.889	0.936	0.918	0.946	0.954	0.962	0.958
FB	1.296	1.290	1.249	1.233	1.108	1.082	1.036	1.028
2M	1.503	1.625	1.704	1.821	1.799	1.947	1.946	2.084

Method	vpoor		poor		mod		high	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
MLEM	0.746	0.715	0.709	0.640	0.687	0.625	0.697	0.614
IW	1.149	1.106	1.201	1.177	1.187	1.137	1.138	1.106
LD	0.762	0.677	0.750	0.658	0.756	0.666	0.753	0.686
FI	0.742	0.656	0.719	0.645	0.724	0.651	0.720	0.647
MI	0.956	0.960	0.923	0.910	0.926	0.928	0.935	0.927
FB	1.184	1.165	1.169	1.150	1.192	1.148	1.144	1.105
2M	1.828	1.952	1.765	1.980	1.720	1.906	1.639	1.719

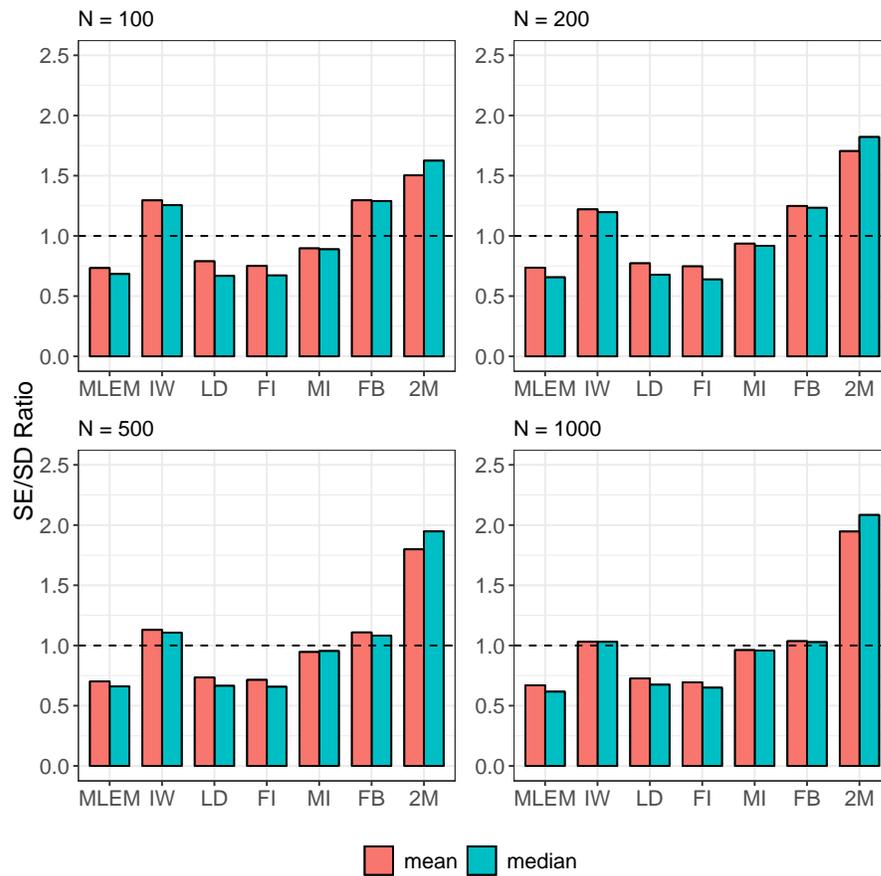


Figure 4.9. Overall means and medians of the SE/SD ratios by ME and SS (Missing data methods). The results from the full data analysis methods are included in the plots as a basis for comparison.

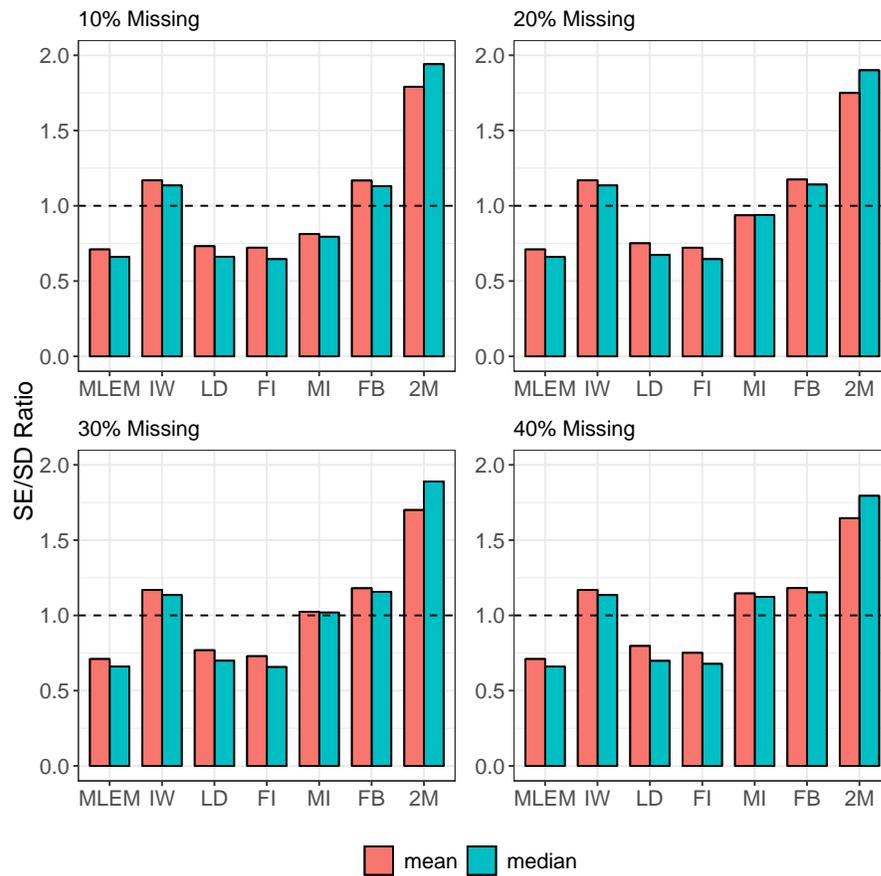


Figure 4.10. Overall means and medians of the SE/SD ratios by ME and MR (Missing data methods). The results from the full data analysis methods are included in the plots as a basis for comparison.

A factorial ANOVA of the standard error bias measures revealed a significant two-way ME×MR interaction for the intercept variances, slope variances, and residual variances, as well as a significant two-way ME×SS interaction for all the variance measures. With the exception of the residual variances, the majority of the variability (η^2 values between 85% and 93%) was explained by SE/SD differences in the method that was used, so these will be the focus of discussion next, followed by a discussion of the interaction effects that were flagged.

Table 4.30

Factorial ANOVA Results for SE/SD Ratios of Estimates of Means (Missing Data Methods)

Factor	μ_{intC1}		μ_{intC2}		μ_{slopC1}		μ_{slopC2}	
	<i>F</i>	η^2	<i>F</i>	η^2	<i>F</i>	η^2	<i>F</i>	η^2
ME	6319	92.4	2713.4	85.8	4515	91.3	6648.7	92.8

Note. Values in bold are discussed in further detail. MM: missing mechanism; MR: missing rate; SP: class separation; SS: sample size; ME: analysis method.

Table 4.31

Factorial ANOVA Results for SE/SD Ratios of Estimates of Variances (Missing Data Methods)

Factor	Ψ_{int}		Ψ_{slop}		Ψ_{12}		Θ	
	<i>F</i>	η^2	<i>F</i>	η^2	<i>F</i>	η^2	<i>F</i>	η^2
ME×SS	119.9	9.0	297.3	18.4	408.1	21.2	327.7	53.8
ME×MR	66.3	6.7	94.4	7.8	-	-	76	16.6
ME	2382.3	59.8	3130.5	64.7	3542	61.2	-	-

Note. Cells containing a “-” indicate non-significance (p -value > 0.05) and/or effect sizes less than 6%. Values in bold are discussed in further detail. MM: missing mechanism; MR: missing rate; SP: class separation; SS: sample size; ME: analysis method.

4.2.4.1 Results of the Main Effects

The SE/SD ratios for each method averaged over all conditions for intercept means and slope means is presented in Table 4.32. With the exception of differences in SE/SD ratios between the LD and FI methods, all other comparisons between methods produced significant differences. Similar to what was observed in the overall summary of the ratios, MLEM-based methods (LD, FI, and MI) produced underestimated standard errors (SE/SD ratios between 0.5 and 0.6) while the FB and 2M methods produced overestimated standard errors (SE/SD ratios between 1.23 and 2.46). Main effects of ME observed for the SE estimates of the variance are not discussed because significant interactions with other conditions were found, which are discussed next.

Table 4.32

Mean SE/SD Ratios Grouped by Analysis Method (ME)

ME	μ_{intC1}	μ_{intC2}	μ_{slopC1}	μ_{slopC2}
LD†	0.53	0.59	0.59	0.53
FI†	0.51	0.59	0.59	0.50
MI	0.78	0.91	0.96	0.88
FB	1.29	1.28	1.24	1.23
2M	2.00	2.05	2.46	2.33

Note. Means are averaged over levels of SP, SS, MR, and MM. All pairwise comparisons were significant at the $\alpha = 0.05$ level unless otherwise noted. †Differences between LD and FI were not significant for all parameters. All p -values were adjusted using the Tukey method.

4.2.4.2 Results of the Interaction Effects

To investigate the two-way interaction effects ($ME \times SS$ and $ME \times MR$), main effect pairwise comparisons between the methods were conducted for the different levels of the second factor. The results of these comparisons are provided in Appendix B, Tables B.3 through B.9. No three-way interactions were flagged. In general, no clear patterns emerged from the pairwise comparisons, although most comparisons were flagged as significantly different.

To make some comparison visible, the SE/SD ratios for any differences that were flagged for significant interaction effects were plotted and are presented in Figure 4.11 to show the $ME \times SS$ interactions and Figure 4.12 to show the $ME \times MR$ interactions. In these plots, the SE/SD ratios of the complete data methods (MLEM and IW) are also provided as a basis of comparison.

Although most comparisons showed significant differences, overall, no clear patterns emerged and the results corroborated what was previously observed. Mainly, MLEM-based method (MLEM, LD, FI, MI) tended to underestimate the standard errors, and Bayesian-based methods (IW, FB, 2M) tended to overestimate standard errors. This can be observed in Figure 4.11, where the intercept variance, slope variance and the intercept slope covariance show the points for MLEM, LD, FI, and MI below the dotted line, and the points for IW, FB, and 2M (with the exception of the intercept variance and residual variance) above the dotted line.

