

UC Merced

UC Merced Electronic Theses and Dissertations

Title

Advanced Methods for Implementation of Bayesian Growth Mixture Modeling

Permalink

<https://escholarship.org/uc/item/4kk918c9>

ISBN

9798197864376

Author

Heo, Ihnwhi

Publication Date

2026-05-08

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, MERCED

Advanced Methods for Implementation of Bayesian Growth Mixture Modeling

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Psychological Sciences

with a specialization in

Quantitative Methods, Measurement, and Statistics

by

Ihnwhi Heo

2026

Dissertation Committee:

Professor Sarah Depaoli, Co-Chair
Professor Fan Jia, Co-Chair
Professor Haiyan Liu
Professor Meng Qiu

© Ihnwhi Heo, 2026
All rights reserved.

SIGNATURE PAGE

The dissertation of Ihnwhi Heo is approved, and
it is acceptable in quality and form for publication
on microfilm and electronically:

(Professor Meng Qiu)

(Professor Haiyan Liu)

(Professor Fan Jia, Co-Chair)

(Professor Sarah Depaoli, Co-Chair)

University of California, Merced

2026

DEDICATION

To my father, mother, and sister.

For the pride that has held me from afar,
the prayers that have walked with me each day,
and the steady love that has looked after me
since the very beginning,
the quiet weight behind every page is yours.

EPIGRAPH

As I see the Bayesian, or as I would prefer to call it, the coherent, argument today, it consists of two complementary parts. First, a development of the position from axioms, along the same principles as other parts of mathematics. [...] Secondly, an application of these developments to scientific problems that really works. Theoretical satisfaction and practical implementation are the twin ideals of coherent statistics. In retrospect one can see that both these aspects are present in Jeffreys.

Dennis V. Lindley, 1980

TABLE OF CONTENTS

	Signature Page	iii
	Dedication	iv
	Epigraph	v
	Table of Contents	vi
	List of Figures	ix
	List of Tables	xi
	Acknowledgements	xii
	Vita	xiv
	Abstract	xvi
Chapter 1	Overview of Dissertation	1
	1.1 Preview of Studies	1
	1.2 Organization	2
Chapter 2	Bayesian Growth Mixture Modeling: A Methodological Framework	3
	2.1 Modeling Intraindividual Change: Latent Growth Curve Models	3
	2.2 Accommodating Heterogeneity: Latent Growth Mixture Models	5
	2.3 Updating Beliefs with Data: Bayesian Inference	6
	2.3.1 Overview of Bayesian Estimation	6
	2.3.2 Markov Chain Monte Carlo	7
	2.4 Integrating Model and Inference: Bayesian Growth Mixture Modeling	8
	2.4.1 Advantages and Considerations	8
	2.4.2 Prior Specifications	9
	2.5 Implementation Gaps in Bayesian Growth Mixture Modeling	10
Chapter 3	Bayesian Variable Selection via Shrinkage Priors in Growth Mixture Models	11
	3.1 Introduction	11
	3.1.1 Modeling Change and Uncovering Heterogeneity	11
	3.1.2 Describing and Predicting Heterogeneous Growth	12
	3.1.3 Problem of Predictor Selection and Bayesian Solutions	13
	3.1.4 Goals and Structure of the Present Investigation	15

3.2	Incorporating Predictors into Growth Mixture Models . . .	15
3.3	Bayesian Variable Selection via Shrinkage Priors	17
3.4	Shrinkage Prior Specifications Under Investigation	19
3.4.1	Ridge	19
3.4.2	Lasso	21
3.4.3	Hyperlasso	22
3.4.4	Elastic Net	22
3.4.5	Horseshoe	23
3.4.6	Regularized Horseshoe	23
3.4.7	Spike-and-Slab	24
3.5	Performance of Shrinkage Priors in Bayesian Latent Variable Modeling	25
3.6	Simulation Design	27
3.6.1	Population Model	27
3.6.2	Sample Size	29
3.6.3	Class Proportion	29
3.6.4	Class Separation	29
3.6.5	Predictor Effect Pattern	30
3.6.6	Predictor Intercorrelation	31
3.6.7	Shrinkage Prior Specification	31
3.6.8	Analytic Procedure	33
3.6.9	Outcome Measures	33
3.7	Simulation Results	36
3.7.1	Convergence Rate	36
3.7.2	Estimation Accuracy	37
3.7.3	Variable Selection Accuracy	43
3.7.4	Classification Accuracy	52
3.7.5	Predictive Performance	53
3.8	Discussion	53
3.8.1	Shrinkage Behavior and the Role of Effect Magnitude	55
3.8.2	Variable Selection: Power, False Positives, and the Heterogeneous Pattern	57
3.8.3	Convergence, Classification, and Predictive Performance	58
3.8.4	Advice for Applied Researchers	59
3.8.5	Future Research Avenues	63
3.8.6	Concluding Remarks	64
Chapter 4	Bayesian Nonparametric Multiple Imputation via the Chinese Restaurant Process for Growth Mixture Models	66
4.1	Introduction	66
4.1.1	Heterogeneity in Longitudinal Change and Growth Mixture Models	66

	4.1.2	Missing Data in Growth Mixture Models	67
	4.1.3	Bayesian Nonparametric Multiple Imputation and the Present Study	69
	4.2	Missing Data Issues in Growth Mixture Models	70
	4.2.1	Missing Data Mechanisms	71
	4.2.2	Review of Existing Approaches	72
	4.2.3	Summary of the Current Landscape	76
	4.3	Bayesian Nonparametric Multiple Imputation	77
	4.3.1	From the Dirichlet Distribution to the Dirichlet Process	77
	4.3.2	Dirichlet Process Mixture Models	78
	4.3.3	The Chinese Restaurant Process	78
	4.3.4	Multiple Imputation via the Chinese Restaurant Pro- cess	79
	4.4	Simulation Design	82
	4.4.1	Population Model	83
	4.4.2	Sample Size	83
	4.4.3	Class Proportion	84
	4.4.4	Class Separation	84
	4.4.5	Missing Data Proportion	84
	4.4.6	Missing Data Methods	85
	4.4.7	Analytic Settings	88
	4.4.8	Outcome Measures	89
	4.5	Simulation Results	90
	4.5.1	Convergence and Degenerate Solutions	91
	4.5.2	Parameter Recovery	93
	4.5.3	Classification Accuracy	105
	4.6	Discussion	110
	4.6.1	Convergence Difficulties and Degenerate Solutions .	111
	4.6.2	Recovery of Growth Factor Parameters	112
	4.6.3	Classifying Individuals While Handling Missingness	114
	4.6.4	Implications for Applied Practice	115
	4.6.5	Open Questions and Future Directions	116
	4.6.6	Concluding Remarks	117
Chapter 5		General Discussion	119
	5.1	Contributions	119
	5.2	Limitations and Prospects	121
	5.3	Outlook	122

Online Supplemental Material. All supplementary results are available on the Dissertation Open Science Framework (OSF) page: <https://osf.io/8pbd5/>.

LIST OF FIGURES

Figure 3.1:	Conceptual diagram of three pathways through which predictors can influence outcomes in a growth mixture modeling framework: latent class membership through multinomial logistic regression, within-class growth factors through linear regression, and repeated measures directly as a time-varying covariate effect.	16
Figure 3.2:	Prior densities of the seven focal shrinkage priors, plotted under illustrative hyperparameter settings. The top row shows priors with frequentist origins (ridge, lasso, hyperlasso, and elastic net), and the bottom row shows priors developed within the Bayesian framework (horseshoe, regularized horseshoe, and spike-and-slab). Ridge: $\sigma^2/\lambda = 1$. Lasso: scale = 1. Hyperlasso: $\nu = 1/2$, $\lambda \sim \text{half-Cauchy}(0, 1)$. Elastic net: $\lambda_1 = 1/4$, $\lambda_2 = 1$. Horseshoe: $\tau_j, \lambda \sim \text{half-Cauchy}(0, 1)$. Regularized horseshoe: $\tau_j, \lambda \sim \text{half-Cauchy}(0, 1)$, $c = 2$. Spike-and-Slab: $\omega = 0.5$, $\phi_0^2 = 0.01$, $\tau_j^2 = 1$	20
Figure 3.3:	Path diagram of the two-class linear conditional growth mixture model used as the population model. The dashed rectangle indicates the ten candidate predictors, with a generic predictor x_j entering through regression coefficients γ_j (latent class membership) and $\beta_{\ell j}^{(c)}$ (within-class growth factors).	28
Figure 3.4:	Relative Bias for Class Membership Coefficients.	38
Figure 3.5:	Relative Bias for Intercept Growth Factor Regression Coefficients Under the Homogeneous Pattern.	40
Figure 3.6:	Relative Bias for Slope Growth Factor Regression Coefficients Under the Homogeneous Pattern.	41
Figure 3.7:	Relative Bias Comparison Between Homogeneous and Heterogeneous Patterns at $n = 900$: Intercept Coefficients.	42
Figure 3.8:	Relative Bias Comparison Between Homogeneous and Heterogeneous Patterns at $n = 900$: Slope Coefficients.	43
Figure 3.9:	Root Mean Square Error for Class Membership Coefficients. . . .	44
Figure 3.10:	Root Mean Square Error for Intercept Growth Factor Regression Coefficients Under the Homogeneous Pattern.	45
Figure 3.11:	Root Mean Square Error for Slope Growth Factor Regression Coefficients Under the Homogeneous Pattern.	46
Figure 3.12:	Selection Rates for Class Membership Coefficients.	47
Figure 3.13:	Selection Rates for Intercept Growth Factor Regression Coefficients Under the Homogeneous Pattern.	48
Figure 3.14:	Selection Rates for Slope Growth Factor Regression Coefficients Under the Homogeneous Pattern.	49
Figure 3.15:	Selection Rates for Slope Regression Coefficients Comparing Homogeneous and Heterogeneous Patterns ($n = 150$ and 900). . . .	50

Figure 3.16: Sensitivity of Selection Rates for Slope Growth Factor Regression Coefficients to Selection Criterion ($n = 150$ and 900 , Equal Proportions, Homogeneous Pattern, MD = 3.0).	51
Figure 3.17: Workflow for Selecting a Shrinkage Prior for Variable Selection in Growth Mixture Models.	61
Figure 4.1: Conceptual illustration of the MI-CRP procedure. The top panel shows an incomplete data set in which individuals n_j and n_k have missing values (denoted by “?”) at later time points due to attrition. The middle panel depicts multiple imputation rounds, where the CRP assigns individuals to latent clusters ($C_1, C_2, C_3, \dots$). The number of clusters is not fixed in advance and may vary across rounds, as indicated by the ellipsis. Cluster assignments differ across imputation rounds, reflecting the stochastic nature of the CRP. The bottom panel shows the resulting completed data sets, in which missing values have been replaced by draws from the cluster-specific posterior predictive distributions. Each completed data set contains different imputed values, propagating the uncertainty due to missing data into downstream analyses.	80
Figure 4.2: Path diagram of the two-class linear growth mixture model used as the population model.	83
Figure 4.3: Relative Bias for Growth Factor Means at MD = 2.0.	94
Figure 4.4: Relative Bias for Growth Factor Means at MD = 3.0.	95
Figure 4.5: Relative Bias for Growth Factor Means at MD = 4.0.	96
Figure 4.6: Relative Bias for Growth Factor Variances and Covariance at MD = 2.0.	97
Figure 4.7: Relative Bias for Growth Factor Variances and Covariance at MD = 3.0.	98
Figure 4.8: Relative Bias for Growth Factor Variances and Covariance at MD = 4.0.	99
Figure 4.9: Relative Bias for Class 1 Proportion (Equal Proportions).	100
Figure 4.10: Relative Bias for Class 2 Proportion (Unequal Proportions).	101
Figure 4.11: Root Mean Square Error for Growth Factor Means at MD = 2.0.	102
Figure 4.12: Root Mean Square Error for Growth Factor Means at MD = 3.0.	103
Figure 4.13: Root Mean Square Error for Growth Factor Means at MD = 4.0.	104
Figure 4.14: Root Mean Square Error for Growth Factor Variances and Covariance at MD = 2.0.	105
Figure 4.15: Root Mean Square Error for Growth Factor Variances and Covariance at MD = 3.0.	106
Figure 4.16: Root Mean Square Error for Growth Factor Variances and Covariance at MD = 4.0.	107
Figure 4.17: Root Mean Square Error for Class Proportions.	108
Figure 4.18: Correct Class Assignment Rates Across Simulation Conditions.	109

LIST OF TABLES

Table 3.1:	Summary of Simulation Design Factors and Their Levels.	27
Table 3.2:	Shrinkage Prior Specifications and Hyperparameter Settings. . . .	32
Table 3.3:	Convergence Rates (%) Across Shrinkage Priors, Sample Sizes, and Class Separation Levels.	37
Table 3.4:	Mean Adjusted Rand Index Across Shrinkage Priors, Sample Sizes, and Class Separation Levels.	53
Table 3.5:	Mean WAIC Across Shrinkage Priors, Sample Sizes, Predictor In- tercorrelations, and Class Separation Levels.	54
Table 3.6:	Mean LOOIC Across Shrinkage Priors, Sample Sizes, Predictor In- tercorrelations, and Class Separation Levels.	55
Table 3.7:	Summary of Shrinkage Prior Performance Across Key Dimensions.	60
Table 4.1:	Summary of Simulation Design.	82
Table 4.2:	Prior Specifications for Growth Factor Means.	86
Table 4.3:	Convergence, Degenerate Solution, and Valid Replication Rates from Selected Conditions.	92

ACKNOWLEDGEMENTS

In Bayesian inference, to borrow Lindley’s 1972 phrase, today’s posterior is tomorrow’s prior. This dissertation, for me, is at once the *end of a beginning* and the *beginning of an end*. Every such update is built from countless data: moments too many to name, some that once seemed unrelated, some I thought would last forever, others as fleeting as a breath, every joy and every struggle, all gathered into this moment.

I owe my sincerest gratitude to my dissertation committee: Sarah Depaoli, Fan Jia, Haiyan Liu, and Meng (Chris) Qiu. My advisors, Sarah and Fan, have stood by me from the beginning of my doctoral journey through its completion. They offered more than mentorship. They supported me, protected me when it was needed, brought opportunities to me, and backed the directions I wanted to pursue. A doctorate is not simply about study and research but also about the many professional and personal decisions it entails. They listened and helped me make better choices at each of those turns. I am so grateful for every one of those moments. Haiyan has always kept her door open to whatever questions or thoughts I brought her, whether about research, life, or anything else. I have learned from her both through the depth of her knowledge and through the warmth she has extended as a person. I have grown because of her in ways I cannot measure. Chris contributed his expertise wherever this dissertation called for it, stayed open to every question, and pointed me toward just the right resources when I needed them most. My thanks go to him for all of it.

Beyond my committee, two quant faculty members, Ren Liu and Yueqi Yan, deserve special mention. Ren has given me encouragement and warmth at every turn, always finding the right words to help me believe in myself. I deeply appreciate that. Yueqi has offered steady advice, always kindly asked how I was, and generously shared support and research resources whenever I needed them. I thank him for his kindness.

From outside the quant program, I also thank two faculty members, Rose Scott in the developmental area and Matthew Zawadzki in the health area. Rose, as a member of my faculty advisory committee, checked in on my progress and remained available whenever I needed guidance and advice. Matthew taught me how to communicate and collaborate with applied researchers. Working with him showed me what it means to be part of a team, and the exchanges we had helped me grow as a collaborative methodologist.

Within the program, my fellow graduate students, from the seniors who welcomed me when I arrived to the juniors I now leave behind, have given me a space. I thank Marieke and Luca, who cared for me when I was in my adaptation phase; Yue and Evelyn, who were wonderful office mates; and Alex and Fei, whose check-ins were always a joy. To Minjeong and Tianrun, I entrust the lab going forward, hoping they find in it the same welcome I once did and pass it on in turn. Having come up through the Depaoli and Jia labs, I extend my warmest wishes to their past and present members. Though my visits to SSM 303 grew less frequent in my final year, the memories of those working hours remain.

Let me emphasize that this dissertation was supported in part by the Dissertation Research Award from the American Psychological Association and the Dissertation Research Grant from the Society of Multivariate Experimental Psychology. It is a profound honor to receive funding support from these two organizations. My research could not have been completed without the computational resources their grants made possible. Both organizations have my deepest thanks, and I hope this dissertation advances the field for the academic communities it reaches.

Outside of psychology, I thank my friends Jan-Willem, Bart-Jan, Jinhyun, Jaeyoung, and Soo, whose steady support as friends has sustained me through these years. And, always, my father, mother, and sister. The dedication says what this page cannot.

A decade has passed since I first opened a statistics textbook on my own in 2016. The path since then has carried me through Sungkyunkwan University, Utrecht University, and the University of California, Merced, and will soon carry me into a new phase as Assistant Professor of Quantitative Psychology at the University of Oklahoma. Many mentors, friends, and colleagues have crossed my path along the way, and it is the accumulation of those moments that has brought me here. The posterior I carry forward is richer than any prior I could have placed on myself. So it seems.