The 2M method did not display any noticeable pattern across parameters. Although the overall summary revealed that the 2M method generally overestimated

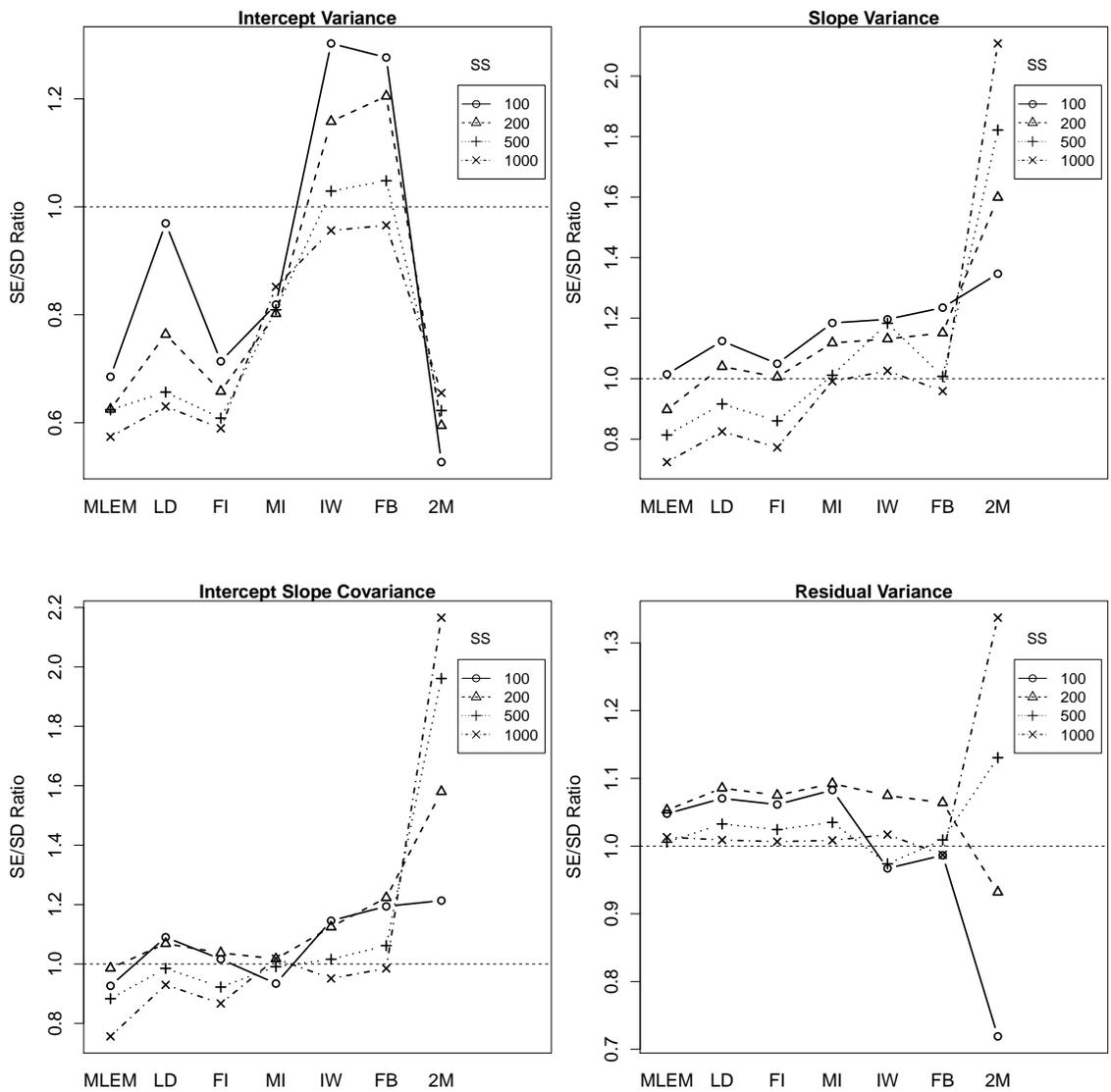


Figure 4.11. Interaction plots for ME \times SS of SE/SD ratios for intercept variance, slope variance, intercept slope covariance, and residual variance.

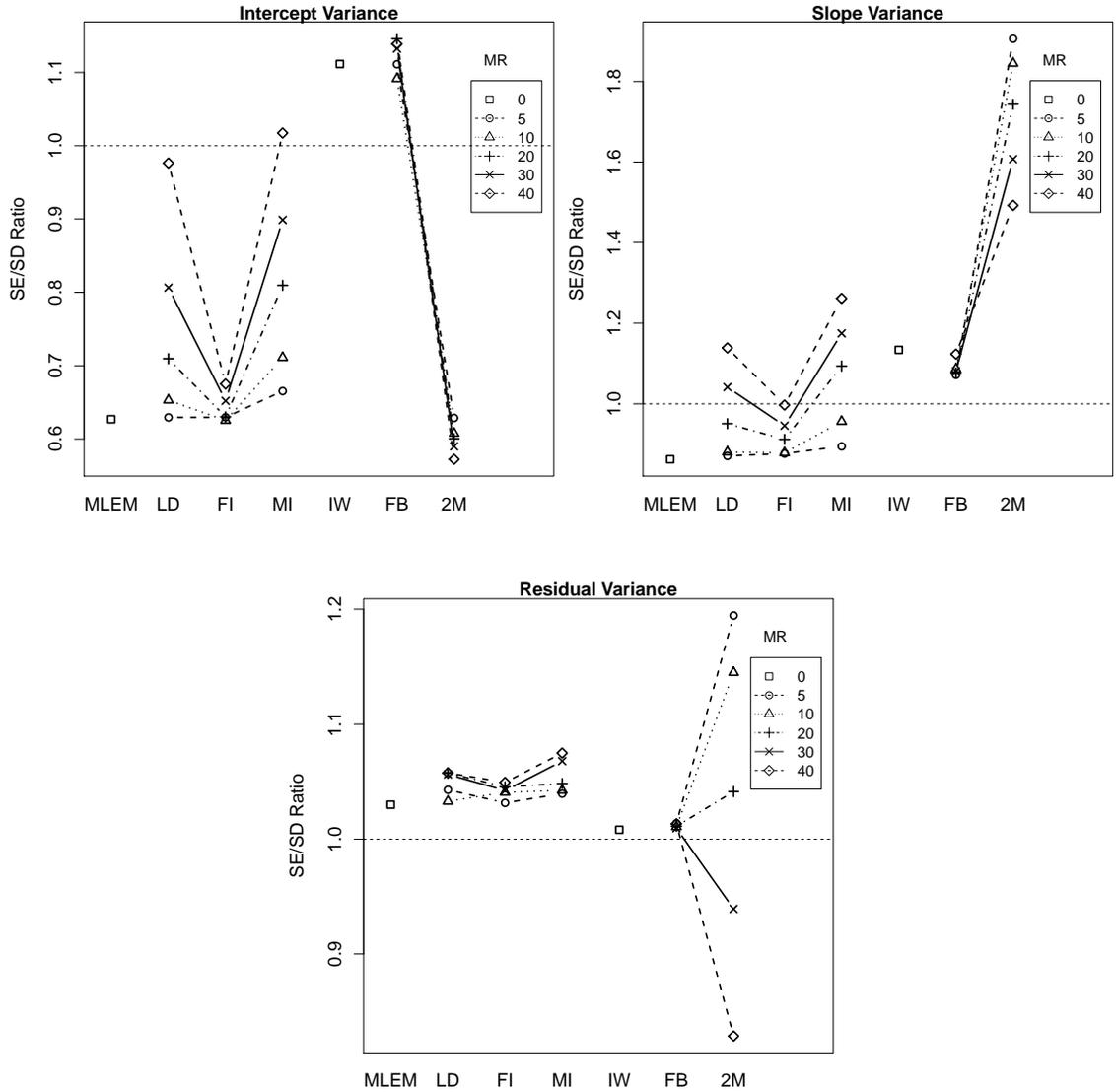


Figure 4.12. Interaction plots for $ME \times MR$ of SE/SD ratios for intercept variance, slope variance, and residual variance.

the estimates of the SEs, the plot of the intercept variance and residual variance for each sample size revealed the opposite; SEs for the 2M method were consistently underestimated for the intercept variance and were over or underestimated for the residual variance depending on the sample size. For example, for the residual variances, at sample sizes of 100 and 200, the SE/SD ratios for the 2M method were 0.72 and 0.93 (underestimated), respectively, but at sample sizes of 500 and 1000, these ratios were 1.13 and 1.34 (overestimated), respectively. A similar lack of pattern was observed in Figure 4.12, where the intercept variance for the 2M method was consistently underestimated, and the residual variance was over or underestimated depending the rate of missing data. To say the least, the estimates of the standard error for the 2M method were quite difficult to make sense of compared to other methods because the direction and severity of the under or over estimation was dependant on the parameter as well as the level of the secondary factor being compared.

An observation that was questionable was that the lower sample sizes (100 and 200) for the MLEM-based methods seemed to correct for the underestimated standard errors, because the complete data method (MLEM) was producing underestimated standard errors to begin with, so this correction may have been an artifact of larger sample sizes inflating the standard errors rather than the methods producing more accurate estimates of the standard error for smaller sample sizes. The FB method produced the closest estimates of the standard errors compared to the complete data method for all parameters.

4.3 Summary of Main Findings

This final section of chapter 4 is provided as a way to summarize the results by enumerating the main findings. The main findings from the simulation study are as follows:

4.3.1 Full Data Analysis

1. Convergence and classification accuracy rates: Using the full data, Bayesian inference methods produced higher convergence rates and classification rates. The inverse Wishart and data-driven prior methods produced the highest rates of convergence among the Bayesian methods, while the classification accuracy was similar among the Bayesian methods.
2. Relative bias: Among the different priors that were tested for the Bayesian method, the improper inverse Wishart priors for the variance parameters produced overall less biased parameters. In general, the variance parameters like the slope variance and intercept slope covariance parameters were the most severely biased when using the MLEM method and Bayesian method using PW and DD priors. The MLEM method also produced severely biased estimates of the slope means and proportion parameters when the class separation not high. Based on these observations, the Bayesian method using IW priors outperformed all other methods for the complete data methods.
3. Standard errors: Overall, standard errors were underestimated for the MLEM

and Bayesian inference using the data-driven MLEM-based priors methods. The improper inverse Wishart and proper inverse Wishart priors produced overestimated standard errors that decreased and reached near-perfect accuracy at sample sizes of 500 and 1000.

4.3.2 Missing Data Analysis

1. Convergence and classification accuracy rates: The convergence rates were generally smaller across the missing data conditions compared to the convergence rates that were observed from the complete data analysis. However, the fully Bayesian and two-stage imputation method maintained convergence rates close to the rates from the full data analysis using IW priors. The listwise deletion and full-information maximum likelihood methods produced even lower convergence rates than when the full data was used. Classification accuracy decreased in general as separation decreased. The fully Bayesian approach maintained the closest accuracy rates to the full data method. Accuracy rates decreased when the two-stage imputation method was used. Using listwise deletion, FIML, and single-stage MI had a very minimal impact on the classification accuracy rates compared the rates produced when using MLEM.
2. Relative bias: Overall, the relative bias of the FB and 2M methods were closest to the relative bias that was observed when missing data was not a factor. The LD, FI and single stage MI method produced greater over bias than MLEM, but this bias decreased as sample size increased. The large relative bias could

be attributed to severely biased slope means and severely biased variances at specific sample sizes and class separation conditions, similar to what was observed in the full data analysis. The estimates of the variance parameters created the most bias for most methods, but the the MLEM-based methods were the most susceptible to the addition of missing data conditions.

3. Standard errors: In general, the MLEM-based methods produced underestimated standard errors by about a factor of 0.5. The FB method produced accurate standard errors with larger samples and slightly overestimated standard errors when sample sizes decreased and varying missing data rate conditions. The 2M method produced overestimated standard errors by almost a factor of 2 across all conditions.

Chapter 5: Discussion

The results from the analysis without any of the missing data conditions was as expected in that the ML method was more susceptible to non-convergence than the Bayesian method with different priors. The use of data-driven priors produced lower convergence rates than when noninformative improper priors and proper priors were used, but this was expected because the priors had to be obtained from an initial ML run. The data-driven priors were still slightly higher than the results from the ML method because the convergence criteria for obtaining the priors from the ML estimates were more relaxed. For example, it was not of concern if the standard errors from the ML estimation were very large. These types of results were omitted from the ML results. The higher convergence rates observed from the noninformative and proper prior methods were expected because ranges were not being restricted on the posterior distributions of the parameters by means of any specific prior.

The classification accuracy rates did improve with increased sample sizes and increased class separation, and the Bayesian methods did improve the accuracy rates slightly over ML-EM (approximately 2-12% increase). These accuracy rates are related to the estimates of the proportion and variance parameters, which were shown to be biased for the ML-EM method. If the class proportion was incorrect,

and variance parameters were biased, then the susceptibility of individuals being placed into the wrong class would be greater.

In general, bias across parameters was found to be greatest when using the MLEM method, especially for the variance estimates, which were generally underestimated. This was actually consistent with the growth modeling literature, which has pointed to the issue that even with complete data, some maximum likelihood methods under a growth modeling SEM framework will underestimate factor variances and standard errors (Browne & Draper, 2006), even though large sample sizes should correct for these discrepancies. In this regard, Browne and Draper (2006) and also McNeish and Harring (2017), who conducted a similar study with a focus on samples of less than 100, recommend using a restricted version of the ML method to obtain unbiased parameter estimates. This finding was also partially supported by the results presented in Depaoli (2013), which show that slope variances and sometimes the intercept and intercept slope variances were consistently underestimated when using the standard ML-EM method (the default in Mplus) was used.

The discrepancies found among the Bayesian method using different prior specifications was different than what was found in the literature. In general, the literature on choosing appropriate priors in the context of GMMs is still mainly unclear, and the results that were observed from the full data analysis were in some way a testament to this. In general, the relative bias of the parameters were the least biased when noninformative (improper inverse Wishart) priors were used, which was partially contrary to what Depaoli (2013) and McNeish (2016)¹ found. Actually, the

¹Prior specifications tested were for small sample sizes (less than 100) in the context of latent

superior performance of the noninformative inverse Wishart (diffuse) priors were more in line with suggestions by [Browne and Draper \(2006\)](#), although the priors tested in that particular study were in the context of longitudinal multilevel models. In any case, the purpose of the current study was not to find the priors that produced the least biased parameter estimates, so using the noninformative default Mplus priors was justifiably good enough for the purposes of the current study. In hindsight, however, although unrealistic, a better approach would have been to use the population values to inform the hyperparameters so that any effects of missing data on GMM estimation could be isolated.

Overall standard errors were underestimated for the MLEM method, generally overestimated for the noninformative (improper inverse Wishart) and proper identity inverse Wishart prior method, and generally underestimated for the data-driven prior method. These results were actually consistent with what was observed in [McNeish \(2016\)](#), where the Bayesian MCMC methods using noninformative priors produced overly wide coverage intervals (overestimated standard errors) while FIML consistently had coverage intervals that were too short (underestimated standard errors). The reverse impact that the priors had on the standard errors of the intercept means was interesting (standard error ratios converged to 1 as sample sizes increased using noninformative priors, while they deviated upwards from 1 as sample sizes increased using data driven priors). In addition, standard errors were severely underestimated for the intercept means when using data driven priors, which may

growth models, which may not have been appropriate for larger samples and in mixture settings, where the number of parameters to account for is greater.

have been an artifact of the priors specified for the variance parameters since priors were not specified for any of the means.

The purpose of analyzing the full-data methods using different prior specifications was to find an appropriate basis of comparison for the missing data methods, which were grounded on MLEM (LD, FI, ML) or Bayesian inference (FB, 2M). Unfortunately, this investigation was not exhaustive or perfect in its execution, and a more thorough investigation of different priors would have been ideal. However, since this was not the main focus of the study, the analysis of the missing data methods was conducted in light of the results that were obtained from testing these suggested prior specifications, which in the end, prompted the use of the noninformative priors for the Bayesian methods.

The accuracy rates across methods when missing data conditions were introduced were mainly indistinguishable for very poor class separations. The FB method, however, was clearly producing the highest accuracy rates with poor, moderate, and high class separation conditions while the other methods produced similar accuracy rates. The lower accuracy rates that were observed for the 2M method were somewhat surprising given that the parameter estimates were quite comparable to the FB method. In retrospect, this may have been an artifact of how these accuracy rates were collected for the 2M method during the simulation, which was different from how accuracy rates were collected for non-imputation methods. This observation warrants a separate investigation.

With the introduction of missing data conditions, there was a downward bias of the parameter estimates across methods with increased missing data rates. This

bias was only attributed to a few parameters, however, as could be seen in the large discrepancies between the means and medians of the relative bias measures. The variable-by-variable comparison indicated severe problems when estimating the intercept variances and slope variances at sample sizes of 100 and 200. While the FB and 2M methods were able to keep the bias controlled with the introduction of some missing data conditions, LD, FI, and MI methods further biased the estimates.

Furthermore, the standard errors were consistently overestimated for the 2M method, while in general, the FB method produced SE to SD ratios closest to 1, albeit overestimated standard errors were observed when sample sizes were at 100 and 200. The LD, MI, and FI methods, which were grounded on FIML-EM, produced underestimated standard errors.

LD and MI methods performed relatively poorly compared to FI, FB and 2M, which was consistent with [Enders and Gottschall \(2011\)](#), who suggested avoiding MI for mixture models altogether. Based on research by [Harel et al. \(2013\)](#), however, 2M seemed to be a viable alternative to MI.

Another important point that should not go ignored is that the differences when imposing an MCAR as supposed to a CMAR+ missingness mechanism was minimal, and whether the missing data approaches did anything to improve estimation when missingness was the more severe CMAR+ missingness was not very clear from the results. With multiple group models, [Enders and Gottschall \(2011\)](#) as well as [Sterba \(2016\)](#) suggested that the missingness mechanism was inconsequential to how poorly the MI method would perform. In this regard, it made sense that there was very little difference observed between the estimates from the two missingness

mechanisms. However, these results are far from conclusive because the MLEM estimates were already producing poor estimates to begin with, making the MI method incomparable to the 2M method.

5.1 Recommendations

Based on the findings from the simulation study, a few recommendations can be made. First, the Bayesian inference and two-stage imputation methods may serve as viable alternatives to MLEM and single-stage MI in terms of producing less biased parameter estimates for GMMs. However, this comparison warrants further investigation given that the full-data analysis step of the single-stage MI method was conducted using MLEM, which produced relatively more biased estimates compared to the Bayesian method. Although choosing priors can be an intricate step, the advantages in terms of producing higher convergence rates, higher classification accuracy rates, less biased parameter estimates, and more accurate standard errors (for the Bayesian approach) may be worth the additional analysis. One disadvantage of using the two-stage imputation method is that although the tools are available in Mplus, the setup is more involved and requires the researcher to take additional steps outside of Mplus. The steps that were taken to conduct the two-stage imputation method for this study are outlined in Appendix C.

Even with the use of noninformative flat priors, the fully Bayesian approach was superior in regards to the evaluation criteria that were considered, especially with the estimation of the standard errors. While the two-stage imputation was

relatively better than single-stage imputation and FIML in terms of recovering parameter estimates, the standard errors were consistently overestimated, which in practical applications would have led to inflated Type II error rates.

Nonetheless, these recommendations should be taken with caution because several problems arose during the simulation study, which could have compromised some of the results. In addition, there were several factors of the study that kept the results from this study from being generalized to all kinds of situations. These issues and other related limitations are discussed next.

5.2 Limitations and Future Studies

The first part of the simulation study was an attempt to find a useful, non-biasing, and realistic set of parameter priors. This precursory analysis was far from a comprehensive sensitivity analysis that may have been useful. Defining priors through a sensitivity analysis is a popular topic of interest in Bayesian literature because there are several ways to define priors. Three prior specifications found in the literature were used in the current study, one of them being the use of the MLEM estimates for the data-driven method, which a popular way of defining priors. Although unrealistic, it may have been best to forgo this step and use the population values as priors in order to isolate the effects of missing data conditions alone. However, this also poses the problem that the results from the Bayesian methods would then produce overly favorable results, perhaps enough to damper any effects of the missing data conditions. In this regard, using the most naive approaches for

both (using the IW priors, which was the default Mplus setting) may have been appropriate. Nevertheless, a useful future study would be to conduct a sensitivity analysis using various prior specification methods that are currently available in the context GMMs with missing data.

The method of choosing priors becomes even murkier when implementing a multiple imputation approach. For example, during the study it became evident that the choice existed between using MLEM and Bayesian inference for analyzing the multiply imputed data. This choice of analysis presented itself at both stages of the imputation for two-stage imputation, and at the first stage for the single-stage imputation method. After imputation, there was the additional choice of analyzing the datasets with imputed data using MLEM or Bayesian inference. For the current study, the default ML method was used for single-stage MI after the imputation step. Unfortunately, this will have produced incomparable results between the single-stage and two-stage imputation. This was an obvious limitation and would be a topic to investigate in the near future.

Using Mplus for the simulation presented problems due to convergence issues. For example, for numerous replications, the estimates of the variances needed to be monitored because Mplus would terminate successfully without flagging any negative variances. For some replications, Mplus even allowed extremely large standard errors to go unnoticed, so results from such replications needed to be filtered out of the results. Convergence issues were even more problematic when running the multiple imputation methods, where it became unclear as to when and how many of the estimations from the imputed datasets produced sensible results. To ensure

that all estimations for imputations converged and produced good estimated values, extra imputations were created and estimated until the desired number of imputations per replication were obtained. The same procedure was used for the two-stage imputation method, where additional imputations at both stages were created and estimated until the total number of desired imputations were collected.

The type of model that was tested was also a limitation in itself. Some other conditions of the model that are typically manipulated but were fixed for the current study are: total number of classes, varying class proportions, total number of data collection points (timepoints), structure of the intervals between timepoints, structure of the variance components of the model, the inclusion of time-invariant/time-varying covariates, and the overall shape of the model. For example, simulation studies involving GMMs have regularly tested assumptions and data conditions with more complex models such as higher order models like quadratic growth models or piece-wise models (see, e.g., [Ning & Luo, 2017](#); [Wu, Zumbo, & Siegel, 2011](#)), otherwise known as knot models (see, e.g., [Kohli, 2011](#)). These models are often more complex and the effect of missing data on these types of GMMs has gone largely unexplored.

The GMM model that was tested in this study was also highly restrictive in the sense that between class random effects were kept invariant, which has been noted to be highly untenable in practice ([Bauer & Curran, 2003](#)). Setting random effects to be invariant across classes is actually an Mplus default, which has been set as so in order to lessen the computational burdens of estimation, given that any specific invariances across class variance parameters may not be of interest to most applied

researchers. In contrast, more restrictive models, like the latent class growth model (Nagin, 1999), which does not impose a variance covariance structure of the growth parameters, is another popular approach for when individual growth trajectories are not of particular interest. An investigation on how these models would perform under missing data conditions could be another future study.

The number of timepoints in the current study was also fixed at 4 and these timepoints were also fixed to have the same distance across timepoints. This type of data collection is rarely the norm, as the number of measurements and the intervals when these measurements are collected at each timepoint often vary. This data design often leads to different time-residual variance covariance structures that need to be addressed. Simulation studies involving longitudinal models, however, have shown that imputation methods differ when the times between timepoints vary. Several variations of the JM approach and FCS approach have been developed for this purpose and investigating these methods under GMMs would be a useful study.

Another crucial step of general finite mixture modeling that was completely overlooked in the study was the class enumeration process, which is in itself an intricate part of the methodology. This process involves comparing the fit indices of models with a varying number of classes, in an effort to identify the correct number of subpopulations. This is an important step because as Bauer and Curran (2003) demonstrated, even data from a homogeneous population may exhibit the presence of multiple classes when the population is non-normal. Bauer (2007) also enumerated several violations of the model that can adversely affect class enumeration, including the presence of missing data that is non-ignorable. This research was mainly using

MLEM and fit indices available for MLEM, but fit indices for Bayesian methods are different (see, e.g., [Ning & Luo, 2017](#)) and the literature is unclear on how these indices perform when different missing data conditions plague the data. In addition, there are no established procedures for evaluating fit when using multiple imputation, making the class enumeration process even more unclear for imputation methods.

The number of imputations for the two-stage imputation method is also unclear. [McGinniss and Harel \(2016\)](#) conducted a simulation for a three stage model, but only on the means. How the estimates of the variances are affected by the number of first and second stage imputations is still unclear. This hole in the literature leaves the question of whether or not estimates of the variances would have been better if more imputations were done at each stage. In addition, it is unclear if these suggested number of imputations are affected by different class separations.

Finally, several useful missing data handling methods like pattern mixture models and predictive mean matching were not considered in the current study. For example, [Bauer and Curran \(2003\)](#) suggested using a pattern mixture model for GMMs when any missing data are assumed to be MAR. There is also a growing literature on the use of non-parametric tree imputation techniques using classification and regression trees ([Breiman, 2017](#)) or random forest models ([Ho, 1995](#)). These alternative methods for handling missing data in the growth mixture context could be useful but the extent of their usefulness would need to be further investigated.

Although the popularity of GMMs has grown in the last decade, the same cannot be said about methods to handle missing data, despite the reality that most

realistic datasets are rife of missing data. This study was nowhere exhaustive of all the possible scenarios that one might encounter when dealing with missing data with GMMs. However, the main goal was not to be exhaustive, but rather, to investigate a few conditions that were previously considered in similar studies, and more importantly, to begin a discussion on better ways one might be able to deal with missing data in the context of GMMs.