VITA

Education

University of California, Merced, USA

Ph.D. in Quantitative Methods, Measurement, and Statistics

May 2026

Dissertation: *Advanced methods for implementation of Bayesian growth mixture modeling*

Advisors: Dr. Sarah Depaoli and Dr. Fan Jia

Utrecht University, The Netherlands

M.Sc. in Methodology and Statistics, *Cum Laude*

July 2021

Thesis: *All models are uncertain, but averaging is useful: Bayesian multi-model inference in structural equation models with bridge sampling*

Supervisor: Prof. Dr. Eric-Jan Wagenmakers (University of Amsterdam)

Sungkyunkwan University, South Korea

B.A. in Psychology, *Highest Honors*

August 2019

Publications

Heo, I., Jia, F., & Depaoli, S. (2026). Bayesian variable selection via shrinkage priors in growth mixture models. *Multivariate Behavioral Research*. Advance online publication. <https://doi.org/10.1080/00273171.2026.2673279>

Heo, I., & Jeon, M. (2026). Rethinking the evaluation of psychological theories: A commentary on Calderon et al.'s registered replication report. *Advances in Methods and Practices in Psychological Science*, 9(2), 1–7. <https://doi.org/10.1177/25152459261440485>

Depaoli, S., **Heo, I.**, Jauregui, M., Liu, H., & Jia, F. (2026). A comprehensive evaluation of model selection indices for class enumeration in Bayesian latent growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, 33(2), 157–176. <https://doi.org/10.1080/10705511.2025.2566135>

Heo, I., Liu, R., Liu, H., Depaoli, S., & Jia, F. (2026). A study of latent state-trait theory framework in piecewise growth models. *Applied Psychological Measurement*, 50(1–2), 21–32. <https://doi.org/10.1177/01466216251360565>

Heo, I., Simons, J.-W., & Liu, H. (2025). A tutorial on Bayesian model averaging for exponential random graph models. *British Journal of Mathematical and Statistical Psychology*. Advance online publication. <https://doi.org/10.1111/bmsp.70007>

- Heo, I.**, Jia, F., & Depaoli, S. (2025). Recovering knot placements in Bayesian piecewise growth models with missing data. *Behavior Research Methods*, 57(7), 1–27. <https://doi.org/10.3758/s13428-025-02716-0>
- Liu, H., **Heo, I.**, Ivanov, A., & Depaoli, S. (2025). Model assumption violations in Bayesian latent mediation analysis: An exploration of Bayesian SEM fit indices and PPP. *Structural Equation Modeling: A Multidisciplinary Journal*, 32(5), 866–896. <https://doi.org/10.1080/10705511.2025.2503789>
- Liu, H., **Heo, I.**, Depaoli, S., & Ivanov, A. (2025). Parameter recovery for misspecified latent mediation models in the Bayesian framework. *Structural Equation Modeling: A Multidisciplinary Journal*, 32(4), 618–637. <https://doi.org/10.1080/10705511.2025.2475490>
- Heo, I.**, Pfadt, J. M., & Wagenmakers, E.-J. (2025). Contributed discussion of “Sparse Bayesian factor analysis when the number of factors is unknown”. *Bayesian Analysis*, 20(1), 295–296. <https://doi.org/10.1214/24-BA1423>
- Heo, I.**, Depaoli, S., Jia, F., & Liu, H. (2024). Bayesian approach to piecewise growth mixture modeling: Issues and applications in school psychology. *Journal of School Psychology*, 107, 101366. <https://doi.org/10.1016/j.jsp.2024.101366>
- Heo, I.**, Jia, F., & Depaoli, S. (2024). Performance of model fit and selection indices for Bayesian piecewise growth modeling with missing data. *Structural Equation Modeling: A Multidisciplinary Journal*, 31(3), 455–476. <https://doi.org/10.1080/10705511.2023.2264514>
- Depaoli, S., Jia, F., & **Heo, I.** (2023). Detecting model misspecification in Bayesian piecewise growth models. *Structural Equation Modeling: A Multidisciplinary Journal*, 30(4), 574–591. <https://doi.org/10.1080/10705511.2022.2144865>
- Heo, I.**, Jia, F., & Depaoli, S. (2023). Book review of *Longitudinal structural equation modeling with Mplus: A latent state-trait perspective* by Geiser. *Psychometrika*, 88(2), 733–737. <https://doi.org/10.1007/s11336-022-09897-z>
- Liu, R., **Heo, I.**, Liu, H., Shi, D., & Jiang, Z. (2023). Applying negative binomial distribution in diagnostic classification models for analyzing count data. *Applied Psychological Measurement*, 47(1), 64–75. <https://doi.org/10.1177/01466216221124604>

ABSTRACT OF THE DISSERTATION

Advanced Methods for Implementation of Bayesian Growth Mixture Modeling

by
Ihnwhi Heo

Doctor of Philosophy in Psychological Sciences
Specialization in Quantitative Methods, Measurement, and Statistics
University of California, Merced, 2026

Professor Sarah Depaoli, Co-Chair
Professor Fan Jia, Co-Chair

Growth mixture models (GMMs) constitute a versatile statistical framework for capturing population heterogeneity in longitudinal data by identifying latent classes of growth trajectories. Since their introduction in 1999, the models have been extended and disseminated through sustained methodological advancements, yet uncharted gaps remain in their implementation. This dissertation develops and evaluates advanced Bayesian methods to address two such gaps, predictor selection and missing data handling, across two studies. In Study 1, I focused on predictor selection in conditional GMMs. Beyond the usefulness of GMMs in describing heterogeneous growth patterns, predicting these patterns offers an enhanced understanding of substantive phenomena. However, the question of how best to select important predictors has remained unaddressed in the GMM literature. As a principled approach to this issue, I proposed Bayesian variable selection via shrinkage priors and evaluated seven priors (ridge, lasso, hyperlasso, elastic net, horseshoe, regularized horseshoe, and spike-and-slab) through a Monte Carlo simulation. Results showed that the horseshoe, regularized horseshoe, and spike-and-slab offered the best balance between detecting true nonzero predictors, shrinking small effects, and controlling false selections. In addition, the regularized horseshoe, spike-and-slab, and horseshoe achieved superior predictive accuracy, though the regularized horseshoe and spike-and-slab came at the cost of lower convergence rates. The ridge prior showed limited effectiveness in selecting important predictors and high false selection rates. Practical recommendations are provided for choosing shrinkage priors, with caveats and other implementation considerations. In Study 2, I addressed missing data in GMMs under attrition scenarios common in longitudinal research. Existing missing data techniques either discard incomplete cases, require the number of latent classes to be pre-specified as part of the missing data treatment, or ignore the population heterogeneity. To overcome these limitations, I developed a Bayesian nonparametric multiple imputation approach using the Chinese restaurant process (MI-CRP). This new approach retains existing values, imputes missing values, and accounts for population heterogeneity

without requiring the number of classes to be predetermined. A Monte Carlo simulation compared MI-CRP with four existing methods, including full information maximum likelihood, Bayesian estimation, multiple imputation via joint modeling, and multiple imputation via fully conditional specification. Results showed that MI-CRP adequately recovered the growth factor means and avoided extreme covariance bias. MI-CRP also achieved higher convergence and valid replication rates at low class separation with smaller sample sizes, and higher classification accuracy with less variability under challenging conditions. Overall, the advanced Bayesian methods developed and evaluated in this dissertation fill unresolved methodological gaps in GMMs and equip applied and methodological researchers with principled tools for implementing these models.

Chapter 1

Overview of Dissertation

Studying change has been an enduring topic of critical importance in psychology specifically and the social and behavioral sciences more broadly. Growth mixture models (GMMs; B. Muthén, 2004; B. Muthén & Shedden, 1999) represent a class of statistical techniques designed to capture change in a nuanced way. These models identify unobserved subgroups, called latent classes, that exhibit qualitatively different patterns of growth over time. GMMs relax the single-population assumption of latent growth curve models (Bollen & Curran, 2006; Meredith & Tisak, 1990) and instead posit that the observed sample is drawn from a finite mixture of latent classes. Each class then has its own developmental trajectory and corresponding parameters. This reformulation yields a richer picture of longitudinal substantive phenomena. Since their introduction in 1999 (B. Muthén & Shedden, 1999), the methods have been disseminated across wide academic communities concerned with the analysis of longitudinal data (e.g., Harring & Hodis, 2016; Heo, Depaoli, et al., 2024; Jung & Wickrama, 2008; Ram & Grimm, 2009). Importantly, as an inferential framework, Bayesian inference has demonstrated substantial utility for GMMs. The Bayesian approach provides advantages such as improved estimation efficiency, tractable estimation of complex nonlinear models, more reliable convergence, stable estimation with small samples, and the ability to incorporate substantive background information (Depaoli, 2013, 2014; Depaoli et al., 2017; Heo, Depaoli, et al., 2024; van de Schoot et al., 2018). Despite these sustained methodological advancements, I identify uncharted gaps that remain in the implementation of these models. Across two studies, this dissertation develops and evaluates advanced Bayesian methods that address these gaps in predictor selection and missing data handling.

1.1 Preview of Studies

Study 1 is entitled “Bayesian Variable Selection via Shrinkage Priors in Growth Mixture Models.” GMMs are useful for researchers to describe heterogeneous growth patterns of latent classes. Beyond this usefulness, researchers can further predict these patterns, which offers an enhanced understanding of substantive phenomena.

However, the methodological question of how to rigorously select important predictors has remained unaddressed in the GMM literature. As a principled approach to this issue, Study 1 proposes Bayesian variable selection via shrinkage priors, which retain substantively important predictors and regularize small or null effects toward zero. The primary goals are to (1) evaluate the performance of seven shrinkage priors (ridge, lasso, hyperlasso, elastic net, horseshoe, regularized horseshoe, and spike-and-slab) through a Monte Carlo simulation study and (2) provide practical recommendations for researchers seeking to select predictors in GMMs.

Study 2 is entitled “Bayesian Nonparametric Multiple Imputation via the Chinese Restaurant Process for Growth Mixture Models.” Missing data, particularly attrition, is a pervasive concern in longitudinal research that can compromise parameter recovery and classification accuracy in GMMs. However, existing missing data techniques for GMMs face limitations. They either discard incomplete cases, require the number of latent classes to be pre-specified as part of the missing data treatment, or ignore the population heterogeneity. To overcome these limitations, Study 2 develops a Bayesian nonparametric multiple imputation approach using the Chinese restaurant process (MI-CRP). The primary goals are to (1) develop a multiple imputation approach that retains existing values, imputes missing values, and accounts for population heterogeneity without requiring the number of classes to be predetermined and (2) evaluate the proposed approach relative to existing methods through a Monte Carlo simulation study.

1.2 Organization

I begin in Chapter 2 by introducing growth mixture modeling as a modeling framework and Bayesian inference as an inferential framework. This chapter integrates these frameworks into Bayesian growth mixture modeling as the underlying methodology. I then present two studies, each in a dedicated chapter (Chapter 3 and Chapter 4). Each chapter includes background specific to the topic, simulation design, simulation results, and discussion. The dissertation concludes with a general discussion in Chapter 5, which articulates the contributions, reflects on limitations and prospects, and closes with an outlook.

Chapter 2

Bayesian Growth Mixture Modeling: A Methodological Framework

The two studies presented in this dissertation share a common methodological foundation that this chapter develops across four stages. The latent growth curve model provides a statistical framework for describing intraindividual change over time. Extending this single-population model to the growth mixture model accommodates population heterogeneity. Bayesian inference then serves as the inferential framework of this dissertation. I close the chapter by integrating these components into Bayesian growth mixture modeling, the foundation on which the two studies build.

2.1 Modeling Intraindividual Change: Latent Growth Curve Models

Longitudinal research draws insights from repeated measurements of the same individuals across multiple time points. One of its goals is to describe intraindividual change, that is, how each individual changes over time. To model this change, researchers need a statistical representation of the individual trajectory. The latent growth curve model (hereafter, latent GCM; Bollen & Curran, 2006; McArdle & Epstein, 1987; Meredith & Tisak, 1990) treats repeated measurements as manifest indicators of continuous latent variables that represent individuals' trajectories. A trajectory is characterized by an intercept factor capturing initial status and one or more slope factors capturing growth over time, collectively known as growth factors. This latent variable formulation carries an important advantage. Because the growth factors are embedded within the structural equation modeling (SEM) framework, the model easily extends to accommodate finite mixture distributions in which individuals are sorted into discrete latent trajectory classes (B. Muthén, 2004; B. Muthén & Shedden, 1999). Although equivalent specifications of the basic growth model are

available within the multilevel modeling framework, it is the latent variable architecture of the SEM approach that provides the most intuitive framework for extending the single-population model to its mixture counterpart (Grimm et al., 2016). Therefore, I adopt the SEM-based latent GCM as the foundational framework and briefly review its specification, establishing the notation that extends into the mixture model section.

The latent GCM decomposes T repeated measures for each individual into a set of continuous latent growth factors and time-specific residuals. In matrix notation, the measurement model can be expressed as:

$$\mathbf{y}_i = \mathbf{\Lambda}_y \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i, \quad (2.1)$$

where $\mathbf{y}_i = (y_{1i}, y_{2i}, \dots, y_{Ti})^\top$ is the $T \times 1$ vector of repeated measures for individual i ($i = 1, \dots, n$), $\boldsymbol{\eta}_i$ is the $m \times 1$ vector of latent growth factors, $\mathbf{\Lambda}_y$ is the $T \times m$ matrix of factor loadings, and $\boldsymbol{\varepsilon}_i$ is the $T \times 1$ vector of time-specific residuals. The simplest formulation of the latent GCM is a linear trajectory with $m = 2$ growth factors: a latent intercept η_{1i} and a latent linear slope η_{2i} . Equation (2.1) can then be written in expanded form as:

$$\begin{bmatrix} y_{1i} \\ y_{2i} \\ \vdots \\ y_{Ti} \end{bmatrix} = \begin{bmatrix} 1 & \lambda_1 \\ 1 & \lambda_2 \\ \vdots & \vdots \\ 1 & \lambda_T \end{bmatrix} \begin{bmatrix} \eta_{1i} \\ \eta_{2i} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \vdots \\ \varepsilon_{Ti} \end{bmatrix}, \quad (2.2)$$

where the factor loadings for the latent intercept are all fixed to 1, reflecting that the intercept contributes equally to each measurement occasion, and the factor loadings λ_t for the latent slope define the temporal spacing of the observations. For instance, with $T = 4$ equally spaced occasions, the slope loadings can be fixed to $\boldsymbol{\lambda} = (0, 1, 2, 3)^\top$, with the first time point conventionally set to zero. Under this specification, the intercept corresponds to the predicted value at the first occasion and the slope represents the expected per-occasion rate of change (Grimm et al., 2016). At the level of a single observation, the model reduces to the scalar form:

$$y_{ti} = \eta_{1i} + \eta_{2i}\lambda_t + \varepsilon_{ti}, \quad (2.3)$$

where $\varepsilon_{ti} \sim \mathcal{N}(0, \sigma_t^2)$ is the time-specific residual, assumed independently distributed across occasions. The residual variance σ_t^2 reflects the magnitude of within-person variability around the modeled trajectory at each time point.

The structural model can then be specified as follows:

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{\zeta}_i, \quad (2.4)$$

where $\boldsymbol{\alpha}$ is the $m \times 1$ vector of population means of the growth factors and $\boldsymbol{\zeta}_i$ is a vector of normally distributed deviations from the population means with covariance matrix $\boldsymbol{\Psi}$. Continuing with the linear latent GCM, the two growth factors jointly follow a bivariate normal distribution:

$$\boldsymbol{\eta}_i = \begin{pmatrix} \eta_{1i} \\ \eta_{2i} \end{pmatrix} \sim \mathcal{N}_2 \left(\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \boldsymbol{\Psi} \right), \quad (2.5)$$

where α_1 and α_2 are the population means of the latent intercept and slope, respectively, and $\boldsymbol{\Psi}$ is the 2×2 covariance matrix of the latent growth factors. The diagonal elements ψ_{11} and ψ_{22} in $\boldsymbol{\Psi}$ capture between-person variability in initial level and rate of change, and the off-diagonal element ψ_{12} captures the covariance between the two factors.

The latent GCM provides a framework for capturing intraindividual change. However, it rests on the assumption that a single underlying growth process governs the entire population. This assumption becomes untenable when the sample is composed of qualitatively distinct subpopulations following different developmental trajectories (Harring & Hodis, 2016; Heo, Depaoli, et al., 2024; Jung & Wickrama, 2008; Ram & Grimm, 2009). Imposing a single-trajectory model in such settings can yield misleading parameter estimates and substantively incorrect conclusions (Jung & Wickrama, 2008; B. Muthén, 1989). Recognizing this limitation motivates the extension to a mixture framework that can accommodate multiple trajectory classes.

2.2 Accommodating Heterogeneity: Latent Growth Mixture Models

The growth mixture model (GMM; B. Muthén, 2004; B. Muthén & Shedden, 1999) extends the latent GCM by introducing a categorical latent variable that partitions the sample into C latent classes, each characterized by a distinct growth trajectory. In doing so, the GMM relaxes the single-population assumption and instead allows the observed data to reflect a mixture of distinct subpopulations (Ram & Grimm, 2009). Under this framework, a latent class refers to an unobserved subgroup of individuals whose membership is inferred through the mixture model, capturing a grouping structure that is not directly measured in the data.