Appendix A: Complete Data Methods Relative Bias and SE/SD Ratio Tables

Table A.1

Method	Variable	N = 100					N = 200					N = 500					N = 1000					
		vpoor	poor	mod	high	vpoor	poor	mod	high	vpoor	poor	mod	high	vpoor	poor	mod	high	vpoor	poor	mod	high	
MLEM	μ_{intC1}	0.10	0.07	0.05	0.03	0.10	0.08	0.06	0.02	0.06	0.04	0.02	0.02	0.06	0.04	0.02	0.02	0.06	0.04	0.01	0.01	0.01
IW		0.03	0.01	-0.01	-0.03	0.03	0.01	-0.01	-0.02	0.02	0.01	-0.01	-0.01	0.01	0.01	-0.01	-0.01	0.03	0.01	0.01	-0.01	-0.02
PW		0.05	0.03	0.01	0.00	0.05	0.03	0.01	0.00	0.04	0.02	0.00	0.00	0.04	0.02	0.00	0.00	0.03	0.02	0.00	0.00	-0.01
DD		0.03	0.02	0.00	-0.02	0.03	0.02	0.00	-0.01	0.02	0.01	0.02	0.01	-0.01	0.02	0.01	-0.01	0.02	0.00	0.00	-0.02	-0.02
MLEM	μ_{intC2}	-0.05	-0.04	-0.01	0.01	-0.04	-0.03	0.00	0.01	-0.03	-0.01	-0.03	0.00	0.01	-0.03	0.00	0.02	-0.03	-0.03	-0.03	0.00	0.03
IW		-0.03	-0.01	0.01	0.04	-0.02	-0.01	0.02	0.03	0.03	0.02	0.03	0.00	0.03	0.00	0.02	0.02	-0.02	-0.02	0.00	0.00	0.02
PW		-0.04	-0.03	-0.01	0.01	-0.04	-0.03	-0.01	0.01	-0.03	-0.03	-0.02	0.00	0.01	-0.03	-0.02	0.00	0.01	0.01	0.01	0.01	0.01
DD		-0.09	-0.05	-0.03	0.00	-0.17	-0.07	-0.08	-0.02	-0.14	-0.20	-0.11	-0.02	-0.26	-0.25	-0.20	-0.05	-0.02	-0.02	-0.01	0.01	0.01
MLEM	μ_{stopC1}	-0.38	-0.33	-0.34	-0.32	-0.32	-0.27	-0.23	-0.15	-0.30	-0.23	-0.16	-0.13	-0.28	-0.22	-0.20	-0.20	-0.15	-0.07	-0.06	-0.02	-0.15
IW		-0.11	-0.05	-0.05	-0.06	-0.07	-0.02	-0.03	-0.02	-0.05	-0.01	-0.02	-0.02	-0.07	-0.01	0.00	0.00	-0.02	-0.01	0.00	0.00	-0.01
PW		-0.11	-0.07	-0.07	-0.07	-0.08	-0.04	-0.07	-0.07	-0.08	-0.07	-0.06	-0.04	-0.09	-0.06	-0.07	-0.06	-0.07	-0.09	-0.06	-0.07	-0.06
DD		-0.14	-0.09	-0.11	-0.11	-0.08	-0.07	-0.06	-0.06	-0.07	-0.01	-0.05	-0.04	-0.05	0.00	0.00	-0.02	-0.02	-0.05	0.00	-0.02	-0.02
MLEM	μ_{stopC2}	0.19	0.24	0.29	0.35	0.15	0.26	0.24	0.24	0.22	0.30	0.25	0.23	0.12	0.24	0.24	0.25	0.12	0.24	0.11	0.12	0.25
IW		0.10	0.17	0.16	0.15	0.07	0.14	0.14	0.14	0.05	0.11	0.13	0.13	0.08	0.11	0.12	0.13	0.05	0.11	0.08	0.11	0.13
PW		0.08	0.17	0.17	0.17	0.08	0.15	0.17	0.17	0.07	0.16	0.16	0.16	0.07	0.16	0.16	0.16	0.08	0.16	0.16	0.16	0.16
DD		0.11	0.17	0.19	0.19	0.07	0.16	0.15	0.16	0.06	0.13	0.15	0.14	0.07	0.13	0.16	0.14	0.07	0.13	0.13	0.16	0.14
MLEM	Ψ_{int}	-0.26	-0.23	-0.09	0.15	-0.25	-0.21	-0.02	0.17	-0.18	-0.14	0.02	0.20	-0.15	-0.11	0.09	0.28	-0.15	-0.11	-0.11	0.09	0.28
IW		-0.06	-0.02	0.17	0.47	-0.14	-0.07	0.14	0.37	-0.12	-0.08	0.12	0.32	-0.15	-0.06	0.14	0.29	-0.06	-0.15	-0.06	0.14	0.29
PW		-0.40	-0.36	-0.27	-0.38	-0.38	-0.36	-0.21	-0.03	-0.29	-0.26	-0.09	0.06	-0.25	-0.18	0.00	0.13	-0.18	-0.25	-0.18	0.00	0.13
DD		-0.19	-0.16	0.04	0.23	-0.20	-0.15	0.01	0.21	-0.16	-0.10	0.07	0.24	-0.12	-0.05	0.17	0.26	-0.12	-0.05	0.17	0.17	0.26
MLEM	Ψ_{stop}	-0.16	-0.13	-0.16	-0.22	-0.16	-0.19	-0.16	-0.18	-0.25	-0.16	-0.17	-0.21	-0.23	-0.16	-0.19	-0.23	-0.23	-0.11	-0.05	-0.06	-0.07
IW		0.01	0.03	0.04	0.01	-0.06	0.00	-0.04	-0.04	-0.07	-0.02	-0.07	-0.06	-0.11	-0.05	-0.06	-0.07	-0.11	-0.05	-0.06	-0.06	-0.07
PW		-0.13	-0.16	-0.17	-0.21	-0.15	-0.14	-0.22	-0.24	-0.18	-0.19	-0.19	-0.19	-0.20	-0.19	-0.21	-0.19	-0.20	-0.19	-0.19	-0.21	-0.19
DD		-0.18	-0.15	-0.21	-0.19	-0.10	-0.12	-0.13	-0.16	-0.16	-0.09	-0.15	-0.12	-0.14	-0.08	-0.13	-0.12	-0.14	-0.08	-0.13	-0.13	-0.12
MLEM	Ψ_{12}	-0.09	-0.31	-0.48	-0.57	0.18	0.08	-0.45	-0.70	0.32	0.07	-0.36	-0.40	0.29	0.00	-0.32	-0.42	0.29	0.00	0.00	-0.32	-0.42
IW		0.29	-0.01	-0.34	-0.62	0.46	0.12	-0.25	-0.51	0.37	0.10	-0.20	-0.37	0.49	0.13	-0.28	-0.38	0.49	0.13	0.13	-0.28	-0.38
PW		0.85	0.78	0.67	0.64	0.87	0.83	0.66	0.53	0.84	0.73	0.43	0.41	0.88	0.59	0.35	0.24	0.88	0.59	0.35	0.35	0.24
DD		0.31	0.22	0.01	-0.14	0.37	0.16	-0.07	-0.13	0.46	0.13	-0.12	-0.20	0.38	0.06	-0.27	0.20	0.38	0.06	-0.27	-0.27	0.20
MLEM	Θ	-0.04	-0.02	-0.02	-0.02	-0.01	-0.01	-0.01	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01
IW		-0.01	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PW		0.01	0.02	0.03	0.03	0.01	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
DD		-0.01	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MLEM	π_{c1}	-0.19	-0.14	-0.11	-0.02	-0.24	-0.20	-0.21	-0.08	-0.10	-0.08	-0.04	-0.01	-0.17	-0.07	-0.06	-0.01	-0.17	-0.07	-0.07	-0.06	-0.01
IW		-0.01	0.00	-0.01	-0.01	-0.02	-0.01	-0.03	-0.01	-0.02	-0.02	-0.02	-0.01	-0.03	-0.03	-0.02	-0.01	-0.03	-0.03	-0.03	-0.02	-0.01
PW		-0.02	-0.02	-0.02	-0.01	-0.02	-0.03	-0.04	-0.02	-0.03	-0.03	-0.03	-0.02	-0.02	-0.02	-0.01	-0.02	-0.02	-0.02	-0.02	-0.01	-0.02
DD		0.05	0.03	0.02	0.00	0.11	0.02	0.05	0.01	0.11	0.18	0.09	0.02	0.23	0.23	0.21	0.07	0.23	0.23	0.23	0.21	0.07
MLEM	π_{c2}	0.19	0.14	0.11	0.02	0.24	0.20	0.21	0.08	0.10	0.08	0.04	0.01	0.17	0.07	0.06	0.01	0.17	0.07	0.07	0.06	0.01
IW		0.01	0.00	0.01	0.01	0.02	0.01	0.03	0.01	0.02	0.02	0.02	0.01	0.03	0.03	0.02	0.01	0.03	0.03	0.03	0.02	0.01
PW		0.02	0.02	0.02	0.01	0.02	0.03	0.04	0.02	0.03	0.02	0.03	0.02	0.01	0.02	0.03	0.02	0.01	0.03	0.02	0.01	0.02
DD		-0.05	-0.03	-0.02	0.00	-0.11	-0.02	-0.05	-0.01	-0.11	-0.18	-0.09	-0.02	-0.23	-0.23	-0.21	-0.07	-0.23	-0.23	-0.23	-0.21	-0.07

Note. MLEM = full data ML, IW = full data Bayesian with improper IW priors, PW = full data Bayesian with proper IW priors, DD = full data Bayesian with data-driven priors. vpoor is MD = 0.5, poor is MD = 1.0, mod is MD = 0.5, high is MD = 0.5.

Table A.2

SE/SD Ratios for All Parameters by All Conditions Using Full Data Methods

Method	Variable	N = 100					N = 200					N = 500					N = 1000									
		vpoor	poor	mod	high	high	vpoor	poor	mod	high	high	vpoor	poor	mod	high	high	vpoor	poor	mod	high	high	vpoor	poor	mod	high	high
MLEM IW PW DD	$\mu_{int_{c1}}$	0.50	0.47	0.47	0.51	0.48	0.52	0.48	0.48	0.48	0.48	0.51	0.52	0.50	0.56	0.56	0.45	0.49	0.50	0.56	0.56	0.45	0.49	0.46	0.47	0.47
		1.36	1.55	1.41	1.41	1.30	1.20	1.20	1.17	1.30	1.02	1.02	1.02	1.14	1.02	1.02	0.90	1.22	1.09	1.14	1.02	0.90	1.22	0.99	0.78	0.78
		0.96	0.96	1.05	0.93	0.92	0.83	0.95	1.01	0.92	0.93	0.93	0.83	0.83	0.83	0.83	0.85	0.99	0.83	0.83	0.83	0.85	0.99	0.84	0.74	0.74
		1.07	1.13	1.15	1.12	1.00	1.12	1.02	1.07	1.00	1.34	1.34	1.51	1.14	1.05	1.05	1.62	1.59	1.70	1.70	1.70	1.62	1.59	1.70	0.98	0.98
MLEM IW PW DD	$\mu_{int_{c2}}$	0.68	0.55	0.58	0.78	0.61	0.73	0.70	0.65	0.61	0.74	0.74	0.66	0.64	0.67	0.67	0.67	0.55	0.58	0.58	0.58	0.67	0.55	0.58	0.71	0.71
		1.54	1.56	1.66	1.78	1.47	1.47	1.40	1.49	1.55	1.33	1.33	1.21	1.33	1.13	1.05	1.05	1.38	1.10	1.10	1.05	1.05	1.38	1.10	0.82	0.82
		1.13	1.15	1.17	1.21	1.00	1.13	1.21	1.14	1.00	1.02	1.02	0.97	0.92	0.84	1.08	1.08	0.87	0.94	0.78	0.78	1.08	0.87	0.94	0.78	0.78
		0.13	0.18	0.20	0.24	0.17	0.10	0.12	0.13	0.17	0.09	0.09	0.09	0.10	0.15	0.09	0.09	0.09	0.10	0.10	0.12	0.09	0.09	0.10	0.12	0.12
MLEM IW PW DD	$\mu_{stop_{c1}}$	0.49	0.50	0.51	0.54	0.59	0.62	0.59	0.59	0.59	0.59	0.59	0.62	0.57	0.55	0.54	0.54	0.54	0.58	0.58	0.61	0.54	0.54	0.58	0.61	0.61
		1.31	1.52	1.35	1.26	1.28	1.28	1.49	1.13	1.24	1.27	1.27	1.12	1.13	1.26	1.07	1.07	1.26	1.17	1.17	1.14	1.07	1.26	1.17	1.14	1.14
		1.07	1.00	1.03	1.09	0.97	1.01	1.10	0.98	0.97	1.07	1.07	1.00	1.06	0.88	1.07	1.07	0.93	1.09	0.77	0.77	1.07	0.93	1.09	0.77	0.77
		1.13	0.97	1.05	0.91	0.94	0.96	0.92	0.91	0.94	0.95	0.95	0.79	0.79	0.92	0.82	0.82	0.82	0.77	0.81	0.81	0.82	0.82	0.77	0.81	0.81
MLEM IW PW DD	$\mu_{stop_{c2}}$	0.53	0.56	0.52	0.57	0.50	0.67	0.59	0.52	0.50	0.51	0.51	0.53	0.48	0.49	0.73	0.73	0.59	0.61	0.56	0.56	0.62	0.59	0.61	0.56	0.56
		1.15	1.25	1.46	1.43	1.25	1.32	1.21	1.55	1.25	1.29	1.29	1.47	1.40	1.12	1.06	1.06	1.23	0.94	0.94	1.07	1.06	1.23	0.94	1.07	1.07
		1.09	0.97	1.17	0.97	1.23	1.26	1.21	1.25	1.23	1.20	1.20	1.07	1.12	0.94	0.94	0.94	0.88	1.00	0.83	0.83	0.94	0.99	1.00	0.83	0.83
		1.01	1.02	0.90	1.21	1.15	1.04	1.05	1.11	1.15	1.37	1.37	0.99	0.99	1.16	1.06	1.06	1.07	0.93	0.76	0.76	1.06	1.07	0.93	0.76	0.76
MLEM IW PW DD	Ψ_{int}	0.75	0.71	0.69	0.60	0.64	0.66	0.60	0.64	0.59	0.70	0.70	0.66	0.59	0.54	0.62	0.62	0.59	0.59	0.49	0.49	0.62	0.59	0.59	0.49	0.49
		1.12	1.21	1.40	1.49	1.24	1.00	1.13	1.24	1.26	1.10	1.10	0.99	1.08	0.95	1.05	1.05	1.06	0.99	0.72	0.72	1.05	1.06	0.99	0.72	0.72
		0.98	1.06	0.94	0.99	0.84	0.84	0.94	0.96	0.87	0.85	0.85	0.82	0.79	0.73	0.88	0.88	0.80	0.79	0.68	0.68	0.88	0.80	0.79	0.68	0.68
		0.84	0.79	0.92	1.00	0.70	0.70	0.78	0.90	0.84	0.90	0.90	0.87	0.74	0.87	0.79	0.79	0.69	0.83	0.67	0.67	0.79	0.69	0.83	0.67	0.67
MLEM IW PW DD	Ψ_{stop}	1.01	0.92	1.03	1.09	0.85	0.95	0.89	0.85	0.90	0.86	0.91	0.73	0.73	0.75	0.73	0.73	0.73	0.69	0.75	0.75	0.73	0.73	0.69	0.75	0.75
		1.24	1.15	1.23	1.16	1.14	1.23	1.07	1.14	1.08	1.21	1.07	1.23	1.23	1.22	0.97	0.97	1.15	1.01	0.98	0.98	0.97	1.15	1.01	0.98	0.98
		1.20	1.29	1.12	1.16	1.22	1.15	1.29	1.22	1.31	1.14	1.19	1.25	1.11	1.02	0.89	0.89	0.89	0.99	0.76	0.76	0.89	0.89	0.99	0.76	0.76
		0.88	0.84	0.93	0.93	0.82	0.85	0.77	0.83	0.82	0.99	0.71	0.67	0.87	0.87	0.70	0.70	0.66	0.59	0.62	0.62	0.66	0.59	0.62	0.62	0.62
MLEM IW PW DD	Ψ_{12}	1.04	0.99	0.87	0.81	1.00	1.07	0.99	1.00	0.88	1.06	0.88	0.88	0.77	0.81	0.86	0.86	0.78	0.71	0.67	0.67	0.86	0.78	0.71	0.67	0.67
		1.17	1.21	1.12	1.09	1.08	1.19	1.15	1.08	1.07	1.08	1.01	1.06	0.91	0.90	0.90	0.90	1.05	0.98	0.88	0.88	0.90	1.05	0.98	0.88	0.88
		1.12	1.06	1.10	1.09	1.09	1.13	1.43	1.27	1.00	1.14	1.11	1.11	0.88	1.06	0.79	0.79	0.83	0.71	0.71	0.71	1.06	0.79	0.83	0.71	0.71
		1.25	1.25	1.08	1.05	1.05	1.15	1.18	1.05	1.06	1.07	0.98	0.77	0.98	0.98	0.85	0.85	0.76	0.80	0.72	0.72	0.85	0.76	0.80	0.72	0.72
MLEM IW PW DD	Θ	1.11	1.04	1.02	1.02	1.07	1.04	1.02	1.07	1.09	0.96	0.99	1.00	1.07	0.99	1.04	0.99	1.05	0.98	1.04	1.04	0.99	1.05	0.98	1.04	1.04
		0.97	1.00	0.94	0.95	1.08	1.08	1.08	1.04	1.10	0.98	0.97	0.97	0.97	0.97	1.00	1.00	1.00	1.02	1.04	1.04	1.00	1.00	1.02	1.04	1.04
		0.96	0.97	0.95	0.95	1.05	1.05	1.13	1.09	1.06	0.99	0.97	0.95	0.97	0.95	0.97	0.95	0.99	1.02	1.05	1.05	0.99	0.99	1.02	1.05	1.05
		0.99	0.98	1.01	0.99	1.06	1.06	1.06	1.00	0.99	0.99	1.01	0.99	0.99	0.99	0.96	0.96	0.98	0.98	1.04	1.04	0.99	0.98	0.98	1.04	1.04

Note. MLEM = full data ML, IW = full data Bayesian with improper IW priors, PW = full data Bayesian with proper IW priors, DD = full data Bayesian with data-driven priors. vpoor is MD = 0.5, poor is MD = 1.0, mod is MD = 0.5, high is MD = 0.5.

Appendix B: Missing Data Methods SE/SD Ratios and Pairwise Comparison Tables

Table B.1

SE/SD Ratios For Mean Estimates Grouped by ME and SS (Missing Data Methods)

Method	Variable	N = 100	N = 200	N = 500	N = 1000
MLEM	μ_{intC_1}	0.488	0.490	0.522	0.488
IW		1.432	1.181	1.067	1.432
LD		0.514	0.537	0.559	0.514
FI		0.507	0.517	0.582	0.507
MI		0.684	0.732	0.841	0.684
FB		1.425	1.359	1.239	1.425
2M		1.852	1.947	2.099	1.852
MLEM		μ_{intC_2}	0.648	0.674	0.677
IW	1.636		1.476	1.252	1.636
LD	0.532		0.607	0.580	0.532
FI	0.592		0.581	0.593	0.592
MI	0.872		0.881	0.954	0.872
FB	1.475		1.377	1.163	1.475
2M	1.939		2.210	1.973	1.939
MLEM	μ_{stopC_1}		0.509	0.595	0.584
IW		1.360	1.285	1.196	1.360
LD		0.554	0.577	0.624	0.554
FI		0.519	0.617	0.654	0.519
MI		0.779	0.945	1.049	0.779
FB		1.398	1.198	1.231	1.398
2M		2.269	2.335	2.551	2.269
MLEM		μ_{stopC_2}	0.546	0.568	0.501
IW	1.321		1.333	1.321	1.321
LD	0.557		0.541	0.513	0.557
FI	0.520		0.492	0.476	0.520
MI	0.816		0.897	0.877	0.816
FB	1.376		1.413	1.108	1.376
2M	2.157		2.437	2.230	2.157

Table B.2

SE/SD Ratios For Variance Estimates Grouped ME and SS (Missing Data Methods)

Method	Variable	N = 100	N = 200	N = 500	N = 100
MLEM	Ψ_{int}	0.685	0.625	0.624	0.685
IW		1.303	1.158	1.029	1.303
LD		0.890	0.733	0.646	0.890
FI		0.700	0.652	0.608	0.700
MI		0.819	0.802	0.809	0.819
FB		1.277	1.205	1.048	1.277
2M		0.527	0.595	0.623	0.527
MLEM	Ψ_{slop}	1.014	0.897	0.813	1.014
IW		1.196	1.131	1.183	1.196
LD		1.077	1.010	0.914	1.077
FI		1.020	0.983	0.860	1.020
MI		1.184	1.118	1.011	1.184
FB		1.235	1.151	1.007	1.235
2M		1.347	1.599	1.821	1.347
MLEM	Ψ_{12}	0.926	0.986	0.883	0.926
IW		1.145	1.124	1.016	1.145
LD		1.153	1.107	1.002	1.153
FI		1.105	1.077	0.920	1.105
MI		0.934	1.017	0.991	0.934
FB		1.194	1.223	1.062	1.194
2M		1.213	1.580	1.961	1.213
MLEM	Θ	1.048	1.053	1.006	1.048
IW		0.967	1.075	0.974	0.967
LD		1.034	1.066	1.035	1.034
FI		1.043	1.058	1.022	1.043
MI		1.083	1.092	1.035	1.083
FB		0.986	1.064	1.009	0.986
2M		0.719	0.932	1.131	0.719

Table B.3

Main Effect Pairwise Comparisons Corresponding to $ME \times SS$ Interaction of Intercept Variances for SE/SD Ratios

Contrast (ME)	SS	$\bar{X}_{ME=j,SS=k}$	$\bar{X}_{ME=j',SS=k'}$	Difference
LD - FI	100	0.97	0.71	0.26
LD - MI		0.97	0.82	0.15
LD - FB		0.97	1.28	-0.31
LD - 2M		0.97	0.53	0.44
FI - MI		0.71	0.82	-0.11
FI - FB		0.71	1.28	-0.56
FI - 2M		0.71	0.53	0.19
MI - FB		0.82	1.28	-0.46
MI - 2M		0.82	0.53	0.29
FB - 2M		1.28	0.53	0.75
LD - FI	200	0.76	0.66	0.11
LD - MI		0.76	0.80	-0.04
LD - FB		0.76	1.20	-0.44
LD - 2M		0.76	0.59	0.17
FI - MI		0.66	0.80	-0.14
FI - FB		0.66	1.20	-0.55
FI - 2M		0.66	0.59	0.06
MI - FB		0.80	1.20	-0.40
MI - 2M		0.80	0.59	0.21
FB - 2M		1.20	0.59	0.61
LD - FI	500	0.66	0.61	0.05
LD - MI		0.66	0.81	-0.15
LD - FB		0.66	1.05	-0.39
LD - 2M		0.66	0.62	0.03
FI - MI		0.61	0.81	-0.20
FI - FB		0.61	1.05	-0.44
FI - 2M*		0.61	0.62	-0.01
MI - FB		0.81	1.05	-0.24
MI - 2M		0.81	0.62	0.19
FB - 2M		1.05	0.62	0.43
LD - FI	1000	0.63	0.59	0.04
LD - MI		0.63	0.85	-0.22
LD - FB		0.63	0.97	-0.34
LD - 2M*		0.63	0.66	-0.02
FI - MI		0.59	0.85	-0.26
FI - FB		0.59	0.97	-0.38
FI - 2M		0.59	0.66	-0.07
MI - FB		0.85	0.97	-0.11
MI - 2M		0.85	0.66	0.20
FB - 2M		0.97	0.66	0.31

Note. Means are averaged over levels of SP, MR, and MM. All pairwise comparisons were significant at the $\alpha = 0.05$ level unless otherwise noted by a *. All p -values were adjusted using the Tukey method.