To formalize, the GMM introduces a discrete latent variable $c_i \in \{1, 2, \dots, C\}$ indexing the class membership of individual i . Each class c occurs with probability $\pi_c = \Pr(c_i = c)$, where the mixing proportions satisfy $\sum_{c=1}^C \pi_c = 1$. The observed data for individual i are then modeled as a finite mixture density:

$$f(\mathbf{y}_i | \boldsymbol{\Omega}) = \sum_{c=1}^C \pi_c f_c(\mathbf{y}_i | \boldsymbol{\theta}_c), \quad (2.6)$$

where each component density $f_c(\mathbf{y}_i | \boldsymbol{\theta}_c)$ is weighted by the corresponding mixing proportion π_c , and $\boldsymbol{\Omega} = (\boldsymbol{\pi}^\top, \boldsymbol{\theta}_1^\top, \dots, \boldsymbol{\theta}_C^\top)^\top$ collects all model parameters. The class-specific density $f_c(\cdot)$ is assumed to be multivariate normal with class-specific mean and covariance defined below.

Building on the SEM framework established for the latent GCM, the measurement model extends to accommodate class-specific parameters. Within each class,

the relationship between observed measures and latent growth factors takes the form:

$$\mathbf{y}_{ic} = \mathbf{\Lambda}_y \boldsymbol{\eta}_{ic} + \boldsymbol{\varepsilon}_{ic}, \quad \boldsymbol{\varepsilon}_{ic} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Theta}_c), \quad (2.7)$$

where the factor loading matrix $\mathbf{\Lambda}_y$ remains invariant across classes, as specified in Equation (2.2), and the residual covariance matrix $\boldsymbol{\Theta}_c$ is permitted to vary by class.

At the structural model level, the single-population structural model in Equation (2.4) generalizes to allow class-specific parameters:

$$\boldsymbol{\eta}_{ic} = \boldsymbol{\alpha}_c + \boldsymbol{\zeta}_{ic}, \quad \boldsymbol{\zeta}_{ic} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_c), \quad (2.8)$$

where the population mean vector $\boldsymbol{\alpha}$ is replaced by the class-specific mean vector $\boldsymbol{\alpha}_c = (\alpha_1^{(c)}, \alpha_2^{(c)})^\top$, with $\alpha_1^{(c)}$ and $\alpha_2^{(c)}$ representing the mean intercept and mean slope for class c , respectively. The covariance matrix $\boldsymbol{\Psi}_c$ captures within-class variability in the growth factors. Combining Equations (2.7) and (2.8), the model-implied mean and covariance for class c are

$$\boldsymbol{\mu}_c = \mathbf{\Lambda}_y \boldsymbol{\alpha}_c, \quad \boldsymbol{\Sigma}_c = \mathbf{\Lambda}_y \boldsymbol{\Psi}_c \mathbf{\Lambda}_y^\top + \boldsymbol{\Theta}_c, \quad (2.9)$$

which define the class-specific density $f_c(\cdot)$ in Equation (2.6) as multivariate normal with mean $\boldsymbol{\mu}_c$ and covariance $\boldsymbol{\Sigma}_c$. The class indices on $\boldsymbol{\Psi}_c$ and $\boldsymbol{\Theta}_c$ permit the within-class covariance structure to vary across latent classes. Although equality constraints across classes are not required by the mixture formulation, researchers frequently impose them in practice to simplify estimation and stabilize parameter recovery.

The GMM specified without predictors is referred to as the unconditional GMM because class membership and growth factor distributions are determined solely by the observed repeated measures. When predictors are incorporated to explain class membership or within-class growth, the model becomes a conditional GMM. Chapter 3 builds on the conditional GMM to examine predictor selection via shrinkage priors, and Chapter 4 builds on the unconditional GMM to address missing data handling.

2.3 Updating Beliefs with Data: Bayesian Inference

The preceding sections established the model specification, progressing from the latent GCM to the GMM that accommodates population heterogeneity. I now present the Bayesian inferential framework that underlies this dissertation. This section provides an overview of Bayesian estimation, followed by an introduction to Markov chain Monte Carlo as the computational machinery that makes Bayesian inference tractable for most models of interest.

2.3.1 Overview of Bayesian Estimation

The Bayesian approach to statistical inference combines prior knowledge with observed data to obtain updated beliefs about the parameters of interest (Gelman,

Carlin, et al., 2014; Kaplan, 2023). Central to this approach is Bayes' theorem, expressed as

$$p(\boldsymbol{\theta} \mid \mathbf{Y}) = \frac{p(\mathbf{Y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{Y})}, \quad (2.10)$$

where $p(\mathbf{Y} \mid \boldsymbol{\theta})$ is the likelihood function quantifying the density of the observed data given the parameters, $p(\boldsymbol{\theta})$ is the prior distribution reflecting beliefs about the parameters before observing the data, and $p(\mathbf{Y}) = \int p(\mathbf{Y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$ is the marginal likelihood serving as a normalizing constant that ensures the posterior distribution integrates to one. The left-hand side, $p(\boldsymbol{\theta} \mid \mathbf{Y})$, is the posterior distribution, representing updated beliefs about the parameters after observing the data. Because $p(\mathbf{Y})$ does not depend on $\boldsymbol{\theta}$, Bayes' theorem is often expressed in its unnormalized form:

$$p(\boldsymbol{\theta} \mid \mathbf{Y}) \propto p(\mathbf{Y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}). \quad (2.11)$$

The posterior distribution thus synthesizes two sources of information: what the data reveal through the likelihood and what is known or assumed a priori through the prior.

Posterior inference proceeds by summarizing the posterior distribution, typically through point estimates (e.g., the posterior mean or median) and interval estimates (e.g., the highest posterior density interval or quantile-based credible intervals such as the 95% equal-tailed credible interval). For most models of interest, however, the posterior distribution does not admit a closed-form expression. The next subsection introduces Markov chain Monte Carlo methods, which provide a general computational strategy for approximating the posterior when closed-form derivation is not possible.

2.3.2 Markov Chain Monte Carlo

When the posterior distribution lacks a closed-form expression, simulation offers a general route to approximation. Markov chain Monte Carlo (MCMC) is an iterative sampling scheme that produces draws from the posterior distribution through a Markov chain whose stationary distribution is the target posterior (Gamerman & Lopes, 2006; Robert & Casella, 2004). Rather than deriving the posterior analytically, MCMC generates a sequence of parameter draws that, after a sufficient number of iterations, can be treated as a sample from $p(\boldsymbol{\theta} \mid \mathbf{Y})$. These draws then serve as the basis for posterior summaries such as point estimates and credible intervals.

Among various MCMC algorithms, the Gibbs sampler is a widely used choice when the full conditional distribution of each parameter can be derived. Let $\boldsymbol{\theta}$ denote a parameter vector comprising p components, $\boldsymbol{\theta}^{(s)} = (\theta_1^{(s)}, \theta_2^{(s)}, \dots, \theta_p^{(s)})^\top$, where the superscript (s) indexes the $s = 1, 2, \dots, S$ iteration of the chain. Starting from an initial set of values $\boldsymbol{\theta}^{(0)}$, the sampler updates each parameter in turn by drawing from its full conditional distribution given the most recent values of the remaining parameters and the observed data. The general procedure at iteration $s + 1$ can be

expressed as

$$\begin{aligned}
\theta_1^{(s+1)} &\sim p(\theta_1 \mid \theta_2^{(s)}, \theta_3^{(s)}, \dots, \theta_p^{(s)}, \mathbf{Y}), \\
\theta_2^{(s+1)} &\sim p(\theta_2 \mid \theta_1^{(s+1)}, \theta_3^{(s)}, \dots, \theta_p^{(s)}, \mathbf{Y}), \\
\theta_3^{(s+1)} &\sim p(\theta_3 \mid \theta_1^{(s+1)}, \theta_2^{(s+1)}, \theta_4^{(s)}, \dots, \theta_p^{(s)}, \mathbf{Y}), \\
&\vdots \\
\theta_p^{(s+1)} &\sim p(\theta_p \mid \theta_1^{(s+1)}, \theta_2^{(s+1)}, \dots, \theta_{p-1}^{(s+1)}, \mathbf{Y}),
\end{aligned} \tag{2.12}$$

where each parameter is sampled from its full conditional using the most recently updated values for the other parameters and the data $\mathbf{Y}$. Iterating this procedure produces a Markov chain of posterior samples $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(S)}\}$.

The initial draws in the chain do not yet reflect the target posterior because the sampler has not traveled far enough from its starting point. An initial segment of the chain is therefore discarded as burn-in, and the remaining draws are used for posterior inference. Convergence of the chain is assessed through diagnostics such as the Gelman-Rubin $\hat{R}$ statistic (Gelman & Rubin, 1992), with values sufficiently close to one indicating that the chain has reached its stationary distribution.

2.4 Integrating Model and Inference: Bayesian Growth Mixture Modeling

Having established the GMM as a statistical framework for heterogeneous growth and Bayesian inference as the inferential framework of this dissertation, I now integrate these components in Bayesian growth mixture modeling. This section first elaborates on the advantages of applying Bayesian methods to GMMs along with the considerations that accompany their use, and then presents the prior specifications commonly adopted for GMM parameters.

2.4.1 Advantages and Considerations

The Bayesian framework has several distinct advantages for estimating GMMs. Because Bayesian estimation does not rely on large-sample approximations for valid inference, it is well suited to applied settings where sample sizes are limited (van de Schoot et al., 2018). Bayesian methods can improve estimation efficiency, convergence, parameter recovery, and class assignment, with benefits most pronounced when sample sizes are small or class separation is low (Depaoli, 2013, 2014). These gains are consequential for recovering small but substantively meaningful latent classes that may be difficult to detect under conventional estimation. The Bayesian framework can further accommodate more complex GMM specifications involving nonlinear trajectory structures (Heo, Depaoli, et al., 2024).

Central to these advantages is the role of the prior distribution $p(\boldsymbol{\theta})$, as reflected in Equation (2.10). The choice of prior can have direct consequences for estimation in GMMs. Accurately specified informative priors have been shown to improve parameter recovery and class assignment (Depaoli, 2013), and growth parameter estimates appear relatively robust to misspecified prior means provided that the variance hyperparameters are reasonably large (Depaoli, 2014). At the same time, substantive conclusions from mixture models can shift as prior settings are modified, underscoring the importance of sensitivity analysis (Depaoli et al., 2017). For empirical applications, van de Schoot et al. (2018) provided a practical roadmap for deriving informative priors from systematic literature searches and expert elicitation.

2.4.2 Prior Specifications

For Bayesian GMMs, prior distributions are specified for each set of model parameters (see Depaoli, 2021, for details):

$$\boldsymbol{\pi} \sim \text{Dirichlet}(d_1, \dots, d_C), \quad (2.13)$$

$$\alpha_{\ell c} \sim \mathcal{N}(\mu_{\alpha_{\ell c}}, \sigma_{\alpha_{\ell c}}^2), \quad (2.14)$$

$$\sigma_t^2 \sim \text{Inverse-Gamma}(a_t, b_t), \quad (2.15)$$

$$\boldsymbol{\Psi}_c \sim \text{Inverse-Wishart}(\mathbf{S}, \nu). \quad (2.16)$$

The Dirichlet prior is placed on the class proportion vector $\boldsymbol{\pi}$, and its hyperparameters $(d_1, \dots, d_C)$ encode prior expectations about the relative sizes of the latent classes. Larger hyperparameter values concentrate prior mass around the implied proportions, whereas a symmetric choice such as $(1, 1, \dots, 1)$ supplies minimal information about relative class sizes. In practice, the hyperparameters may be specified in terms of the proportion or the number of cases expected in each class.

The normal prior is placed on each class-specific growth factor mean $\alpha_{\ell c}$ for growth factor ℓ in class c . The mean hyperparameter $\mu_{\alpha_{\ell c}}$ anchors the center of the prior, and the variance hyperparameter $\sigma_{\alpha_{\ell c}}^2$ captures the dispersion of plausible values around that center. Small values of $\sigma_{\alpha_{\ell c}}^2$ concentrate prior mass near $\mu_{\alpha_{\ell c}}$ and convey stronger prior beliefs, and larger values produce a more diffuse prior.

The inverse-gamma prior is placed on each time-specific residual variance σ_t^2 . Its hyperparameters a_t and b_t are the shape and scale parameters of the inverse-gamma distribution, both constrained to be strictly positive to ensure a proper prior density. Different combinations of these hyperparameters yield priors ranging from weakly informative to strictly informative for the residual variance.

The inverse-Wishart prior is placed on the within-class growth factor covariance matrix $\boldsymbol{\Psi}_c$. Its hyperparameters are the positive-definite scale matrix $\mathbf{S}$ and the degrees-of-freedom parameter ν . The scale matrix encodes prior expectations about the magnitude and correlation of within-class variability in the growth factors, and the degrees of freedom ν govern the informativeness of the prior. With ν set to small values, the inverse-Wishart prior becomes more diffuse, and with larger values the prior conveys stronger prior beliefs.

2.5 Implementation Gaps in Bayesian Growth Mixture Modeling

This dissertation addresses two implementation gaps that have persisted in Bayesian growth mixture modeling. In Study 1 (Chapter 3), I propose Bayesian variable selection via shrinkage priors as a principled approach to predictor selection in the conditional GMM, and examine the performance of seven shrinkage priors (ridge, lasso, hyperlasso, elastic net, horseshoe, regularized horseshoe, and spike-and-slab) through a Monte Carlo simulation. In Study 2 (Chapter 4), I develop a Bayesian nonparametric multiple imputation approach using the Chinese restaurant process for handling attrition-pattern missing data in the unconditional GMM, and compare this approach against existing missing data methods through a Monte Carlo simulation. Together, these studies extend the methodological foundation developed in this chapter and aim to advance the implementation of Bayesian growth mixture modeling for applied and methodological research.

Chapter 3

Bayesian Variable Selection via Shrinkage Priors in Growth Mixture Models

3.1 Introduction

3.1.1 Modeling Change and Uncovering Heterogeneity

One of the central questions of longitudinal research in the social and behavioral sciences concerns how individuals or other units of analysis change over time. With substantive research questions and longitudinal data collected, researchers ask how to model change using an appropriate statistical framework. Among many, the latent growth curve model (hereafter, latent GCM; Bollen & Curran, 2006; Meredith & Tisak, 1990) has become one of the standard tools for this purpose. The latent GCM captures interindividual differences in intraindividual change by assuming a specific form (e.g., linear, quadratic, piecewise) for growth as a function of time. Here, the growth factors that characterize each individual's trajectory, such as initial status and rate of change, are treated as latent variables whose values are inferred from the pattern of repeated observations over time. This framework is flexible, as it can be extended by relating other observed and latent variables for more complex models (Bollen & Curran, 2006; Grimm et al., 2016).

The latent GCM assumes that all individuals in the sample belong to a single population described by one set of growth parameters (Bollen & Curran, 2006; Jung & Wickrama, 2008; Ram & Grimm, 2009). In practice, populations may comprise individuals who follow qualitatively different trajectories of change. Researchers have long been interested in understanding such heterogeneity in longitudinal data by partitioning populations into subgroups (Curran et al., 2010; Lubke & Muthén, 2005). One common strategy is to define subgroups using observed characteristics such as demographic variables (e.g., gender, ethnicity) or ad hoc decision rules such as arbitrary cutoff scores (e.g., classifying individuals who score above a certain threshold

as “high risk”). When appropriate grouping variables are available and well justified, this approach can be informative.

However, for many substantive phenomena, important grouping variables may remain unobserved altogether (Lubke & Muthén, 2005; B. Muthén, 1989). Indeed, substantively meaningful heterogeneity in longitudinal data may exist along dimensions that are not directly captured by observed grouping variables (Harring & Hodis, 2016; Heo, Depaoli, et al., 2024; Jung & Wickrama, 2008; Ram & Grimm, 2009). In such scenarios, a complementary strategy is to identify subgroup structure empirically, based on the patterns present in the data, rather than relying on observed grouping variables or predefined classifications. When researchers seek to study longitudinal change within such empirically identified subgroups, they can turn to growth mixture models (GMMs; B. Muthén, 2004; B. Muthén & Shedden, 1999). The GMM posits a finite number of latent classes of growth trajectories, each with its own set of growth parameters. Understanding and examining these classes is the primary focus of the GMM framework, which I elaborate on in the following subsection.

3.1.2 Describing and Predicting Heterogeneous Growth

Describing longitudinal phenomena is a foundational step toward understanding developmental processes. In GMMs, description of who belongs to which classes and how those classes express distinct developmental patterns is essential to any substantive interpretation of the mixture solution and forms the basis on which subsequent analysis builds. To illustrate, researchers may distinguish classes by their trajectory shapes (e.g., increasing, declining, or stable), relative class sizes, and growth parameter estimates to understand the nature of heterogeneity in growth. Yet substantive interest more often than not extends beyond mere description and further focuses on predicting developmental phenomena (Curran et al., 2010; Nylund-Gibson & Choi, 2018; Petras & Masyn, 2010).¹ When the focus shifts toward prediction, the central question concerns what predicts latent class membership and what accounts for within-class variation in growth. Petras and Masyn (2010) articulated this motivation, noting that GMMs gain much of their practical utility when predictors are introduced to explain and contextualize the identified classes in addition to description.