Table B.4

Main Effect Pairwise Comparisons Corresponding to $ME \times SS$ Interaction of Slope Variances for SE/SD Ratios

Contrast (ME)	SS	$\bar{X}_{ME=j,SS=k}$	$\bar{X}_{ME=j',SS=k'}$	Difference
LD - FI	100	1.12	1.05	0.07
LD - MI		1.12	1.18	-0.06
LD - FB		1.12	1.23	-0.11
LD - 2M		1.12	1.35	-0.22
FI - MI		1.05	1.18	-0.13
FI - FB		1.05	1.23	-0.19
FI - 2M		1.05	1.35	-0.30
MI - FB		1.18	1.23	-0.05
MI - 2M		1.18	1.35	-0.16
FB - 2M		1.23	1.35	-0.11
LD - FI*	200	1.04	1.00	0.03
LD - MI		1.04	1.12	-0.08
LD - FB		1.04	1.15	-0.11
LD - 2M		1.04	1.60	-0.56
FI - MI		1.00	1.12	-0.11
FI - FB		1.00	1.15	-0.15
FI - 2M		1.00	1.60	-0.59
MI - FB*		1.12	1.15	-0.03
MI - 2M		1.12	1.60	-0.48
FB - 2M		1.15	1.60	-0.45
LD - FI	500	0.92	0.86	0.06
LD - MI		0.92	1.01	-0.10
LD - FB		0.92	1.01	-0.09
LD - 2M		0.92	1.82	-0.91
FI - MI		0.86	1.01	-0.15
FI - FB		0.86	1.01	-0.15
FI - 2M		0.86	1.82	-0.96
MI - FB*		1.01	1.01	0.00
MI - 2M		1.01	1.82	-0.81
FB - 2M		1.01	1.82	-0.81
LD - FI	1000	0.82	0.77	0.05
LD - MI		0.82	0.99	-0.17
LD - FB		0.82	0.96	-0.13
LD - 2M		0.82	2.11	-1.28
FI - MI		0.77	0.99	-0.22
FI - FB		0.77	0.96	-0.19
FI - 2M		0.77	2.11	-1.34
MI - FB*		0.99	0.96	0.03
MI - 2M		0.99	2.11	-1.12
FB - 2M		0.96	2.11	-1.15

Note. Means are averaged over levels of SP, MR, and MM. All pairwise comparisons were significant at the $\alpha = 0.05$ level unless otherwise noted by a *. All p -values were adjusted using the Tukey method.

Table B.5

Main Effect Pairwise Comparisons Corresponding to $ME \times SS$ Interaction of Intercept Slope Covariances for SE/SD Ratios

Contrast (ME)	SS	$\bar{X}_{ME=j,SS=k}$	$\bar{X}_{ME=j',SS=k'}$	Difference
LD - FI	100	1.09	1.02	0.07
LD - MI		1.09	0.93	0.16
LD - FB		1.09	1.19	-0.10
LD - 2M		1.09	1.21	-0.12
FI - MI		1.02	0.93	0.08
FI - FB		1.02	1.19	-0.18
FI - 2M		1.02	1.21	-0.20
MI - FB		0.93	1.19	-0.26
MI - 2M		0.93	1.21	-0.28
FB - 2M		1.19	1.21	-0.02
LD - FI	200	1.07	1.04	0.03
LD - MI		1.07	1.02	0.05
LD - FB		1.07	1.22	-0.15
LD - 2M		1.07	1.58	-0.51
FI - MI		1.04	1.02	0.02
FI - FB		1.04	1.22	-0.19
FI - 2M		1.04	1.58	-0.54
MI - FB		1.02	1.22	-0.21
MI - 2M		1.02	1.58	-0.56
FB - 2M		1.22	1.58	-0.36
LD - FI	500	0.98	0.92	0.06
LD - MI		0.98	0.99	-0.01
LD - FB		0.98	1.06	-0.08
LD - 2M		0.98	1.96	-0.98
FI - MI		0.92	0.99	-0.07
FI - FB		0.92	1.06	-0.14
FI - 2M		0.92	1.96	-1.04
MI - FB		0.99	1.06	-0.07
MI - 2M		0.99	1.96	-0.97
FB - 2M		1.06	1.96	-0.90
LD - FI	1000	0.93	0.87	0.06
LD - MI		0.93	1.02	-0.09
LD - FB		0.93	0.98	-0.06
LD - 2M		0.93	2.17	-1.24
FI - MI		0.87	1.02	-0.15
FI - FB		0.87	0.98	-0.12
FI - 2M		0.87	2.17	-1.30
MI - FB		1.02	0.98	0.03
MI - 2M		1.02	2.17	-1.15
FB - 2M		0.98	2.17	-1.18

Note. Means are averaged over levels of SP, MR, and MM. All pairwise comparisons were significant at the $\alpha = 0.05$ level unless otherwise noted by a *. All p -values were adjusted using the Tukey method.

Table B.6

Main Effect Pairwise Comparisons Corresponding to $ME \times SS$ Interaction of Residual Variances for SE/SD Ratios

Contrast (ME)	SS	$\bar{X}_{ME=j,SS=k}$	$\bar{X}_{ME=j',SS=k'}$	Difference
LD - FI*	100	1.07	1.06	0.01
LD - MI*		1.07	1.08	-0.01
LD - FB*		1.07	0.99	0.08
LD - 2M		1.07	0.72	0.35
FI - MI*		1.06	1.08	-0.02
FI - FB		1.06	0.99	0.07
FI - 2M		1.06	0.72	0.34
MI - FB		1.08	0.99	0.10
MI - 2M		1.08	0.72	0.36
FB - 2M		0.99	0.72	0.27
LD - FI*	200	1.09	1.08	0.01
LD - MI*		1.09	1.09	-0.01
LD - FB*		1.09	1.06	0.02
LD - 2M		1.09	0.93	0.15
FI - MI*		1.08	1.09	-0.02
FI - FB*		1.08	1.06	0.01
FI - 2M		1.08	0.93	0.14
MI - FB*		1.09	1.06	0.03
MI - 2M		1.09	0.93	0.16
FB - 2M		1.06	0.93	0.13
LD - FI*	500	1.03	1.02	0.01
LD - MI*		1.03	1.03	0.00
LD - FB*		1.03	1.01	0.02
LD - 2M		1.03	1.13	-0.10
FI - MI*		1.02	1.03	-0.01
FI - FB*		1.02	1.01	0.02
FI - 2M		1.02	1.13	-0.11
MI - FB*		1.03	1.01	0.03
MI - 2M		1.03	1.13	-0.10
FB - 2M		1.01	1.13	-0.12
LD - FI*	1000	1.01	1.01	0.00
LD - MI*		1.01	1.01	0.00
LD - FB*		1.01	0.99	0.02
LD - 2M		1.01	1.34	-0.33
FI - MI*		1.01	1.01	0.00
FI - FB		1.01	0.99	0.02
FI - 2M		1.01	1.34	-0.33
MI - FB*		1.01	0.99	0.02
MI - 2M		1.01	1.34	-0.33
FB - 2M		0.99	1.34	-0.35

Note. Means are averaged over levels of SP, MR, and MM. All pairwise comparisons were significant at the $\alpha = 0.05$ level unless otherwise noted by a *. All p -values were adjusted using the Tukey method.

Table B.7

Main Effect Pairwise Comparisons Corresponding to ME × MR Interaction of Intercept Variances for SE/SD Ratios

Contrast (ME)	MR	$\bar{X}_{ME=j,MR=k}$	$\bar{X}_{ME=j',MR=k'}$	Difference
LD - FI*	5	0.63	0.63	0.00
LD - MI*		0.63	0.67	-0.04
LD - FB		0.63	1.11	-0.48
LD - 2M*		0.63	0.63	0.00
FI - MI*		0.63	0.67	-0.04
FI - FB		0.63	1.11	-0.48
FI - 2M*		0.63	0.63	0.00
MI - FB		0.67	1.11	-0.45
MI - 2M*		0.67	0.63	0.04
FB - 2M		1.11	0.63	0.48
LD - FI*	10	0.65	0.63	0.03
LD - MI		0.65	0.71	-0.06
LD - FB		0.65	1.09	-0.44
LD - 2M		0.65	0.61	0.05
FI - MI		0.63	0.71	-0.09
FI - FB		0.63	1.09	-0.47
FI - 2M*		0.63	0.61	0.02
MI - FB		0.71	1.09	-0.38
MI - 2M		0.71	0.61	0.10
FB - 2M		1.09	0.61	0.48
LD - FI	20	0.71	0.63	0.08
LD - MI		0.71	0.81	-0.10
LD - FB		0.71	1.15	-0.44
LD - 2M		0.71	0.60	0.11
FI - MI		0.63	0.81	-0.18
FI - FB		0.63	1.15	-0.52
FI - 2M*		0.63	0.60	0.03
MI - FB		0.81	1.15	-0.34
MI - 2M		0.81	0.60	0.21
FB - 2M		1.15	0.60	0.55
LD - FI	30	0.81	0.65	0.15
LD - MI		0.81	0.90	-0.09
LD - FB		0.81	1.13	-0.33
LD - 2M		0.81	0.59	0.22
FI - MI		0.65	0.90	-0.25
FI - FB		0.65	1.13	-0.48
FI - 2M		0.65	0.59	0.06
MI - FB		0.90	1.13	-0.23
MI - 2M		0.90	0.59	0.31
FB - 2M		1.13	0.59	0.54
LD - FI	40	0.98	0.67	0.30
LD - MI		0.98	1.02	-0.04
LD - FB		0.98	1.14	-0.16
LD - 2M		0.98	0.57	0.40
FI - MI		0.67	1.02	-0.34
FI - FB		0.67	1.14	-0.46
FI - 2M		0.67	0.57	0.10
MI - FB		1.02	1.14	-0.12
MI - 2M		1.02	0.57	0.45
FB - 2M		1.14	0.57	0.57

Note. Means are averaged over levels of SP, SS, and MM. All pairwise comparisons were significant at the $\alpha = 0.05$ level unless otherwise noted by a *. All p -values were adjusted using the Tukey method.

Table B.8

Main Effect Pairwise Comparisons Corresponding to ME \times MR Interaction of Slope Variances for SE/SD Ratios

Contrast (ME)	MR	$\bar{X}_{ME=j,MR=k}$	$\bar{X}_{ME=j',MR=k'}$	Difference
LD - FI*	5	0.87	0.88	-0.01
LD - MI*		0.87	0.89	-0.02
LD - FB		0.87	1.07	-0.20
LD - 2M		0.87	1.91	-1.04
FI - MI*		0.88	0.89	-0.02
FI - FB		0.88	1.07	-0.20
FI - 2M		0.88	1.91	-1.03
MI - FB		0.89	1.07	-0.18
MI - 2M		0.89	1.91	-1.01
FB - 2M		1.07	1.91	-0.83
LD - FI*	10	0.88	0.88	0.00
LD - MI		0.88	0.96	-0.08
LD - FB		0.88	1.09	-0.21
LD - 2M		0.88	1.84	-0.96
FI - MI		0.88	0.96	-0.08
FI - FB		0.88	1.09	-0.21
FI - 2M		0.88	1.84	-0.97
MI - FB		0.96	1.09	-0.13
MI - 2M		0.96	1.84	-0.89
FB - 2M		1.09	1.84	-0.76
LD - FI*	20	0.95	0.91	0.04
LD - MI		0.95	1.09	-0.14
LD - FB		0.95	1.08	-0.13
LD - 2M		0.95	1.74	-0.79
FI - MI		0.91	1.09	-0.18
FI - FB		0.91	1.08	-0.16
FI - 2M		0.91	1.74	-0.83
MI - FB		1.09	1.08	0.02
MI - 2M		1.09	1.74	-0.65
FB - 2M		1.08	1.74	-0.67
LD - FI	30	1.04	0.95	0.10
LD - MI		1.04	1.17	-0.13
LD - FB*		1.04	1.08	-0.04
LD - 2M		1.04	1.61	-0.57
FI - MI		0.95	1.17	-0.23
FI - FB		0.95	1.08	-0.14
FI - 2M		0.95	1.61	-0.66
MI - FB		1.17	1.08	0.09
MI - 2M		1.17	1.61	-0.43
FB - 2M		1.08	1.61	-0.53
LD - FI	40	1.14	1.00	0.14
LD - MI		1.14	1.26	-0.12
LD - FB*		1.14	1.12	0.02
LD - 2M		1.14	1.49	-0.35
FI - MI		1.00	1.26	-0.26
FI - FB		1.00	1.12	-0.13
FI - 2M		1.00	1.49	-0.50
MI - FB		1.26	1.12	0.14
MI - 2M		1.26	1.49	-0.23
FB - 2M		1.12	1.49	-0.37

Note. Means are averaged over levels of SP, SS, and MM. All pairwise comparisons were significant at the $\alpha = 0.05$ level unless otherwise noted by a *. All p -values were adjusted using the Tukey method.