The inclusion of predictors is consequential for the GMM analysis from both applied and methodological perspectives. From an applied perspective, predictors clarify the sources of differences in growth trajectories both across and within classes

¹In some applied and methodological work, predictors have been described as causes or determinants of development. However, as McCormick et al. (2024) emphasized, the validity of such causal language depends on the study design (e.g., experimental or well-defended quasi-experimental designs that address confounding) rather than on the statistical model itself. Regression coefficients in conditional GMMs therefore capture associations between covariates and growth factors, and their interpretation as causal effects requires a design-based warrant that the method alone cannot provide. Accordingly, I use the term *predictors* throughout. When previous work uses terms such as *determinants* or *causes*, the interpretive warrant rests on those authors’ study designs.

and illuminate the determinants of latent class membership, providing context for why individuals follow distinct developmental pathways. This interest is reflected in the applied literature, where researchers across diverse disciplines have incorporated predictors into GMMs to enrich their substantive interpretations (see, e.g., Frankfurt et al., 2016; Gomez et al., 2017; Huang et al., 2010; L. C. Liu et al., 2010; Magee et al., 2013; Wake et al., 2023; Wang & Bodner, 2007).

From a methodological perspective, the specification of predictors carries implications that extend across the entire analysis, from parameter recovery and classification accuracy to class enumeration. As such, a growing body of methodological work has evaluated the impact of predictor inclusion and specification on the performance of GMMs in these regards (e.g., Diallo et al., 2017a, 2017b; Hu et al., 2017; L. Li & Hser, 2011; M. Li & Harring, 2017). These studies show that the choice of which predictors to include, and how to specify their effects, has substantive consequences for the mixture solution. For example, omitting predictors of growth factors can violate within-class distributional assumptions and compromise class enumeration accuracy (Diallo et al., 2017a; L. Li & Hser, 2011). Similarly, introducing predictors at this stage, or misspecifying the path through which they operate, can distort the recovered class solution (Diallo et al., 2017a; Hu et al., 2017). Collectively, these studies reinforce that careful attention to predictors is imperative for both the accuracy and interpretability of the mixture solution.

3.1.3 Problem of Predictor Selection and Bayesian Solutions

Despite this recognition, a systematic guideline for predictor selection in GMMs has not been established, as discussed in Huang et al. (2010). Researchers may hold specific hypotheses about the role of predictors and include them on confirmatory grounds (e.g., Barboza, 2020; Gomez et al., 2017; Magee et al., 2013; B. Muthén et al., 2002; Rodriguez et al., 2005), or they may pursue exploratory analyses to identify potentially important predictors (e.g., Caesens et al., 2020; Nagin, 1999; Wake et al., 2023). In the absence of formal guidelines, the rationale underlying predictor selection in applied GMM research often remains implicit. Research conclusions may then be drawn by comparing model fit between specifications with and without predictors, in the absence of an explicit confirmatory or exploratory justification for the predictors considered. This ambiguity in current practice is not a shortcoming of individual studies but rather a reflection of the lack of available methodological tools. Addressing this gap calls for a principled approach to predictor selection in GMMs.

A natural recourse might be traditional variable selection strategies such as stepwise regression, forward selection, and backward elimination. Yet these methods have well-documented limitations that undermine their utility (e.g., Derksen & Keselman, 1992; Thompson, 1995; Whittingham et al., 2006). These shortcomings include susceptibility to overfitting through the retention of sample-specific predictors, compromised generalization to new data, and inflated parameter estimates with biased standard errors. In GMMs, such risks take on added weight. Because predic-

tor specification can shape the recovered class solution, iteratively refitting the model across candidate predictor sets can lead to class instability, as the number of classes and their substantive interpretation shift from one candidate model to the next. Iterative selection also requires full mixture estimation at each step with multiple random starts, compounding the computational burden and increasing exposure to convergence failures (Hipp & Bauer, 2006; Jung & Wickrama, 2008). What is needed, then, is an approach that provides guidance where the decision has historically been left implicit and that can be integrated into confirmatory and exploratory analytic frameworks alike.

To address the predictor selection problem, I propose regularization, an estimation approach designed to reduce overfitting and enable variable selection (Jacobucci et al., 2023). Regularization adds a penalty term that introduces bias and ultimately balances bias and variance in parameter estimation. Without such a penalty, GMMs with many predictors may achieve low bias but suffer from high variability in the results, limiting generalizability. With a penalty in place, a degree of bias is deliberately introduced in exchange for reduced variability, yielding more stable and generalizable estimates. As a regularization approach for GMMs, I adopt Bayesian variable selection via shrinkage priors (Kaplan, 2023). Shrinkage priors are prior distributions centered around zero that function as penalty terms, controlling the degree of shrinkage applied to each regression coefficient. Under this framework, shrinkage is applied selectively by effect magnitude. Small effects that likely reflect noise (i.e., coefficients with true values at or near zero) are shrunk toward zero, and moderate to large effects that likely represent genuine signals (i.e., coefficients with substantively nonzero true values) are retained with only partial shrinkage. This represents a principled approach to identifying substantively meaningful predictors, and one that can be integrated into both confirmatory and exploratory analytic frameworks (Jacobucci et al., 2019).

As van Erp et al. (2019) outlined, adopting a Bayesian perspective on variable selection offers several practical advantages. First, it is straightforward to implement through shrinkage priors placed on model parameters (e.g., regression coefficients), as prior distributions are inherently a component of Bayesian inference. Second, the approach is flexible, as estimation of model parameters and variable selection proceed simultaneously within a single modeling step. Third, it is versatile, as many different distributions of shrinkage priors can be specified. These properties make shrinkage priors an ideal candidate for addressing the predictor selection problem in GMMs. Indeed, Bayesian regularization via shrinkage priors has already demonstrated its utility across a range of latent variable models, including factor analysis (Jacobucci & Grimm, 2018), latent growth curve models (Harra & Kaplan, 2024; Jacobucci & Grimm, 2018), path analysis (Harra & Kaplan, 2024), multiple-indicators multiple-causes (MIMIC) models (Zhang & Liang, 2024), and mediation models (Shi et al., 2023).

This chapter advances the Bayesian regularization via shrinkage priors framework to GMMs, where predictor selection is complicated by the presence of latent

class structure. In doing so, it contributes to the growing body of regularization research within the latent variable modeling literature. Furthermore, as noted in Chapters 1 and 2, Bayesian estimation offers distinct advantages for growth mixture modeling, further supporting the suitability of a Bayesian approach for addressing the predictor selection problem (Depaoli, 2013, 2014; Depaoli et al., 2017; Heo, Depaoli, et al., 2024; van de Schoot et al., 2018).

3.1.4 Goals and Structure of the Present Investigation

This chapter has two goals. The first is to evaluate the performance of Bayesian shrinkage priors for the selection of important predictor variables in GMMs through a Monte Carlo simulation study. The second is to provide practical guidelines for applied researchers on predictor selection using shrinkage priors, enriching the implementation and application of GMMs with predictors. My hope is that the current findings will help researchers make more informed decisions regarding which predictors to retain and demonstrate the promise of shrinkage priors within the mixture modeling framework.

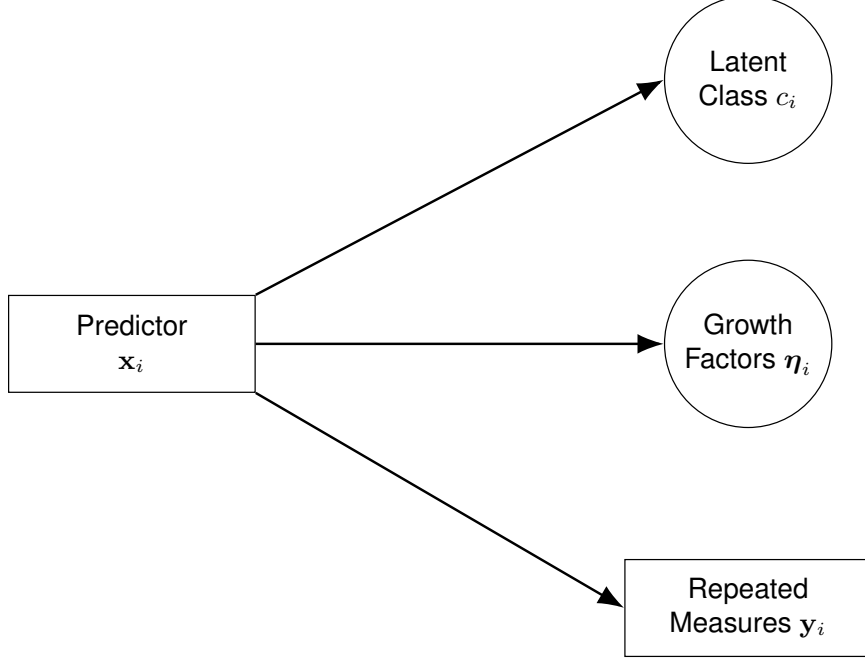
I outline the structure of this chapter in what follows. Building on the unconditional GMM introduced in Chapter 2, I begin by extending the model to the conditional GMM in which predictors are specified for both latent class membership and within-class growth factors. I then introduce Bayesian variable selection via shrinkage priors, leveraging the Bayesian inferential framework presented in Chapter 2. Next, I introduce and review previous research on seven shrinkage prior specifications that serve as the focal priors in this investigation: ridge, lasso, hyperlasso, elastic net, horseshoe, regularized horseshoe, and spike-and-slab. I present the design of a Monte Carlo simulation study and subsequently present the simulation results. Finally, I discuss the main findings, offer practical guidance for applied researchers, and chart directions for future research avenues.

3.2 Incorporating Predictors into Growth Mixture Models

Chapter 2 introduced the unconditional GMM that partitions the population into latent classes based on the observed repeated measures but does not address what predicts class membership or within-class variation in growth. As motivated in the Introduction, incorporating predictors into the GMM serves this purpose by relating observed variables to the latent class structure and the growth process within each class. Predictors can in principle influence three components of the growth mixture model: (a) latent class membership, (b) within-class growth factors, and (c) repeated measures directly as time-varying covariates. Figure 3.1 illustrates these three pathways. Consistent with this chapter’s focus on latent class membership and within-class growth heterogeneity, the first two pathways are examined: Predictors

enter the model through (a) multinomial logistic regression for latent class membership and (b) linear regression for the within-class growth factors.

Figure 3.1: Conceptual diagram of three pathways through which predictors can influence outcomes in a growth mixture modeling framework: latent class membership through multinomial logistic regression, within-class growth factors through linear regression, and repeated measures directly as a time-varying covariate effect.



For the first pathway, let $\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{Ji})^\top$ denote a $J \times 1$ vector of predictors for individual i . The probability of membership in class c is modeled as a multinomial logistic regression:

$$\Pr(c_i = c \mid \mathbf{x}_i) = \frac{\exp(\gamma_{0c} + \boldsymbol{\gamma}_c^\top \mathbf{x}_i)}{\sum_{c'=1}^C \exp(\gamma_{0c'} + \boldsymbol{\gamma}_{c'}^\top \mathbf{x}_i)}, \quad (3.1)$$

where γ_{0c} is the class-specific intercept and $\boldsymbol{\gamma}_c = (\gamma_{1c}, \gamma_{2c}, \dots, \gamma_{Jc})^\top$ is the vector of logistic regression coefficients linking the predictors to class membership. For identification, the parameters of a reference class (e.g., class C) are constrained to zero: $\gamma_{0C} = 0$ and $\boldsymbol{\gamma}_C = \mathbf{0}$. This is needed because the class probabilities depend only on how the γ coefficients differ across classes, so without a reference, many different coefficient sets could produce the same probabilities and the parameters would not be uniquely identified. When $C = 2$, the multinomial logistic regression reduces to a binomial logistic regression with a single set of coefficients.

For the second pathway, the structural model in Equation (2.8) is extended to include predictors as determinants of the latent growth factors within each class:

$$\boldsymbol{\eta}_{ic} = \boldsymbol{\alpha}_c + \mathbf{B}^{(c)} \mathbf{x}_i + \boldsymbol{\zeta}_{ic}, \quad \boldsymbol{\zeta}_{ic} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_c), \quad (3.2)$$

where $\mathbf{B}^{(c)}$ is an $m \times J$ matrix of class-specific regression coefficients. For the linear growth model with $m = 2$ growth factors and J predictors, this can be written as

$$\begin{pmatrix} \eta_{1i}^{(c)} \\ \eta_{2i}^{(c)} \end{pmatrix} = \begin{pmatrix} \alpha_1^{(c)} \\ \alpha_2^{(c)} \end{pmatrix} + \begin{pmatrix} \beta_{11}^{(c)} & \beta_{12}^{(c)} & \cdots & \beta_{1J}^{(c)} \\ \beta_{21}^{(c)} & \beta_{22}^{(c)} & \cdots & \beta_{2J}^{(c)} \end{pmatrix} \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{Ji} \end{pmatrix} + \begin{pmatrix} \zeta_{1i}^{(c)} \\ \zeta_{2i}^{(c)} \end{pmatrix}, \quad (3.3)$$

where $\beta_{\ell j}^{(c)}$ denotes the regression coefficient of predictor j on growth factor ℓ in class c . Because $\mathbf{B}^{(c)}$ is indexed by class, the same predictor can exert different effects on the growth factors depending on the class to which an individual belongs. Substantively, this means that the same predictor can produce effects that vary in magnitude or direction on initial status or change across trajectory subgroups (e.g., Gomez et al., 2017; Wake et al., 2023). Constraining $\mathbf{B}^{(c)}$ to be equal across classes would instead assume homogeneous predictor effects and rule out such cross-class heterogeneity. Together, Equations (3.1) and (3.2) define the conditional GMM, in which the regression coefficients γ_c and $\mathbf{B}^{(c)}$ are the parameters on which shrinkage priors will be placed for the purpose of variable selection.

3.3 Bayesian Variable Selection via Shrinkage Priors

Chapter 2 introduced Bayesian inference as the inferential framework of this dissertation, including Bayes' theorem, posterior summaries, and Markov chain Monte Carlo (MCMC) approximation of the posterior. This section considers a specific application of the prior distribution at the core of this chapter: the regression coefficients γ_c and $\mathbf{B}^{(c)}$ of the conditional GMM introduced in the previous section. In a regression model with J predictors, the variable selection problem concerns determining which of the J coefficients represent genuine effects (i.e., coefficients with nonzero true values) and which reflect noise (i.e., coefficients with true values at or near zero). A common assumption motivating variable selection is *sparsity*, the expectation that only a subset of the available predictors have truly nonzero effects on the outcome, and that the remaining coefficients are zero or negligibly small (Hastie et al., 2015; Kaplan, 2023). Sparsity is especially applicable in GMMs. Whether predictors are included on confirmatory or exploratory grounds, each predictor enters many regression coefficients, only a portion of which are truly nonzero at the population level and substantively meaningful for predicting class membership or within-class growth. Nonzero estimates of the remaining coefficients in any single sample can reflect sampling variation rather than true effects. Formally, this assumption applies to both the multinomial logistic regression coefficients $\gamma_1, \dots, \gamma_J$ predicting class membership and the within-class regression coefficients $\beta_{11}^{(c)}, \dots, \beta_{1J}^{(c)}, \dots, \beta_{m1}^{(c)}, \dots, \beta_{mJ}^{(c)}$ linking each of the m growth factors to each of the J predictors.

Bayesian variable selection via shrinkage priors approaches this problem by placing prior distributions on the regression coefficients that pull small effects toward zero while leaving large effects relatively unaffected (Kaplan, 2023; van Erp et al., 2019). This shrinkage reflects the bias-variance tradeoff inherent in regularization. Without shrinkage, coefficient estimates may exhibit low bias but high variance, particularly when the number of predictors is large relative to the sample size. By introducing a controlled amount of bias through the prior, shrinkage reduces the variance of the estimates, yielding more stable and generalizable results. The degree and form of this shrinkage depend on the specific prior chosen. In general, shrinkage priors on a generic regression coefficient β_j ($j = 1, \dots, J$) can be expressed as scale mixtures of normal distributions (West, 1987),

$$\beta_j \mid \tau_j^2 \sim \mathcal{N}(0, \tau_j^2), \quad \tau_j^2 \sim g(\cdot), \quad (3.4)$$

where τ_j^2 is a predictor-specific variance parameter and $g(\cdot)$ is a mixing distribution that determines the shape of the resulting marginal prior on β_j . Different choices of $g(\cdot)$ yield different shrinkage priors, and this shape governs their shrinkage behavior. The prior’s density near zero controls how strongly small coefficients are shrunk and whether it can produce exact or near-exact zeros. The prior’s tail behavior controls how large coefficients are treated. Heavier tails leave genuine effects largely unattenuated, and lighter tails shrink all coefficients more uniformly regardless of magnitude. Finally, whether $g(\cdot)$ introduces a single global scale parameter or predictor-specific local scales distinguishes priors that apply uniform shrinkage from those that apply adaptive shrinkage across predictors.

In the conditional GMM specified in Equations (3.1) and (3.2), the scale mixture formulation applies to two sets of regression coefficients. For the logistic regression coefficients predicting class membership, the shrinkage prior is placed on each γ_{jc} ,

$$\gamma_{jc} \mid \tau_{\gamma,jc}^2 \sim \mathcal{N}(0, \tau_{\gamma,jc}^2), \quad \tau_{\gamma,jc}^2 \sim g(\cdot), \quad (3.5)$$

and for the within-class regression coefficients predicting the growth factors, the same formulation is applied to each $\beta_{\ell j}^{(c)}$,

$$\beta_{\ell j}^{(c)} \mid \tau_{\beta,\ell j}^{(c)2} \sim \mathcal{N}(0, \tau_{\beta,\ell j}^{(c)2}), \quad \tau_{\beta,\ell j}^{(c)2} \sim g(\cdot). \quad (3.6)$$

Under this formulation, coefficients associated with noise predictors are shrunk toward zero, and coefficients reflecting genuine effects are preserved with minimal attenuation. The choice of $g(\cdot)$ thus directly shapes the variable selection properties of the prior.

Shrinkage priors were originally developed in the regression modeling context and have since been applied across a range of statistical models, including latent variable models. Implementing this principled approach to GMMs, where predictor selection is complicated by the presence of latent class structure, has not yet been undertaken. This gap motivates the present investigation. By applying shrinkage simultaneously to the class-membership coefficients γ_c and within-class regression coefficients $\mathbf{B}^{(c)}$, this approach brings principled variable selection to a modeling context

where predictor relevance shapes both class assignment and the form of within-class trajectories.

3.4 Shrinkage Prior Specifications Under Investigation

To examine how different choices of $g(\cdot)$ perform in the conditional GMM, I consider seven shrinkage prior specifications that span a range of shrinkage behaviors. Among the seven priors, the first four (ridge, lasso, hyperlasso, and elastic net) have their origins in the frequentist literature, where they were introduced as penalty terms in the optimization of the loss function (Griffin & Brown, 2011; Hastie et al., 2015; Hoerl & Kennard, 1970; Tibshirani, 1996; Zou & Hastie, 2005). For their Bayesian counterparts, prior distributions are specified in a way that produces the same penalization behavior. The remaining three priors (horseshoe, regularized horseshoe, and spike-and-slab) were developed within the Bayesian framework itself and do not have direct frequentist analogs (Carvalho et al., 2010; George & McCulloch, 1993; Mitchell & Beauchamp, 1988; Piironen & Vehtari, 2017). For a comprehensive review of Bayesian variable selection methods more broadly, readers may consult O’Hara and Sillanpää (2009).

Throughout, I use β_j (as defined in Equation 3.4), σ^2 for the residual variance,² and λ for the global penalty parameter. The specifications are applied identically to both the logistic regression coefficients for class membership and the within-class regression coefficients for the growth factors in the conditional GMM.

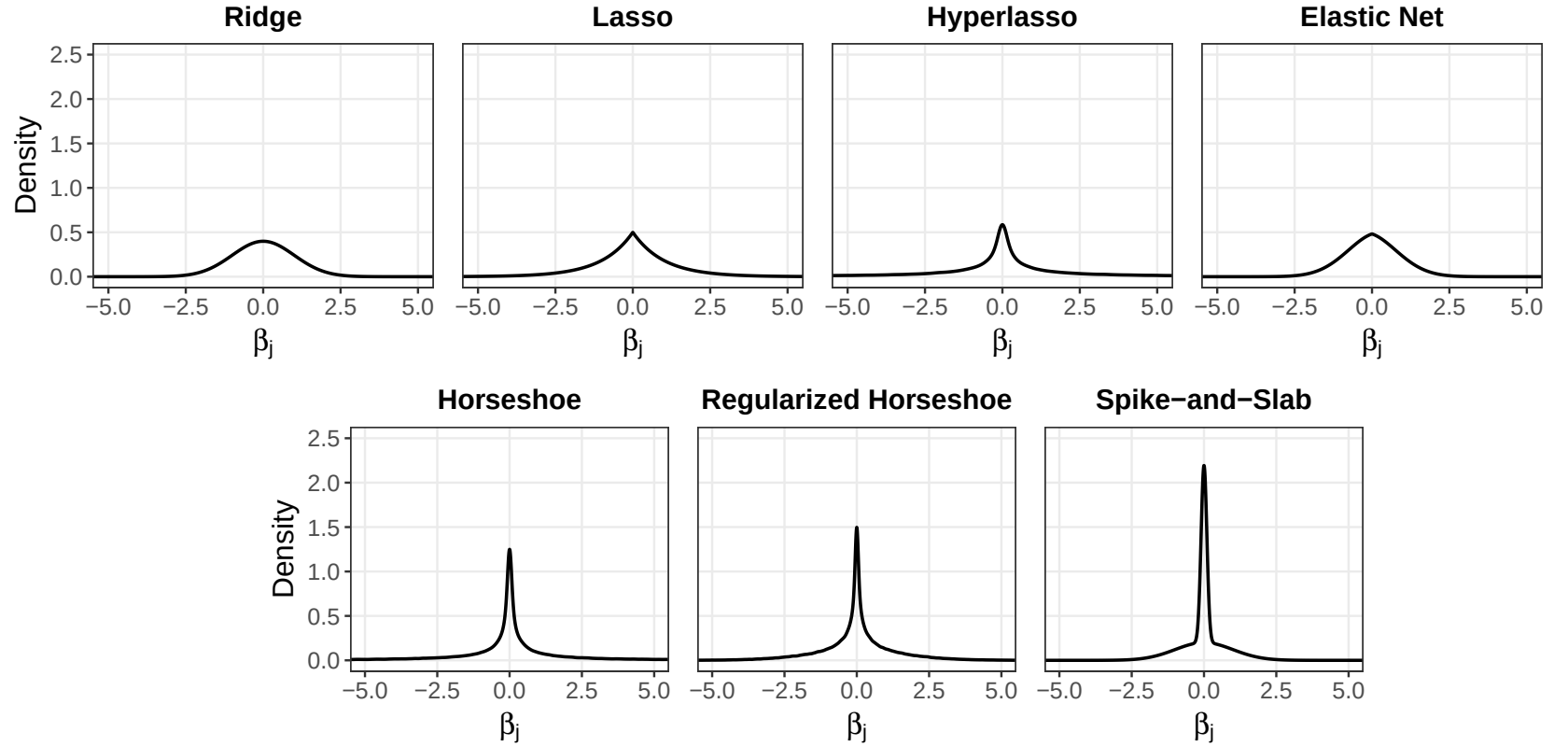
Figure 3.2 provides a visual illustration of the prior density functions for all seven priors. The hyperparameter values underlying the figure are illustrative rather than prescriptive. For priors whose standard formulations involve hyperpriors (horseshoe, regularized horseshoe, and spike-and-slab), I adopt the hyperprior specifications used in van Erp et al. (2019), which match the original proposals (Carvalho et al., 2010; George & McCulloch, 1993; Piironen & Vehtari, 2017). For the remaining priors (ridge, lasso, hyperlasso, and elastic net), I use unit-scale representative values so that the distinguishing shape features of each density are visible on a common scale.

3.4.1 Ridge

The ridge prior is the Bayesian counterpart of the L_2 penalty introduced by Hoerl and Kennard (1970) in the frequentist regularization literature. The Bayesian formulation places a normal distribution centered at zero on each regression coefficient

²Technically speaking, for the GMM formulation, residual variances are class-specific. For notational simplicity, the shrinkage prior specifications are presented using a generic σ^2 . The implementation details for the conditional GMM are described in the simulation design section.

Figure 3.2: Prior densities of the seven focal shrinkage priors, plotted under illustrative hyperparameter settings. The top row shows priors with frequentist origins (ridge, lasso, hyperlasso, and elastic net), and the bottom row shows priors developed within the Bayesian framework (horseshoe, regularized horseshoe, and spike-and-slab). Ridge: $\sigma^2/\lambda = 1$. Lasso: scale = 1. Hyperlasso: $\nu = 1/2$, $\lambda \sim \text{half-Cauchy}(0, 1)$. Elastic net: $\lambda_1 = 1/4$, $\lambda_2 = 1$. Horseshoe: $\tau_j, \lambda \sim \text{half-Cauchy}(0, 1)$. Regularized horseshoe: $\tau_j, \lambda \sim \text{half-Cauchy}(0, 1)$, $c = 2$. Spike-and-Slab: $\omega = 0.5$, $\phi_0^2 = 0.01$, $\tau_j^2 = 1$.



(Hsiang, 1975):

$$\beta_j \mid \sigma^2, \lambda \sim \mathcal{N}\left(0, \frac{\sigma^2}{\lambda}\right), \quad (3.7)$$

where $\lambda > 0$ is a global penalty parameter that controls the degree of shrinkage. Larger values of λ result in smaller prior variances and stronger shrinkage toward zero. The ridge prior applies uniform shrinkage to all coefficients regardless of their magnitude. This behavior arises from two features of the normal form. First, the penalty parameter λ enters the prior variance identically for all coefficients, so the shrinkage factor applied to each depends on λ but not on its magnitude. Second, the light tails of the normal density assign low probability to values far from zero, so the prior pulls small and large coefficients toward zero with similar strength. The prior can therefore overly attenuate substantively meaningful effects (van Erp et al., 2019). Another consequence of the normal form is that the ridge prior does not produce exact zeros; rather, coefficients are continuously shrunk toward zero but never reach it. The ridge prior thus serves primarily as a regularizer that reduces the variance of estimates rather than as a strict variable selector. In other words, the ridge prior cannot be used for variable selection because it never drives coefficients to zero.

3.4.2 Lasso

One limitation of the ridge prior is its inability to shrink coefficients exactly to zero, which means all predictors remain in the model after regularization (Zou & Hastie, 2005). The least absolute shrinkage and selection operator (lasso; Tibshirani, 1996) addresses this limitation through an L_1 penalty that can shrink coefficients to exact zero. The Bayesian counterpart of the lasso, introduced by Park and Casella (2008), replaces the normal prior with a double-exponential (Laplace) distribution:

$$\beta_j \mid \sigma, \lambda \sim \text{Double-Exponential}\left(0, \frac{\sigma}{\lambda}\right). \quad (3.8)$$

Equivalently, the lasso prior can be expressed as a scale mixture of normals with exponential mixing:

$$\beta_j \mid \tau_j^2, \sigma^2 \sim \mathcal{N}(0, \sigma^2 \tau_j^2), \quad \tau_j^2 \mid \lambda^2 \sim \text{Exponential}\left(\frac{\lambda^2}{2}\right). \quad (3.9)$$

The Laplace density has a sharper peak at zero compared to the normal, concentrating more prior mass near zero and thereby promoting sparser solutions. This sharper peak enables posterior estimates that are effectively zero, making the lasso suitable for variable selection. However, the lasso has several known limitations (Polson & Scott, 2011; Polson et al., 2014). Because its global penalty λ applies uniformly to all coefficients, large effects can be overshrunk. The lasso tends to select only one predictor from each group of correlated variables. In addition, when the number of predictors exceeds the sample size, the lasso cannot select more predictors than observations.

3.4.3 Hyperlasso

The hyperlasso prior (Griffin & Brown, 2011) was proposed as a Bayesian alternative to the adaptive lasso that introduces predictor-specific local shrinkage parameters to address the lasso’s tendency to overshrink large effects. Rather than applying a single global penalty uniformly to all coefficients, the hyperlasso places a gamma hyperprior on the local variance of each coefficient:

$$\beta_j \mid \tau_j^2 \sim \text{Double-Exponential}\left(0, \sqrt{2\tau_j^2}\right), \quad \tau_j^2 \mid \lambda \sim \text{Gamma}\left(\frac{1}{2}, \lambda^2\right). \quad (3.10)$$

This hierarchical structure allows the degree of shrinkage to adapt across predictors. Coefficients associated with noise can receive stronger shrinkage, and coefficients reflecting genuine effects can retain their magnitude. The resulting marginal prior on β_j has heavier tails than the lasso, providing greater resistance to overshrinkage of large effects. However, these heavier tails correspond to a non-convex penalty, which implies that the resulting posterior is multimodal (Griffin & Brown, 2011; van Erp et al., 2019). As such, a single MCMC chain can settle into one mode without reaching others. Griffin and Brown (2011) accordingly recommend initializing the sampler from multiple starting values that span different regions of the parameter space, and convergence diagnostics across multiple chains are essential to ensure adequate exploration of the posterior when applying the hyperlasso in practice.

3.4.4 Elastic Net

The elastic net (Zou & Hastie, 2005) combines the L_1 penalty of the lasso with the L_2 penalty of the ridge by introducing two hyperparameters, λ_1 and λ_2 ,³ that control the relative contribution of each component:

$$p(\beta_j \mid \sigma^2, \lambda_1, \lambda_2) \propto \exp\left\{-\frac{1}{2\sigma^2} (\lambda_1|\beta_j| + \lambda_2\beta_j^2)\right\}. \quad (3.11)$$

In this formulation, the $\lambda_1|\beta_j|$ term corresponds to the lasso penalty and the $\lambda_2\beta_j^2$ term corresponds to the ridge penalty. The resulting density is less sharply peaked at zero than the lasso but sharper than the ridge, reflecting its intermediate position between the two (Zou & Hastie, 2005). Through this combination, the elastic net inherits the variable selection capability of the lasso from the λ_1 component and the grouping property of the ridge from the λ_2 component, which encourages correlated predictors to be included or excluded together. Equation (3.11) can equivalently be expressed as a scale mixture of normals:

$$\beta_j \mid \sigma^2, \tau_j^2, \lambda_2 \sim \mathcal{N}\left(0, \frac{\sigma^2}{\lambda_2 + \tau_j^{-2}}\right), \quad \tau_j^2 \mid \lambda_1 \sim \text{Exponential}\left(\frac{\lambda_1^2}{2}\right), \quad (3.12)$$

³The subscripts 1 and 2 here index the two penalty components (L_1 and L_2), not measurement occasions. These parameters are distinct from the factor loadings λ_t defined in the latent GCM specification.

facilitating posterior computation via MCMC. A practical advantage of the Bayesian formulation is that both penalty parameters λ_1 and λ_2 can be estimated simultaneously with the model parameters, avoiding the sequential cross-validation procedures required in the frequentist setting.

3.4.5 Horseshoe

Proposed within the Bayesian framework, the horseshoe prior (Carvalho et al., 2010) belongs to the class of global-local shrinkage priors that separate shrinkage into two components (Polson & Scott, 2011). The first is a global shrinkage parameter λ that controls the overall level of sparsity. The second is predictor-specific local shrinkage parameters τ_j that allow individual coefficients to escape the global pull toward zero. The horseshoe specifies half-Cauchy distributions for both components:

$$\beta_j \mid \tau_j \sim \mathcal{N}(0, \tau_j^2), \quad \tau_j \mid \lambda \sim \text{half-Cauchy}(0, \lambda), \quad \lambda \sim \text{half-Cauchy}(0, \sigma). \quad (3.13)$$

The global parameter λ shrinks all coefficients toward zero, reflecting the prior belief that the true model is sparse. At the same time, the local parameters τ_j allow coefficients with large effects to resist this global shrinkage. The horseshoe derives its name from the horseshoe-shaped density of the implied shrinkage weight $\kappa_j = 1/(1 + \tau_j^2)$, which places substantial mass near both zero and one (Carvalho et al., 2010). This produces a bimodal shrinkage profile in which small coefficients receive κ_j values close to one and are heavily shrunk toward zero, and large coefficients receive κ_j values close to zero and remain unaffected. The heavy tails of the half-Cauchy distribution ensure that genuine signals are preserved even under strong global shrinkage, a property that distinguishes the horseshoe from the ridge and lasso.

3.4.6 Regularized Horseshoe

The regularized horseshoe (Piironen & Vehtari, 2017), also known as the Finnish horseshoe, was proposed to address a specific limitation of the original horseshoe prior from Carvalho et al. (2010). The half-Cauchy distribution has polynomially decaying tails, meaning that extreme values remain probable rather than vanishing rapidly. As a result, the local parameters τ_j can take on extremely large values, so large coefficients can escape shrinkage entirely. When parameters are weakly identified (i.e., the data alone do not clearly distinguish whether a coefficient is zero or nonzero), this behavior can lead to diffuse posteriors and unstable MCMC sampling (Ghosh et al., 2018; Piironen & Vehtari, 2017). The regularized horseshoe addresses this by introducing a finite slab width c^2 that bounds the effective local variance. The

full specification is:

$$\beta_j \sim \mathcal{N}(0, \tilde{\tau}_j^2), \quad \tilde{\tau}_j^2 = \frac{c^2 \tau_j^2 \lambda^2}{c^2 + \tau_j^2 \lambda^2}, \quad (3.14)$$

$$\tau_j \sim \text{half-Cauchy}(0, 1), \quad \lambda \sim \text{half-Cauchy}(0, \tau_0), \quad (3.15)$$

$$c^2 \sim \text{Inverse-Gamma}\left(\frac{\nu}{2}, \frac{\nu s^2}{2}\right), \quad (3.16)$$

where $\tilde{\tau}_j^2$ replaces the unregularized local variance $\tau_j^2 \lambda^2$ in the horseshoe specification.