Table B.9

Main Effect Pairwise Comparisons Corresponding to ME × MR Interaction of Residual Variances for SE/SD Ratios

Contrast (ME)	MR	$\bar{X}_{ME=j,MR=k}$	$\bar{X}_{ME=j',MR=k'}$	Difference
LD - FI*	5	1.04	1.03	0.01
LD - MI*		1.04	1.04	0.00
LD - FB*		1.04	1.01	0.03
LD - 2M		1.04	1.19	-0.15
FI - MI*		1.03	1.04	-0.01
FI - FB*		1.03	1.01	0.02
FI - 2M		1.03	1.19	-0.16
MI - FB*		1.04	1.01	0.03
MI - 2M		1.04	1.19	-0.16
FB - 2M		1.01	1.19	-0.18
LD - FI*	10	1.03	1.04	-0.01
LD - MI*		1.03	1.04	-0.01
LD - FB*		1.03	1.01	0.02
LD - 2M		1.03	1.14	-0.11
FI - MI*		1.04	1.04	0.00
FI - FB*		1.04	1.01	0.03
FI - 2M		1.04	1.14	-0.10
MI - FB*		1.04	1.01	0.03
MI - 2M		1.04	1.14	-0.10
FB - 2M		1.01	1.14	-0.13
LD - FI*	20	1.06	1.05	0.01
LD - MI*		1.06	1.05	0.01
LD - FB		1.06	1.01	0.05
LD - 2M*		1.06	1.04	0.02
FI - MI*		1.05	1.05	0.00
FI - FB		1.05	1.01	0.03
FI - 2M*		1.05	1.04	0.00
MI - FB		1.05	1.01	0.04
MI - 2M*		1.05	1.04	0.01
FB - 2M*		1.01	1.04	-0.03
LD - FI*	30	1.06	1.04	0.01
LD - MI*		1.06	1.07	-0.01
LD - FB		1.06	1.01	0.05
LD - 2M		1.06	0.94	0.12
FI - MI*		1.04	1.07	-0.03
FI - FB		1.04	1.01	0.03
FI - 2M		1.04	0.94	0.10
MI - FB		1.07	1.01	0.06
MI - 2M		1.07	0.94	0.13
FB - 2M		1.01	0.94	0.07
LD - FI*	40	1.06	1.05	0.01
LD - MI*		1.06	1.07	-0.02
LD - FB		1.06	1.01	0.04
LD - 2M		1.06	0.83	0.23
FI - MI*		1.05	1.07	-0.03
FI - FB		1.05	1.01	0.04
FI - 2M		1.05	0.83	0.22
MI - FB		1.07	1.01	0.06
MI - 2M		1.07	0.83	0.25
FB - 2M		1.01	0.83	0.18

Note. Means are averaged over levels of SP, SS, and MM. All pairwise comparisons were significant at the $\alpha = 0.05$ level unless otherwise noted by a *. All p -values were adjusted using the Tukey method.

Appendix C: Code to Implement the Two-Stage Imputation Method in Mplus

The following steps are provided as guidance to implement the two-stage imputation method for GMMs using Mplus. Note that this procedure is illustrated using a two class model GMM, but the procedure can be generalized to more than two classes as well as other types of mixture models. Assuming M first stage imputations (classes) and N second stage imputations (manifest variables), the steps are:

1. Run the first stage imputation code:

```
TITLE: GMM Stage 1 MI for 2-stg MI (2M)
DATA:
FILE=data.dat;
VARIABLE:
  NAMES = id y1-y4;
  USEV = y1-y4;
  MISSING = ALL (-99);
  CLASSES = c(2);
  IDVARIABLE = id;
ANALYSIS:
  TYPE = MIXTURE;
  ESTIMATOR = BAYES;
  PROCESSORS = 1;
  CHAINS = 1;
  STVALUES = ML;
MODEL:
%overall%
  [y1-y4@0];
  i s | y1@0 y2@1 y3@2 y4@3;
  y1-y4;
  i(vi);
  s(vs);
  i WITH s;
%c#1%
  [i](i1);
  [s](s1);
```

```

%c#2%
  [i](i2);
  [s](s2);
MODEL CONSTRAINT:
  i1 > i2;
  s2 > s1;
SAVEDATA:
FILE = stg1probs.dat;
SAVE = fscores(30 1000);
OUTPUT:
TECH8;

```

In this code, $M = 30$ and can be modified by changing the first value of the line

```
SAVE = fscores(30 1000);
```

2. Extract the latent class plausible values from the *stg1probs.dat* file. This file should contain M sets of imputed class variables.
3. Take one set of plausible class values (m) out of the M sets of imputed class values and create a new data file that includes the original data with missing values and a single column for the plausible class values.
4. Separate the classes to different data files so that the next set of imputations are conditional on the same class. In other words, for each dataset m , create separate datasets for each class. For each group dataset, use the following Mplus code to multiply impute N sets of the missing values:

```

TITLE: Stage 2 GROUP 1 or 2 2-stg MI (2M)
DATA:
  FILE = stg2_group1_data.dat;
VARIABLE:
  NAMES = id y1-y4 g;
  USEV = y1-y4;
  MISSING = ALL (-99);
  IDVARIABLE = id;
  AUXILIARY = g;
DATA IMPUTATION:
  IMPUTE = y1-y4;
  Ndatasets = 10;
  SAVE = stg2_g1_imp*.dat;
  THIN = 1000;
ANALYSIS:
TYPE = BASIC;

```

This should create $N \times 2$ fully imputed datasets. In this code, $N = 10$ and can be modified by changing the value of

```
Ndatasets = 10;
```

Combine the corresponding datasets back together (group 1 imputed data and group 2 imputed data), resulting in a total of N fully imputed datasets for dataset m .

5. Run a multiple-group (two class model with known classes) growth mixed effects model for each dataset in N using the following code:

```
TITLE:
  Stage 2 combined MI with complete data (2M)
DATA:
  FILE = stg2_g1g2Combined_data.dat;
VARIABLE:
  NAMES = y1-y4 g id;
  USEV = y1-y4 g;
  CLASSES = c(2);
  KNOWNCLASS = c(g = 1 g = 2);
  IDVARIABLE = id;
ANALYSIS:
  TYPE = MIXTURE;
  ESTIMATOR = BAYES;
  PROCESSORS = 1;
  CHAINS = 1;
MODEL:
%overall%
  [y1-y4@0];
  i s | y1@0 y2@1 y3@2 y4@3;
  y1-y4(yerr);
  i;
  s;
  i WITH s(v);
%c#1%
  [i](i1);
  [s](s1);
%c#2%
  [i](i2);
  [s](s2);
MODEL CONSTRAINT:
  i1 > i2;
  s2 > s1;
```

6. Repeat steps 3 to 5 for each imputed class variable m and for each imputed dataset n .

7. Combine parameter estimates by averaging across all M by N results.
8. Combine standard errors by following rules outlined in Section [2.4.4](#).

Bibliography

- Anderson, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association*, *52*(278), 200–203.
- Anderson, T. W., & Bahadur, R. R. (1962). Classification into two multivariate normal distributions with different covariance matrices. *The Annals of Mathematical Statistics*, *33*(2), 420–431. doi: 10.1214/aoms/1177704568
- Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 243–278). Hillsdale, NJ: Lawrence Erlbaum.
- Asparouhov, T., & Muthén, B. O. (2010). Bayesian analysis using Mplus: Technical implementation. *Manuscript submitted for publication*.
- Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, *42*(4), 757–786.
- Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*(3), 338–363.
- Beale, E. M. L., & Little, R. J. A. (1975). Missing values in multivariate analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, *37*(1), 129–145.
- Breiman, L. (2017). *Classification and regression trees*. New York: Routledge.

- Brown, R. L. (1994). Efficacy of the indirect approach for estimating structural equation models with missing data: A comparison of five methods. *Structural Equation Modeling: A Multidisciplinary Journal*, 1(4), 287–316. doi: 10.1080/10705519409539983
- Browne, W. J., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, 1(3), 473–514. doi: 10.1214/06-BA117
- Buck, S. F. (1960). A method of estimation of missing values in multivariate data suitable for use with an electronic computer. *Journal of the Royal Statistical Society, Series B*(22), 302–307.
- Cai, J.-H., Song, X.-Y., & Hser, Y.-I. (2010). A Bayesian analysis of mixture structural equation models with non-ignorable missing responses and covariates. *Statistics in Medicine*, 29(18), 1861–1874.
- Celeux, G., Hurn, M., & Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, 95(451), 957–970.
- Chen, F. (2011). The RANDOM statement and more: Moving on with PROC MCMC. In *SAS Global Forum 2011 Conference*. Cary, NC: SAS Institute, Inc.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Colder, C. R., Campbell, R. T., Ruel, E., Richardson, J. L., & Flay, B. R. (2002). A finite mixture model of growth trajectories of adolescent alcohol use: Predictors

- and consequences. *Journal of Consulting and Clinical Psychology*, 70(4), 976–985.
- 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), 330–351.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 16–29.
- de Leeuw, E. D., Hox, J. J., & Huisman, M. (2003). Prevention and treatment of item nonresponse. *Journal of Official Statistics*, 19(2), 153–176.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1–38.
- Depaoli, S. (2013). Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus Bayesian estimation. *Psychological Methods*, 18(2), 186–219.
- Diebolt, J., & Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(2), 363–375.
- Diggle, P., & Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43(1), 49–93.

- Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: The Guilford Press.
- Enders, C. K., & Gottschall, A. C. (2011). Multiple imputation strategies for multiple group structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(1), 35–54.
- Enders, C. K., & Tofighi, D. (2008). The impact of misspecifying class-specific residual variances in growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*(1), 75–95.
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466–491. doi: 10.1037/1082-989X.9.4.466
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721–741. doi: 10.1109/TPAMI.1984.4767596
- Glynn, R. J., Laird, N. M., & Rubin, D. B. (1986). Selection modeling versus mixture modeling with nonignorable nonresponse. In *Drawing inferences from self-selected samples* (pp. 115–142). New York, NY: Springer.
- Gold, M. S., & Bentler, P. M. (2000). Treatments of missing data: A Monte Carlo

- comparison of RBHDI, iterative stochastic regression imputation, and expectation-maximization. *Structural Equation Modeling: A Multidisciplinary Journal*, 7(3), 319–355.
- Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 10(1), 80–100.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60(1), 549–576.
- Graham, J. W., Taylor, B. J., & Cumsille, P. E. (2001). Planned missing-data designs in analysis of change. In *New methods for the analysis of change* (pp. 335–353). Washington, DC, US: American Psychological Association. doi: 10.1037/10409-011
- Harel, O. (2003). *Strategies for data analysis with two types of missing values* (Unpublished doctoral dissertation). Pennsylvania State University, State College, PA.
- Harel, O. (2007). Inferences on missing information under multiple imputation and two-stage multiple imputation. *Statistical Methodology*, 4(1), 75–89.
- Harel, O. (2009). The estimation of R² and adjusted R² in incomplete data sets using multiple imputation. *Journal of Applied Statistics*, 36(10), 1109–1118.
- Harel, O., Chung, H., & Miglioretti, D. (2013). Latent class regression: Inference and estimation with two-stage multiple imputation: Two-stage MI for LCR. *Biometrical Journal*, 55(4), 541–553.

- Harel, O., & Schafer, J. L. (2002). Multiple imputation in two stages. In *ASA Proceedings of the Joint Statistical Meetings*. New York, NY.
- Harel, O., & Schafer, J. L. (2009). Partial and latent ignorability in missing-data problems. *Biometrika*, *96*(1), 37–50.
- Harring, J. R. (2012). Finite mixtures of nonlinear mixed effects models. In J. R. Harring & G. R. Hancock (Eds.), *Advances in longitudinal methods in the social and behavioral sciences*. Charlotte, NC: Information Age Publishing, Inc.
- Hedeker, D., & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, *2*(1), 64–78. doi: 10.1037/1082-989X.2.1.64
- Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(2), 202–226.
- Hesser, H., Hedman, E., Lindfors, P., Andersson, E., & Ljótsson, B. (2017). The specific effect of systematic exposure in irritable bowel syndrome: complier average causal effect analysis using growth mixture modeling. *Psychological Medicine*, *47*(15), 2653–2662.
- Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods*, *11*(1), 36–53.
- Ho, T. K. (1995). Random decision forests. In *Proceedings of the third international conference on document analysis and recognition* (Vol. Volume 1, pp. 278–282).

Washington, DC, USA: IEEE Computer Society.

Horton, N. J., & Lipsitz, S. R. (2001). Multiple Imputation in Practice. *The American Statistician*, *55*(3), 244–254. doi: 10.1198/000313001317098266

Hughes, R. A., White, I. R., Seaman, S. R., Carpenter, J. R., Tilling, K., & Sterne, J. A. C. (2014). Joint modelling rationale for chained equations. *BMC Medical Research Methodology*, *14*, 28. doi: 10.1186/1471-2288-14-28

Huque, M. H., Carlin, J. B., Simpson, J. A., & Lee, K. J. (2018). A comparison of multiple imputation methods for missing data in longitudinal studies. *BMC Medical Research Methodology*, *18*(1), 168. doi: 10.1186/s12874-018-0615-6

Kalton, G. (1983). *Introduction to survey sampling*. Beverly Hills, CA: Sage Publications.

Kaplan, D. (1989). A study of the sampling variability and z-values of parameter estimates from misspecified structural equation models. *Multivariate Behavioral Research*, *24*(1), 41–57. doi: 10.1207/s15327906mbr2401_3

Kaplan, D. (2002). Methodological advances in the analysis of individual growth with relevance to education policy. *Peabody Journal of Education*, *77*(4), 189–215.

Kish, L. (1965). Sampling organizations and groups of unequal sizes. *American sociological review*, *30*(4), 564–572.

Kohli, N. (2011). *Estimating unknown knots in piecewise linear-linear latent growth mixture models* (Unpublished doctoral dissertation). University of Maryland, College Park, MD.

Koning, M., Hoekstra, T., de Jong, E., Visscher, T. L. S., Seidell, J. C., & Renders,

- C. M. (2016). Identifying developmental trajectories of body mass index in childhood using latent class growth (mixture) modelling: Associations with dietary, sedentary and physical activity behaviors: a longitudinal study. *BMC Public Health*, *16*(1), 1128.
- Lee, D. Y., Harring, J. R., & Stapleton, L. M. (in press). Accounting for respondent attrition in the longitudinal modeling of panel data. *Journal of Experimental Education*.
- Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. Chichester, England: Wiley.
- Lee, S.-Y., & Song, X.-Y. (2003). Model comparison of nonlinear structural equation models with fixed covariates. *Psychometrika*, *68*(1), 27–47. doi: 10.1007/BF02296651
- Lee, S.-Y., & Tang, N.-S. (2006). Analysis of nonlinear structural equation models with nonignorable missing covariates and ordered categorical data. *Statistica Sinica*, *16*(4), 1117–1141.
- Li, M., Chen, N., Cui, Y., & Liu, H. (2017). Comparison of different LGM-based methods with MAR and MNAR dropout data. *Frontiers in Psychology*, *8*. doi: 10.3389/fpsyg.2017.00722
- Li, M., Harring, J. R., & Macready, G. B. (2014). Investigating the feasibility of using Mplus in the estimation of growth mixture models. *Journal of Modern Applied Statistical Methods*, *13*(1), 31.
- Lin, M., Narayan, V., Drevets, W. C., Ye, J., & Li, Q. (2017). Application of growth

- mixture modeling in antidepressant treatment response studies. *Biological Psychiatry*, 81(10), 224.
- Little, R. J. A. (1992). Regression with missing X's: A review. *Journal of the American Statistical Association*, 87(420), 1227–1237.
- Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88(421), 125–134.
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. John Wiley & Sons.
- Liu, M., Hancock, G. R., & Harring, J. R. (2011). Using finite mixture modeling to deal with systematic measurement error: A case study. *Journal of Modern Applied Statistical Methods*, 10(1), 249–261.
- Lord, F. M. (1955). Estimation of parameters from incomplete data. *Journal of the American Statistical Association*, 50(271), 870–876.
- Lu, Z. L., Zhang, Z., & Lubke, G. (2011). Bayesian inference for growth mixture models with latent class dependent missing data. *Multivariate Behavioral Research*, 46(4), 567–597.
- Lubke, G., & Muthén, B. O. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(1), 26–47.
- Lubke, G., & Neale, M. C. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research*, 41(4), 499–532.

- Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2013). *The BUGS book*. Boca Raton, FL: CRC Press.
- Maddala, G. S. (1983). *Limited dependent and qualitative variables in econometrics*. New York: Cambridge University Press.
- Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science of India*, 12, 49–55.
- McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics*, 16(1), 163–200.
- McGinniss, J., & Harel, O. (2016). Multiple imputation in three or more stages. *Journal of Statistical Planning and Inference*, 176, 33–51. doi: 10.1016/j.jspi.2016.04.001
- McLachlan, G., & Krishnan, T. (2007). *The EM algorithm and extensions* (Vol. 382). John Wiley & Sons.
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. Hoboken, NJ: Wiley.
- McNeish, D. (2016). Using data-dependent priors to mitigate small sample bias in latent growth models: A discussion and illustration using Mplus. *Journal of Educational and Behavioral Statistics*, 41(1), 27–56. doi: 10.3102/1076998615621299
- McNeish, D. (2017). Missing data methods for arbitrary missingness with small samples. *Journal of Applied Statistics*, 44(1), 24–39. doi: 10.1080/02664763.2016.1158246
- McNeish, D., & Harring, J. R. (2017). The effect of model misspecification on growth

- mixture model class enumeration. *Journal of Classification*, *34*(2), 223–248.
- McNeish, D., & Matta, T. (2018). Differentiating between mixed-effects and latent-curve approaches to growth modeling. *Behavior Research Methods*, *50*(4), 1398–1414. doi: 10.3758/s13428-017-0976-5
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*(1), 107–122.
- Mistler, S. A., & Enders, C. K. (2017). A comparison of joint model and fully conditional specification imputation for multilevel missing data. *Journal of Educational and Behavioral Statistics*, *42*(4), 432–466. doi: 10.3102/1076998617690869
- Morris, T. P., White, I. R., & Royston, P. (2014). Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Medical Research Methodology*, *14*(1), 75. doi: 10.1186/1471-2288-14-75
- Musu-Gillette, L. E., Wigfield, A., Harring, J. R., & Eccles, J. S. (2015). Trajectories of change in students' self-concepts of ability and values in math and college major choice. *Educational Research and Evaluation*, *21*(4), 343–370.
- Muthén, B. O. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1–33). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Muthén, B. O. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. *The SAGE Handbook of Quantitative Methodology for the Social Sciences*, *345*, 368.

- Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, *2*(4), 371–402.
- Muthén, B. O., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *52*, 431–462.
- Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, *24*(6), 882–891.
- Muthén, B. O., Muthén, L. K., & Asparouhov, T. (2010). Bayesian analysis using Mplus.
- Muthén, B. O., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*(2), 463–469.
- Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide* (Eighth ed.). Los Angeles, CA: Muthén & Muthén.
- Nagin, D. S. (1999). Analyzing developmental trajectories: a semiparametric, group-based approach. *Psychological Methods*, *4*(2), 139.
- Ning, L., & Luo, W. (2017). Class identification efficacy in piecewise GMM with unknown turning points. *The Journal of Experimental Education*, *86*(2), 282–307. doi: 10.1080/00220973.2017.1301354
- Nylund, K. L., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*,

14(4), 535–569.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74*(4), 525–556.

Peugh, J. L., & Fan, X. (2012). How well does growth mixture modeling identify heterogeneous growth trajectories? A simulation study examining GMM's performance characteristics. *Structural Equation Modeling: A Multidisciplinary Journal, 19*(2), 204–226.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581.

Rubin, D. B. (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business and Economic Statistics, 4*(1), 87–94. doi: 10.2307/1391390

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley.

Rubin, D. B. (2003). Nested multiple imputation of NMES via partially incompatible MCMC. *Statistica Neerlandica, 57*(1), 3–18.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147–177.

Schafer, J. L., & Yucel, R. M. (2002). Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics, 11*(2), 437–457. doi: 10.1198/106186002760180608

Scheuren, F. (2005). Multiple Imputation: How it began and continues. *The American*

- Statistician*, 59(4), 315–319.
- Shen, Z. (2000). *Nested multiple imputation* (Unpublished doctoral dissertation).
Harvard University, Department of Statistics, Cambridge, MA.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.
- Song, X.-Y., & Lee, S.-Y. (2002). Analysis of structural equation model with ignorable missing continuous and polytomous data. *Psychometrika*, 67(2), 261–288. doi: 10.1007/BF02294846
- Sterba, S. K. (2014). Handling missing covariates in conditional mixture models under missing at random assumptions. *Multivariate Behavioral Research*, 49(6), 614–632.
- Sterba, S. K. (2016). Cautions on the use of multiple imputation when selecting between latent categorical versus continuous models for psychological constructs. *Journal of Clinical Child and Adolescent Psychology*, 45(2), 167–175.
- Team, R. D. C. (2008). *R: A language and environment for statistical computing* (Tech. Rep.). Vienna, Austria: R Foundation for Statistical Computing.
Retrieved from <http://www.R-project.org>
- Titterton, D. M., Smith, A. F. M., & Makovm, U. E. (1985). *Statistical analysis of finite mixture models*. New York: Wiley.
- Tofghi, D., & Enders, C. K. (2007). Identifying the correct number of classes in growth mixture models. In G. R. Hancock (Ed.), *Advances in Latent Variable Mixture Models* (pp. 317–341). Greenwich: Information Age Publishing, Inc.

- Tueller, S. J., Drotar, S., & Lubke, G. H. (2011). Addressing the problem of switched class labels in latent variable mixture model simulation studies. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(1), 110–131. doi: 10.1080/10705511.2011.534695
- Tueller, S. J., & Lubke, G. H. (2010). Evaluation of structural equation mixture models: Parameter estimates and correct class assignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 17(2), 165–192. doi: 10.1080/10705511003659318
- van Buuren, S. (2011). Multiple imputation of multilevel data. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of Advanced Multilevel Analysis* (pp. 173–196). New York, NY: Routledge.
- van Buuren, S., Brand, J. P., Groothuis-Oudshoorn, C. G., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, 76(12), 1049–1064. doi: 10.1080/10629360600810434
- Vickers, A. J. (2003). How many repeated measures in repeated measures designs? Statistical issues for comparative trials. *BMC Medical Research Methodology*, 3, 22. doi: 10.1186/1471-2288-3-22
- Wilks, S. S. (1932). Moments and distributions of estimates of population parameters from fragmentary samples. *Annals of Mathematical Statistics*, 3(2), 163–195.
- Willett, J. B., Singer, J. D., & Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical

- models and methodological recommendations. *Development and Psychopathology*, *10*(2), 395–426.
- Winship, C., Mare, R. D., & Warren, J. R. (2002). Latent class models for contingency tables with missing data. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied Latent Class Analysis*. Cambridge University Press.
- Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 256–293). Newbury Park, CA: Sage.
- Wothke, W. (2000). Longitudinal and multi-group modeling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multiple group data: Practical issues, applied approaches and specific examples* (pp. 219–240). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Wu, A. D., Zumbo, B. D., & Siegel, L. S. (2011). General piecewise growth mixture model: Word recognition development for different learners in different phases. *Journal of Modern Applied Statistical Methods*, *10*(1), 226–248. doi: 10.22237/jmasm/1304223600