Intuitively, this regularization caps how large the local variance can grow, keeping large coefficients within a bounded range and preserving the standard horseshoe's heavy shrinkage for small ones. The mechanism can be understood more formally through two limiting cases. When a coefficient is small ($\tau_j^2 \lambda^2 \ll c^2$), the regularized variance $\tilde{\tau}_j^2 \approx \tau_j^2 \lambda^2$, and shrinkage operates as in the standard horseshoe. When a coefficient is large ($\tau_j^2 \lambda^2 \gg c^2$), the regularized variance $\tilde{\tau}_j^2 \approx c^2$, imposing a soft upper bound on the prior variance. In this regime, the prior on large coefficients approximates a Student's t distribution with ν degrees of freedom and scale s^2 , providing controlled regularization rather than unbounded freedom (Piironen & Vehtari, 2017). This regularization stabilizes posterior computation and retains the desirable shrinkage properties of the horseshoe for small coefficients.

A key practical consideration is the calibration of the global shrinkage parameter τ_0 in Equation (3.15). Piironen and Vehtari (2017) proposed setting $\tau_0 = (p_0/(p - p_0)) \cdot (\sigma/\sqrt{n})$, where p_0 is the expected number of nonzero coefficients, p is the total number of predictors, σ is the residual standard deviation, and n is the sample size. This calibration encodes the prior belief about sparsity directly into the global shrinkage level. Although this calibration requires specifying p_0 in advance, the regularized horseshoe has been shown to be robust to the choice of p_0 (Piironen & Vehtari, 2017).

3.4.7 Spike-and-Slab

The spike-and-slab prior (George & McCulloch, 1993; Mitchell & Beauchamp, 1988) takes a fundamentally different approach from all of the continuous shrinkage priors described above. Rather than placing a single continuous distribution on each coefficient, the spike-and-slab models variable selection as a discrete mixture of two components. One is a “spike” distribution concentrated near zero for coefficients that reflect noise, and the other is a “slab” distribution with larger variance for coefficients that represent genuine effects:

$$\beta_j \mid \delta_j \sim (1 - \delta_j) \mathcal{N}(0, \phi_0^2) + \delta_j \mathcal{N}(0, \tau_j^2), \quad \delta_j \sim \text{Bernoulli}(\omega), \quad (3.17)$$

where $\delta_j \in \{0, 1\}$ is a binary inclusion indicator, ϕ_0^2 is a small variance for the spike component, τ_j^2 is the slab variance, and ω is the prior inclusion probability. The working mechanism of the spike-and-slab is straightforward. If the data support a

coefficient close to zero, the posterior favors $\delta_j = 0$, assigning the coefficient to the spike and applying strong shrinkage. If the data support a coefficient that deviates from zero, the posterior favors $\delta_j = 1$, assigning the coefficient to the slab, where it receives minimal shrinkage and is free to take on its data-supported value. The posterior inclusion probability $\Pr(\delta_j = 1 \mid \mathbf{Y})$ provides an interpretable measure of the evidence for including each predictor. However, this discrete selection mechanism of separating active and inactive predictors introduces a categorical latent variable for each coefficient, which increases the computational cost relative to the continuous shrinkage priors.

3.5 Performance of Shrinkage Priors in Bayesian Latent Variable Modeling

A growing body of research has begun to evaluate the performance of shrinkage priors within Bayesian latent variable models. This literature is still in its early stages, but the available evidence offers important insights into how different priors behave when applied to the more complex parameter structures that characterize structural equation models and related frameworks.

One of the early works in the latent variable modeling context was conducted by Jacobucci and Grimm (2018), who provided an overview and demonstration of frequentist and Bayesian regularization in factor analysis and latent GCM. Using empirical examples, they showed that Bayesian estimation with small-variance normal priors produces shrinkage equivalent to frequentist ridge regularization, and that the Bayesian lasso with Laplace priors drives small estimates closer to zero than ridge but does not produce exact zeros under full Bayes estimation. They further extended the demonstration to hierarchical Bayesian methods, including the adaptive lasso, horseshoe, and horseshoe+ priors, the latter being an extension of the horseshoe designed for ultra-sparse settings (Bhadra et al., 2017). These demonstrations established that the Bayesian framework offers a flexible platform for regularized estimation in latent variable models, but also raised the question of whether more adaptive priors might outperform simpler alternatives in separating signal from noise.

This question was taken up by Harra and Kaplan (2024), who compared the ridge, lasso, horseshoe, and regularized horseshoe priors in path analysis and latent GCMs. Through empirical examples and simulation studies that varied sample size, they found that the horseshoe and regularized horseshoe produced substantially greater shrinkage than the ridge and lasso, particularly for small samples. The horseshoe-type priors also demonstrated stronger predictive performance as measured by leave-one-out cross-validation, and simultaneously produced more parsimonious models by shrinking noise coefficients more heavily without overshrinking genuine effects. These results provided the first clear evidence that the adaptive shrinkage properties theorized by Carvalho et al. (2010) translate into practical gains within latent variable models.

In a different latent variable context, Zhang and Liang (2024) evaluated Bayesian regularization for covariate selection in MIMIC models. They compared the ridge, lasso, adaptive lasso, horseshoe, and spike-and-slab priors. Their findings corroborated the advantages of the horseshoe. It provided a favorable balance of shrinkage intensity and variable selection accuracy, with strong performance in relative bias and root mean square error for medium to large effect sizes. The ridge prior showed advantages in convergence stability but was less discriminating between active and inactive predictors. The spike-and-slab prior offered discrete selection but at a higher computational cost. Taken together with the results of Harra and Kaplan (2024), these findings suggest that priors with local adaptivity represent a substantive advance over global shrinkage priors for covariate selection in structural models. Beyond structural regression, Shi et al. (2023) applied posterior inclusion probabilities to mediation models for selecting candidate mediators, further illustrating the versatility of Bayesian variable selection across the latent variable modeling landscape.

Several themes emerge from this body of work. First, priors that combine a global shrinkage parameter with predictor-specific local parameters tend to outperform priors that rely on global shrinkage alone. The horseshoe and regularized horseshoe exemplify this principle, consistently achieving favorable variable selection accuracy across the studies reviewed. Second, the relative performance of the priors is shaped by the interplay between sample size, the number of predictors, and the degree of sparsity, indicating that prior performance can vary depending on the analytic conditions. Third, the consequences of prior choice extend beyond variable selection. Shrinkage priors can influence parameter recovery, predictive accuracy, convergence behavior, and classification accuracy, all of which are critical considerations in the mixture modeling context where individuals must be correctly assigned to latent classes.

I note that, to the best of my knowledge, no study has examined the performance of shrinkage priors for variable selection within latent variable mixture modeling, and specifically within GMMs. Moreover, even within simpler latent variable models, priors such as the hyperlasso and elastic net have received little to no empirical evaluation, making their comparative performance in latent variable contexts an open question. The GMM framework introduces layers of complexity that are absent from the latent variable models examined in the literature to date. The model must simultaneously recover the latent class structure, estimate class-specific regression coefficients linking predictors to both class membership and within-class growth, and maintain adequate classification accuracy across individuals. Whether the favorable properties of shrinkage priors documented in simpler latent variable models carry over to this more demanding setting has not been investigated. This chapter addresses these gaps through a Monte Carlo simulation. It provides the first empirical evaluation of shrinkage priors in GMMs, compares all seven priors introduced above (including the underexamined hyperlasso and elastic net), and tests whether their performance in simpler latent variable models holds up under the added complexity of mixture modeling. Details of the simulation design are presented next.

3.6 Simulation Design

I conducted a Monte Carlo simulation study to stress test the seven shrinkage priors for variable selection in a two-class conditional GMM, examining if and when they perform well. The simulation design was a fully crossed factorial with five factors: sample size ($n = 150, 300, 900$), class separation (Mahalanobis distance of 3.0 and 4.0), class proportion (equal at 50:50 and unequal at 70:30), predictor effect pattern (homogeneous and heterogeneous), and predictor intercorrelation ($\rho = 0.20$ and 0.80). This yields $3 \times 2 \times 2 \times 2 \times 2 = 48$ unique data-generating conditions. Each condition was replicated 500 times, and each replicated data set was analyzed with all seven shrinkage priors under examination, resulting in $48 \times 500 \times 7 = 168,000$ total model fits. A summary of the simulation design factors and their levels is presented in Table 3.1.

Table 3.1: Summary of Simulation Design Factors and Their Levels.

Factor	Levels
Sample size	$n = 150$ (small), 300 (medium), 900 (large)
Class proportion	Equal (50:50), Unequal (70:30)
Class separation	MD = 3.0 (moderate), 4.0 (high) Class 1: $\alpha_1^{(1)} = 48, \alpha_2^{(1)} = 3$ (fixed across conditions) Class 2: $\alpha_1^{(2)} = 39, \alpha_2^{(2)} = 0$ (MD = 3.0) $\alpha_1^{(2)} = 33.613, \alpha_2^{(2)} = 0$ (MD = 4.0)
Predictor effect pattern	Homogeneous, Heterogeneous
Predictor intercorrelation	$\rho = 0.20$ (low), 0.80 (high)

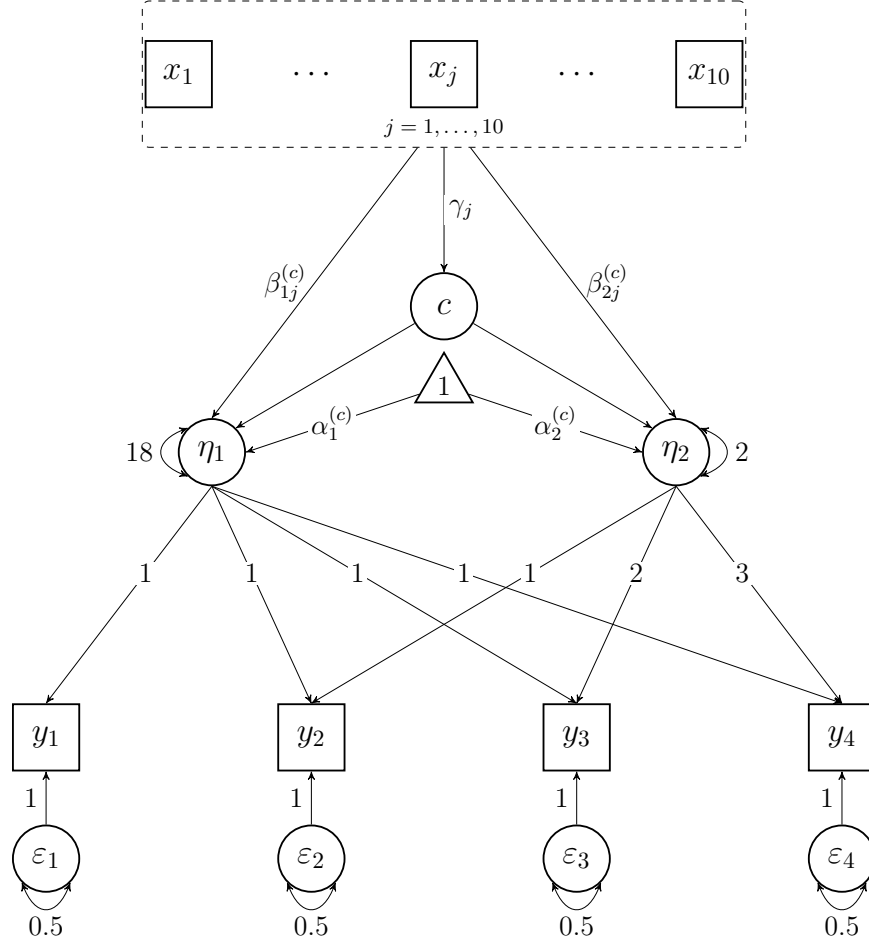
Note. MD = Mahalanobis distance.

3.6.1 Population Model

The population model was designed to reflect a common yet sufficiently comprehensive specification encountered in GMM simulation and applied research (De-paoli, 2013, 2014; Hu et al., 2017; Huang et al., 2010; S. Kim et al., 2022; L. Li & Hser, 2011; Tofighi & Enders, 2007). To this end, the population model was a two-class ($C = 2$) linear conditional GMM with four equally spaced time points ($T = 4$) and two growth factors ($m = 2$): an intercept (η_1) and a linear slope (η_2). The model included $J = 10$ candidate predictors that entered through both latent class membership (via multinomial logistic regression coefficients γ_j) and within-class growth factor regressions (via linear regression coefficients $\beta_{\ell j}^{(c)}$). The choice of $J = 10$

was motivated by previous regularization studies in latent GCMs. Specifically, these studies included 4 (Harra & Kaplan, 2024), 8 (Jacobucci & Grimm, 2018), and 10 (Jacobucci, 2017) predictors, reflecting realistic predictor counts across simulated and applied datasets in growth modeling research. The factor loading matrix was specified as $\mathbf{\Lambda}_y = [\mathbf{1}, (0, 1, 2, 3)^\top]$. A path diagram of the population model is presented in Figure 3.3.

Figure 3.3: Path diagram of the two-class linear conditional growth mixture model used as the population model. The dashed rectangle indicates the ten candidate predictors, with a generic predictor x_j entering through regression coefficients γ_j (latent class membership) and $\beta_{\ell j}^{(c)}$ (within-class growth factors).



The population values for the GMM parameters excluding predictors were determined based on Depaoli (2013, 2014) and Kaplan (2002). For the Class 1 growth factor means, I fixed $\boldsymbol{\alpha}_1 = (48, 3)^\top$, with a mean intercept of 48 and a mean slope of 3. For the Class 2 growth factor means ($\boldsymbol{\alpha}_2$), I varied the values depending on the degree of class separation (described in the Class Separation subsection). For the within-class growth factor covariance matrix and time-specific residual variance, I set $\boldsymbol{\Psi}_c = \text{diag}(18, 2)$ and $\boldsymbol{\Theta}_c = 0.5\mathbf{I}_4$, respectively. Both were held equal across classes

for estimation efficiency.

The ten candidate predictors were generated from a multivariate normal distribution with zero means and unit marginal variances to standardize the predictor scale and facilitate comparison of relative effects. Among these predictors, there was a common pairwise correlation ρ across all predictor pairs (see the Predictor Intercorrelation subsection). In each condition, five of the ten predictors had nonzero effects and the remaining five were noise variables (for the specific designation and the true predictor effects on both class membership and growth factors, see the Predictor Effect Pattern subsection).

3.6.2 Sample Size

Sample size is an important factor in evaluating regularization methods, as smaller samples tend to produce more visible differences across priors and greater sensitivity to prior specification (Harra & Kaplan, 2024; van Erp et al., 2019; Zhang & Liang, 2024). To cover the range of sample sizes typical in GMM research (Depaoli, 2013, 2014; Depaoli, Heo, et al., 2026; S. Kim et al., 2022), I considered three levels of total sample size: $n = 150$ (small), $n = 300$ (medium), and $n = 900$ (large).

3.6.3 Class Proportion

In mixture modeling, the relative sizes of latent classes are known to influence model performance (Tueller & Lubke, 2010). To mirror commonly considered equal and unequal class proportions in GMM research (Depaoli, 2013; Depaoli, Heo, et al., 2026; L. Li & Hser, 2011; M. Li & Harring, 2017), I examined equal ($\pi_1 = \pi_2 = .50$) and unequal ($\pi_1 = .70$, $\pi_2 = .30$) conditions.

3.6.4 Class Separation

Class separation, the degree to which latent classes are similar or dissimilar from one another, affects the performance of GMMs and is important to examine in the context of variable selection. As a measure of class separation, I used the multivariate Mahalanobis distance (MD)⁴:

$$\text{MD} = [(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)^\top \boldsymbol{\Psi}^{-1}(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)]^{1/2}. \quad (3.18)$$

Based on simulation research with GMMs with and without covariates (Depaoli, Heo, et al., 2026; S. Kim et al., 2021, 2022; L. Li & Hser, 2011; M. Li & Harring, 2017), I

⁴The MD in Equation (3.18) follows the conventional definition adopted across the GMM simulation literature, including studies with and without predictors (Depaoli, Heo, et al., 2026; S. Kim et al., 2021, 2022; L. Li & Hser, 2011; M. Li & Harring, 2017), which characterizes class separation through differences in growth factor means and the within-class covariance matrix. This definition does not incorporate predictor effects, and MD reflects intercept-based separation rather than the conditional separation that may arise when predictors have class-heterogeneous effects. I retained the conventional definition for comparability with prior simulation studies.

examined MD = 3.0 (moderate) and MD = 4.0 (high) as two representative levels of class separation. Given the fixed growth factor means for Class 1 at $\alpha_1 = (48, 3)^\top$, the Class 2 means were adjusted to achieve each target MD: $\alpha_2 = (39, 0)^\top$ for MD = 3.0 and $\alpha_2 = (33.613, 0)^\top$ for MD = 4.0. These values were computed using the common covariance matrix $\Psi = \text{diag}(18, 2)$ described in the Population Model subsection.

3.6.5 Predictor Effect Pattern

In the conditional GMM, predictors enter through two pathways: predicting latent class membership and predicting within-class growth factors. For class membership prediction, the coefficients were held constant across all conditions. For the growth factor regressions, I varied whether the active predictor set was the same across classes (homogeneous) or different across classes (heterogeneous). In all conditions, five of the ten predictors had nonzero effects (1 large, 2 medium, and 2 small) and the remaining five were null (zero effects).

Because the population model has two classes, the multinomial logistic regression for class membership reduces to a binomial logistic regression with a single set of coefficients. The true logistic regression coefficients for predicting class membership were:

$$(\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5) = (1.386, 0.916, 0.916, 0.405, 0.405), \quad (\gamma_6, \dots, \gamma_{10}) = (0, \dots, 0). \quad (3.19)$$

These values correspond to odds ratios of 4.0, 2.5, 2.5, 1.5, and 1.5 for γ_1 through γ_5 , respectively, spanning a range of effect sizes from small to large based on conventions for logistic regression (Cohen, 1992; Rosenthal, 1996).

For the growth factor regressions, the nonzero coefficients were set to 0.5, 0.3, and 0.1, reflecting relatively large, medium, and small effects, respectively. These coefficients were applied identically to both the intercept ($\beta_{1j}^{(c)}$) and slope ($\beta_{2j}^{(c)}$) regressions, although they carry different interpretations because the two growth factors are on different metrics. The two patterns are described below.

Homogeneous Pattern

In the homogeneous pattern, the first five predictors (x_1 through x_5) had identical nonzero effects on both the latent intercept and the latent slope across both classes, and the remaining five predictors (x_6 through x_{10}) had zero effects in both classes:

$$(\beta_{1,1}^{(c)}, \dots, \beta_{1,10}^{(c)}) = (\beta_{2,1}^{(c)}, \dots, \beta_{2,10}^{(c)}) = (0.5, 0.3, 0.3, 0.1, 0.1, 0, 0, 0, 0, 0), \quad c = 1, 2. \quad (3.20)$$

Heterogeneous Pattern

In the heterogeneous pattern, the true predictors of the latent growth factors differed between classes. In Class 1, the first five predictors (x_1 through x_5) had

nonzero effects with the same magnitudes as in the homogeneous pattern, and the remaining five (x_6 through x_{10}) had zero effects. In Class 2, this assignment was reversed such that x_6 through x_{10} had nonzero effects with the same magnitudes, and x_1 through x_5 had zero effects:

$$(\beta_{1,1}^{(1)}, \dots, \beta_{1,10}^{(1)}) = (\beta_{2,1}^{(1)}, \dots, \beta_{2,10}^{(1)}) = (0.5, 0.3, 0.3, 0.1, 0.1, 0, 0, 0, 0, 0), \quad (3.21)$$

$$(\beta_{1,1}^{(2)}, \dots, \beta_{1,10}^{(2)}) = (\beta_{2,1}^{(2)}, \dots, \beta_{2,10}^{(2)}) = (0, 0, 0, 0, 0, 0.5, 0.3, 0.3, 0.1, 0.1). \quad (3.22)$$

The heterogeneous pattern represents a scenario in which the active predictor set for one class is distinct from that for the other. Such a setting requires the shrinkage priors to simultaneously identify class-specific sparse solutions.

3.6.6 Predictor Intercorrelation

The degree of correlation among predictors can influence the performance of shrinkage priors. In GMM research with covariates, low to moderate correlations ranging from 0.2 to 0.5 have been considered (Diallo et al., 2017a; M. Li & Harring, 2017), whereas the broader regularization literature has examined higher correlations such as 0.8 or 0.95 (Bainter et al., 2023; Jacobucci et al., 2019). I set a common pairwise correlation across all predictor pairs, as was the case in Diallo et al. (2017a) and M. Li and Harring (2017). In particular, I considered two levels: low ($\rho = 0.20$) and high ($\rho = 0.80$). The high-correlation condition was designed to examine the performance of shrinkage priors under multicollinearity, which is known to inflate estimation variance and complicate variable selection.

3.6.7 Shrinkage Prior Specification

The prior specifications followed the full Bayes (hierarchical) approach (van Erp et al., 2019; Wolpert & Strauss, 1996), in which the penalty parameter λ is treated as an unknown model parameter with its own hyperprior. This approach is preferred because it enables the uncertainty in λ to be propagated through the posterior distribution and allows the penalty parameter to be estimated simultaneously with all other model parameters within an MCMC sampling step. For all priors, a half-Cauchy(0, 1) hyperprior was placed on the global penalty parameter λ , following van Erp et al. (2019) and Polson and Scott (2012). Prior-specific hyperparameter settings are summarized in Table 3.2.

The same shrinkage prior was applied to all three sets of regression coefficients within a given model (the logistic regression coefficients γ_j for class membership and the within-class regression coefficients $\beta_{1j}^{(c)}$ and $\beta_{2j}^{(c)}$ for the intercept and slope growth factors, respectively). For the regularized horseshoe, the expected number of nonzero coefficients was set to $p_0 = 5$ based on the true sparsity level (see also Harra & Kaplan, 2024), following the calibration approach of Piironen and Vehtari (2017).

Noninformative priors were specified for the remaining model parameters. Specifically, growth factor means received $\mathcal{N}(0, 100)$ priors, the within-class growth

Table 3.2: Shrinkage Prior Specifications and Hyperparameter Settings.

Prior	Specification
Ridge	$\beta_j \mid \sigma^2, \lambda \sim \mathcal{N}(0, \sigma^2/\lambda)$ $\lambda \sim \text{half-Cauchy}(0, 1)$
Lasso	$\beta_j \mid \sigma, \lambda \sim \text{DE}(0, \sigma/\lambda)$ $\lambda \sim \text{half-Cauchy}(0, 1)$
Hyperlasso	$\beta_j \mid \tau_j \sim \text{DE}(0, \sqrt{2\tau_j^2})$ $\tau_j^2 \mid \lambda \sim \text{Gamma}(1/2, \lambda^2)$ $\lambda \sim \text{half-Cauchy}(0, 1)$
Elastic net	$\beta_j \mid \sigma^2, \tau_j^2, \lambda_2 \sim \mathcal{N}(0, \sigma^2/(\lambda_2 + \tau_j^{-2}))$ $\tau_j^2 \mid \lambda_1 \sim \text{Exp}(\lambda_1^2/2)$ $\lambda_1, \lambda_2 \sim \text{half-Cauchy}(0, 1)$
Horseshoe	$\beta_j \mid \tau_j \sim \mathcal{N}(0, \tau_j^2)$ $\tau_j \sim \text{half-Cauchy}(0, 1)$ $\lambda \sim \text{half-Cauchy}(0, \sigma)$
Regularized horseshoe	$\beta_j \sim \mathcal{N}(0, \tilde{\tau}_j^2)$ $\tilde{\tau}_j^2 = c^2 \tau_j^2 \lambda^2 / (c^2 + \tau_j^2 \lambda^2)$ $\tau_j \sim \text{half-Cauchy}(0, 1)$ $\lambda \sim \text{half-Cauchy}(0, \tau_0)$ $c^2 \sim \text{IG}(\nu/2, \nu s^2/2)$ with $\nu = 4, s^2 = 2$
Spike-and-slab	$\beta_j \mid \delta_j \sim (1 - \delta_j)\mathcal{N}(0, \phi_0^2) + \delta_j\mathcal{N}(0, \tau_j^2)$ $\delta_j \sim \text{Bernoulli}(0.5)$ $\phi_0^2 = 0.001$ $\tau_j^2 \sim \text{IG}(0.5, 0.5)$

Note. DE = double-exponential (Laplace); Exp = exponential; IG = inverse-gamma. The notation β_j is used generically to refer to the regression coefficients on which shrinkage priors are placed; in the conditional GMM, these priors are applied identically to the class membership coefficients (γ_j) and the within-class growth factor coefficients ($\beta_{\ell_j}^{(c)}$).

factor precision matrices received Wishart priors with an identity scale matrix and 3 degrees of freedom, and time-specific residual precisions received $\text{Gamma}(1, 0.5)$ priors. The logistic regression intercept γ_0 received a $\mathcal{N}(0, 100)$ prior.

3.6.8 Analytic Procedure

Data were generated using *Mplus* version 8.7 (L. K. Muthén & Muthén, 2017). Bayesian estimation was carried out in **JAGS** (Plummer, 2003)⁵ via the **runjags** package (Denwood, 2016) in R (R Core Team, 2025). To mitigate convergence to local solutions, the true parameter values were used as starting values for all MCMC chains (see also Depaoli, Heo, et al., 2026). For each replication, a single MCMC chain was run for 9,000 iterations. The first 1,000 iterations served as adaptation to allow the **JAGS** sampler to optimize sampling efficiency, and the subsequent 4,000 iterations were discarded as burn-in. No thinning was applied.

Convergence was assessed by splitting the single chain in half and computing the Gelman-Rubin $\hat{R}$ statistic (Gelman & Rubin, 1992). A replication was deemed to have converged if $\hat{R} < 1.1$ for all monitored parameters. To prevent within-chain label switching, an identifiability constraint $\alpha_1^{(2)} < \alpha_1^{(1)}$ was imposed on the class-specific intercept means, whereas between-chain label switching was prevented by using a single chain (Cassiday et al., 2021; Depaoli, 2021).

3.6.9 Outcome Measures

I evaluated performance across five domains: convergence rate, estimation accuracy, variable selection accuracy, classification accuracy, and predictive performance. Although my primary interest lies in variable selection accuracy, the choice of shrinkage prior has consequences beyond variable selection, influencing estimation accuracy and predictive performance (Bainter et al., 2023; Harra & Kaplan, 2024; van Erp et al., 2019; Zhang & Liang, 2024) and, for GMMs particularly, classification accuracy as well. All outcome measures were computed for each condition across 500 replications, with all measures except convergence rate based on converged replications only.

Convergence Rate

The convergence rate was the proportion of replications that achieved converged solutions, as defined by the $\hat{R} < 1.1$ criterion described above.

Estimation Accuracy

Estimation accuracy was assessed using relative bias and root mean square error (RMSE) for the regression coefficients. Relative bias for a parameter θ is defined

⁵ **JAGS** parameterizes the normal distribution in terms of precision (the inverse of variance) rather than variance. All prior specifications in Table 3.2 were reparameterized accordingly in the **JAGS** implementation.

as

$$\text{Relative Bias} = \begin{cases} \frac{1}{R} \sum_{r=1}^R \frac{\hat{\theta}_r - \theta}{|\theta|} & \text{if } \theta \neq 0, \\ \frac{1}{R} \sum_{r=1}^R (\hat{\theta}_r - \theta) & \text{otherwise,} \end{cases} \quad (3.23)$$

where $\hat{\theta}_r$ denotes the posterior median obtained in replication r , θ is the corresponding population value, and R is the number of converged replications. The first branch scales the mean deviation by the absolute true value to express bias as a proportion. The second branch applies when $\theta = 0$, in which case the ratio is undefined and the mean deviation alone is reported. I adopted $|\text{Relative Bias}| \leq 0.1$ as an acceptable bound. For nonzero effects, negative relative bias values indicate shrinkage toward zero, with the magnitude expressed as a proportion of the true value. Because this normalization places bias on a common scale, relative bias facilitates comparison across coefficients with different true values. I computed relative bias from raw posterior median estimates without applying the selection threshold, so that the measure reflects the shrinkage imparted by each prior independently of the subsequent selection decision. To capture the combined effect of bias and sampling variability, I additionally computed the RMSE:

$$\text{RMSE} = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\theta}_r - \theta)^2}. \quad (3.24)$$

RMSE is informative because it reflects the bias-variance tradeoff. Specifically, a prior that induces moderate shrinkage may still yield lower RMSE if the accompanying reduction in estimation variance is sufficiently large.

Variable Selection Accuracy

Variable selection accuracy was evaluated for each of the three coefficient sets (class membership coefficients γ_j , intercept regression coefficients $\beta_{1j}^{(c)}$, and slope regression coefficients $\beta_{2j}^{(c)}$) using the proportion of predictors selected, stratified by true effect size (large, medium, small, and null). For nonzero coefficients, the proportion selected reflects the power to detect true effects, and for null coefficients, it reflects the false selection rate.

The primary criterion was based on the posterior median, where a predictor was deemed selected if $|\hat{\theta}_r| > 0.1$ (Feng et al., 2017; van Erp et al., 2019; Zhang & Liang, 2024)⁶. In regularization, however, the final selection can be sensitive to the choice of criterion. Accordingly, I evaluated variable selection under two additional criteria as sensitivity checks. The first applied the same threshold to the posterior

⁶Because shrinkage priors pull estimates toward zero, coefficients with a population value of 0.1 are expected to fall below this threshold. The resulting amount of shrinkage is captured by the bias measures.

mean, and the second used the 95% equal-tailed credible interval excluding zero ($0 \notin [\theta_{r,0.025}, \theta_{r,0.975}]$), where $\theta_{r,0.025}$ and $\theta_{r,0.975}$ denote the lower and upper bounds of the 95% equal-tailed credible interval from replication r (van Erp et al., 2019; Zhang & Liang, 2024).

Classification Accuracy

I quantified classification accuracy by the agreement between the true class labels and the estimated posterior modal class assignments. For each individual, the modal class was determined as the most frequently assigned class across the MCMC posterior draws. Agreement between the true and estimated partitions was measured using the adjusted Rand index (ARI; Hubert & Arabie, 1985), which adjusts the Rand index (Rand, 1971) for the level of agreement expected by chance alone (Milligan & Cooper, 1986; Steinley, 2004). The ARI has been used to evaluate classification accuracy in GMM studies (McNeish, Harring, & Bauer, 2023). For two partitions of n individuals, the ARI is defined as

$$\text{ARI} = \frac{\sum_{ij} \binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}{\frac{1}{2} \left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} \right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}, \quad (3.25)$$

where n_{ij} denotes the number of individuals assigned to class i in the true partition and class j in the estimated partition, $a_i = \sum_j n_{ij}$ and $b_j = \sum_i n_{ij}$ are the corresponding row and column marginals, and n is the total sample size. The ARI equals 1 for perfect agreement and has an expected value of 0 under random partitioning. Higher values indicate stronger classification agreement, with values above 0.80 serving as a common heuristic for good recovery (Steinley, 2004). The ARI was computed using the `mclust` package in R (Scrucca et al., 2023).

Predictive Performance

Predictive performance refers to the ability of a fitted model to accurately predict new or held-out observations. It has been used to evaluate shrinkage priors in latent GCMs (Harra & Kaplan, 2024). In light of this, and considering other predictive performance measures evaluated in GMMs (Depaoli, Heo, et al., 2026), I assessed predictive performance using two information criteria, both of which are scaled on the deviance metric ($-2 \times \log$ -likelihood) so that lower values indicate better predictive performance.

The first is the Watanabe-Akaike information criterion (WAIC; Watanabe, 2010), defined as

$$\text{WAIC} = -2 \text{lppd} + 2 p_{\text{WAIC}}, \quad (3.26)$$

where $\text{lppd} = \sum_{i=1}^n \log \left(S^{-1} \sum_{s=1}^S p(\mathbf{y}_i | \boldsymbol{\pi}^s, \boldsymbol{\theta}^s) \right)$ is the log pointwise predictive density averaged over S posterior draws, and p_{WAIC} is a penalty for model complexity based on the posterior variance of the log-likelihood across observations (Gelman, Hwang, & Vehtari, 2014). I computed WAIC using the `loo` package in R (Vehtari et al., 2017).

The second is the leave-one-out information criterion (LOOIC; Vehtari et al., 2017), which approximates leave-one-out cross-validation without refitting the model:

$$\text{LOOIC} = -2 \sum_{i=1}^n \log p(\mathbf{y}_i | \mathbf{y}_{-i}), \quad (3.27)$$

where $p(\mathbf{y}_i | \mathbf{y}_{-i})$ is the predictive density for observation i given all other observations. To decrease computational burden, LOOIC can be estimated via Pareto-smoothed importance sampling (PSIS; Vehtari et al., 2017) as implemented in the `loo` package in R (Vehtari et al., 2017).

3.7 Simulation Results

Results are organized into five subsections corresponding to the outcome measures. I first report convergence rates, followed by estimation accuracy, variable selection accuracy, classification accuracy, and predictive performance. Unless otherwise noted, all outcome measures other than convergence rate were computed using converged replications only. Given the comprehensive scope of the simulation, I present results in a focused manner that highlights the most informative patterns. The full set of results is available in the Supplementary Materials at <https://osf.io/8pbd5>.

3.7.1 Convergence Rate

The overall mean convergence rate across all conditions was 75.2%, with a median of 83.0%. Table 3.3 presents the convergence rates across shrinkage priors, sample sizes, and class separation levels. Because convergence rates varied minimally across levels of class proportion, predictor effect pattern, and predictor intercorrelation, these factors were collapsed in Table 3.3 for visual clarity.⁷ The full convergence results across all conditions are provided in the Supplementary Materials.

At $n = 150$ with $\text{MD} = 3.0$, convergence rates ranged from 29.3% for the spike-and-slab to 66.8% for the lasso. The ridge (65.6%) and lasso (66.8%) achieved the highest rates, followed by the horseshoe (62.2%), the hyperlasso (55.8%), and the elastic net (54.7%). The regularized horseshoe (41.7%) and the spike-and-slab (29.3%) exhibited the lowest rates. When class separation increased to $\text{MD} = 4.0$,

⁷One modest departure involved the regularized horseshoe and spike-and-slab, for which convergence rates dropped substantially under high predictor intercorrelation ($\rho = 0.8$), with differences roughly ranging from 10 to 45 percentage points relative to $\rho = 0.2$. The remaining five priors showed differences of less than 4 percentage points between the two correlation levels.

Table 3.3: Convergence Rates (%) Across Shrinkage Priors, Sample Sizes, and Class Separation Levels.

Prior	$n = 150$		$n = 300$		$n = 900$	
	MD = 3.0	MD = 4.0	MD = 3.0	MD = 4.0	MD = 3.0	MD = 4.0
Ridge	65.6	96.3	90.4	92.7	87.8	87.5
Lasso	66.8	97.2	91.2	93.2	88.1	88.2
Hyperlasso	55.8	91.7	86.9	91.5	87.8	88.2
Elastic net	54.7	79.4	75.5	81.9	80.2	81.3
Horseshoe	62.2	93.6	87.3	92.7	86.8	88.3
Reg. horseshoe	41.7	55.3	53.0	57.0	59.9	63.4
Spike-and-slab	29.3	54.9	48.1	59.9	58.6	64.6

Note. Reg. horseshoe = regularized horseshoe.

convergence improved markedly at this sample size. The ridge, lasso, hyperlasso, and horseshoe exceeded 90%. The elastic net reached 79.4%. The regularized horseshoe and spike-and-slab improved but remained near 55%.

For larger sample sizes ($n = 300$ and 900), convergence improved across all priors relative to $n = 150$. For the ridge, lasso, hyperlasso, elastic net, and horseshoe, convergence rates ranged from 75.5% to 93.2% across conditions, with higher rates generally observed at greater class separation. Notably, convergence rates at $n = 900$ were slightly lower than those at $n = 300$ for these five priors. The regularized horseshoe and spike-and-slab remained between 48.1% and 64.6%, lagging behind the other five priors.

3.7.2 Estimation Accuracy

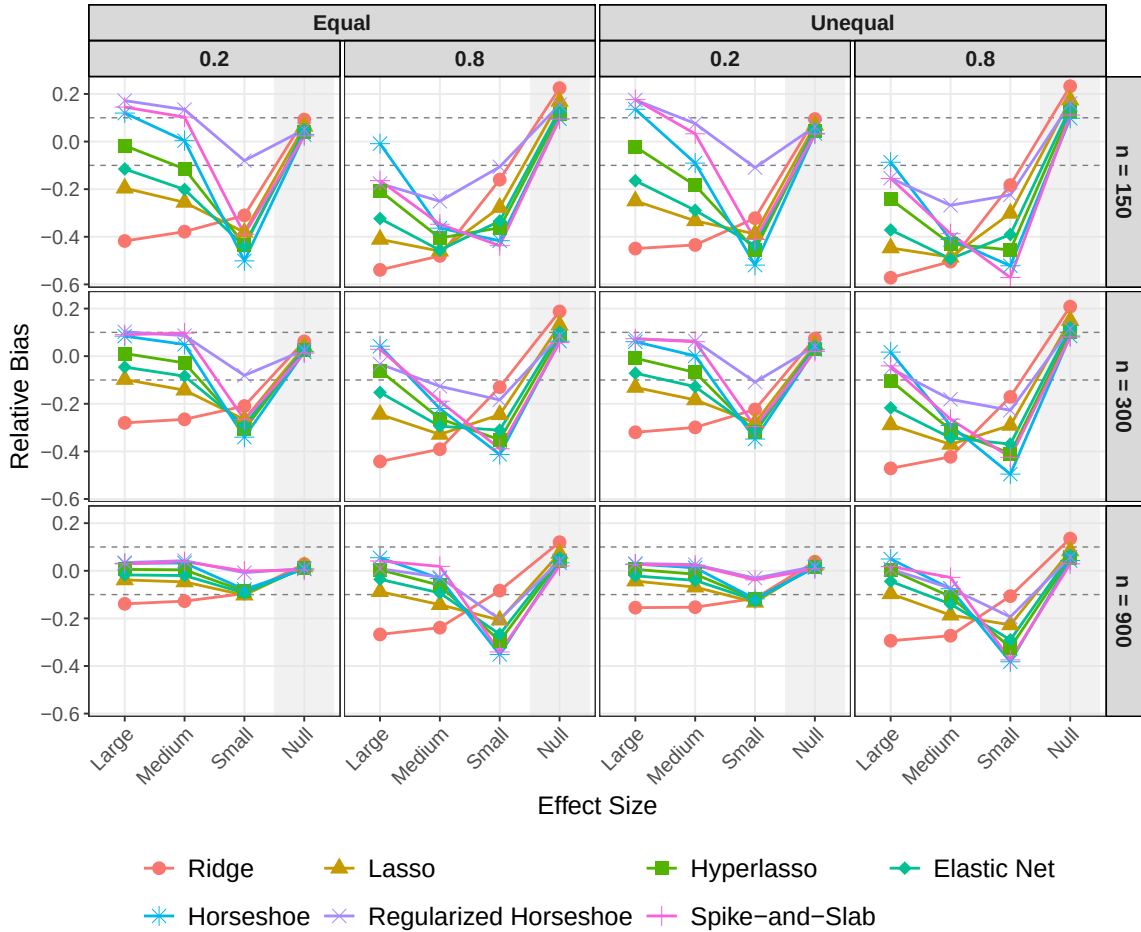
I assessed estimation accuracy using relative bias and RMSE. Relative bias is presented for the class membership coefficients, followed by the growth factor regression coefficients, and then the RMSE results. Results are presented under moderate separation (MD = 3.0), with the corresponding results under high separation (MD = 4.0) provided in the Supplementary Materials. In all relative bias figures, the dashed horizontal lines at ± 0.1 mark the acceptability threshold. For nonzero effects, negative values indicate shrinkage toward zero, expressed as a proportion of the true value. The shaded region highlights the null effect coefficients, for which raw bias is reported.

Relative Bias in Class Membership Coefficients

Figure 3.4 presents the relative bias for the logistic regression coefficients predicting class membership. Each row corresponds to a sample size ($n = 150, 300$, and 900), and the columns are organized by class proportion (Equal and Unequal), each subdivided by predictor intercorrelation ($\rho = 0.2$ and 0.8). Within each panel, the x -axis represents the effect size category (large, medium, small, and null), the y -axis

represents the relative bias, and each line with a different color and shape represents a different shrinkage prior.

Figure 3.4: Relative Bias for Class Membership Coefficients.



intensified shrinkage under high intercorrelation. For small effects, the pattern varied across priors. The spike-and-slab and regularized horseshoe became slightly more negative, and the remaining priors became less negative. For null effects, all priors showed increased positive bias, shifting upward. Under unequal class proportions at $n = 150$, the overall pattern closely resembled the equal proportion condition, and the ordering among priors was consistent with the equal proportion results.

As sample size increased to $n = 300$ and $n = 900$, relative bias generally decreased in magnitude across all priors and effect sizes. Still, the ridge exhibited the strongest negative relative bias for large effects, shrinking estimates toward zero. For null effects, the ridge showed the largest positive bias, and this tendency was more pronounced under high intercorrelation regardless of class proportion.

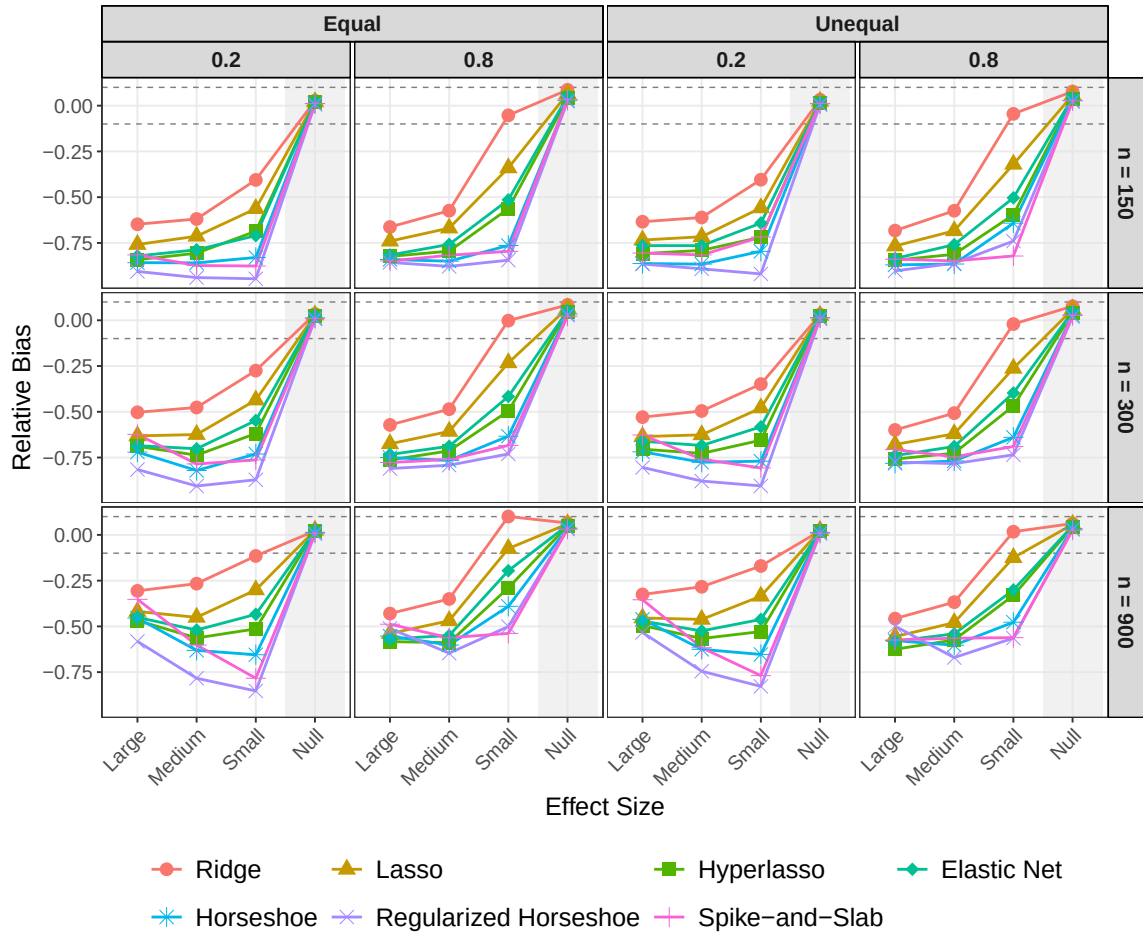
Relative Bias in Growth Factor Regression Coefficients

Figures 3.5 and 3.6 display the relative bias for the intercept and slope growth factor regression coefficients, respectively, under the homogeneous predictor effect pattern. The figures follow the same layout as Figure 3.4. Because both classes share the same true coefficient values under the homogeneous pattern, results are aggregated across classes. Details of class-specific results and those under the heterogeneous pattern are provided in the Supplementary Materials.

In Figure 3.5, beginning with $n = 150$ under equal class proportions with $\rho = 0.2$, all priors produced substantially negative relative bias for the intercept coefficients across all nonzero effect sizes, indicating that estimates were heavily pulled toward zero. The ridge showed the least negative relative bias, and the regularized horseshoe the most. When predictor intercorrelation increased to $\rho = 0.8$, the patterns were largely similar. One notable exception was the ridge, whose relative bias for the small intercept coefficients was near zero, departing from the substantial negative bias observed at $\rho = 0.2$. Under unequal class proportions, the patterns closely resembled those under equal proportions. As sample size increased to $n = 300$ and $n = 900$, negative relative bias decreased in magnitude. However, at $n = 900$ with $\rho = 0.2$, the regularized horseshoe and spike-and-slab produced the largest negative relative bias for small effects, reflecting their shrinkage of small coefficients.

The slope coefficients (Figure 3.6) showed less negative relative bias than the intercept coefficients across all conditions, as reflected in the narrower y -axis range. The most notable difference from the intercept results emerged at larger sample sizes. Beginning at $n = 300$, the adaptive shrinkage behavior of the horseshoe, regularized horseshoe, and spike-and-slab became visible. These priors produced substantial negative relative bias for small effects while preserving large effects with relatively little bias. In contrast, the remaining priors (ridge, lasso, hyperlasso, elastic net) showed less differentiated shrinkage across effect sizes, with negative relative bias for all nonzero effects. This pattern became more pronounced at $n = 900$. Under equal proportions with $\rho = 0.2$, the spike-and-slab and regularized horseshoe produced near-zero relative bias for the large slope effect, falling within the ± 0.1 acceptability band, while simultaneously showing the largest negative relative bias for small

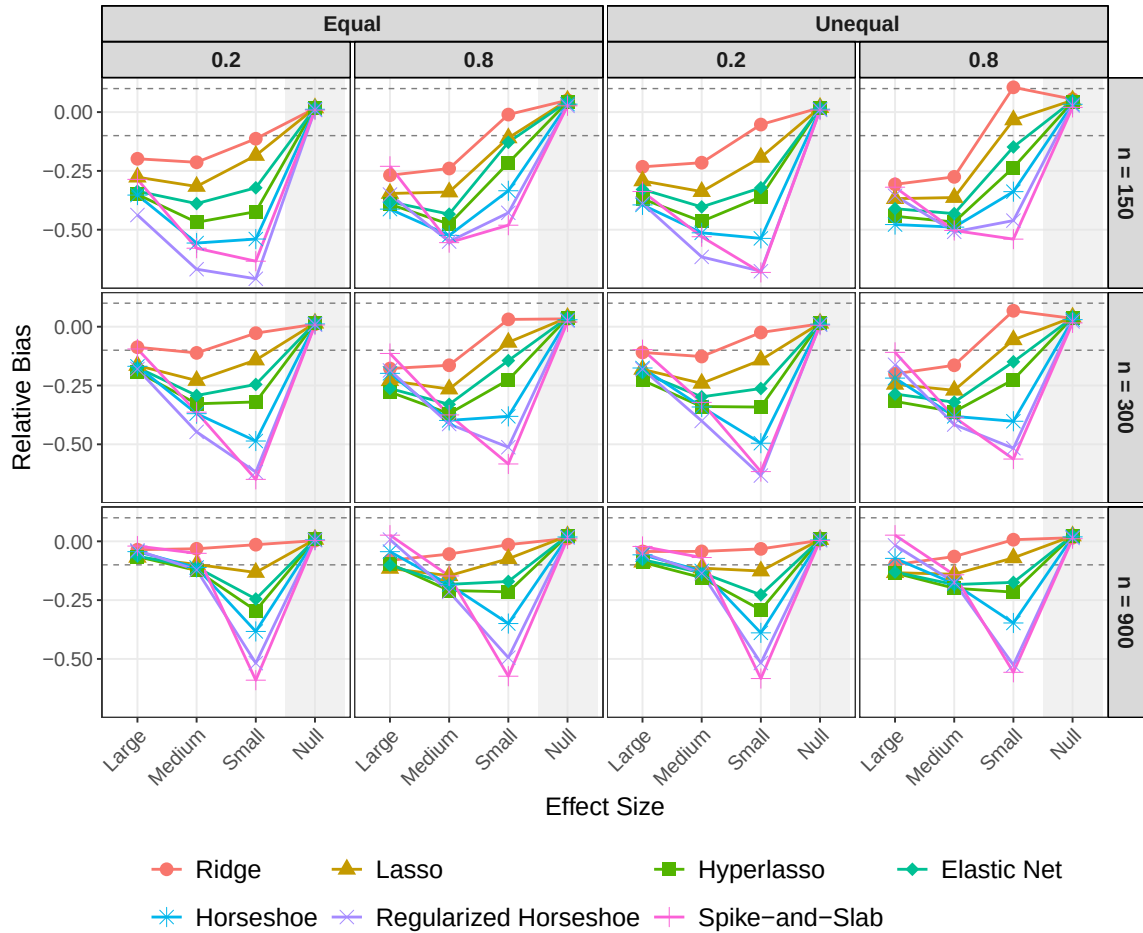
Figure 3.5: Relative Bias for Intercept Growth Factor Regression Coefficients Under the Homogeneous Pattern.



effects. Under equal proportions with $\rho = 0.8$ at $n = 900$, shrinkage generally intensified across effect sizes, though the spike-and-slab and regularized horseshoe produced slight positive relative bias for the large slope effect. Under unequal proportions, these patterns were similar. Overall, the regularized horseshoe and spike-and-slab showed the strongest adaptive shrinkage for the slope coefficients, most pronounced at the largest sample size.

The heterogeneous predictor effect pattern produced results closely matching the homogeneous pattern for most priors (see Supplementary Materials for the full results). One exception involved the horseshoe, regularized horseshoe, and spike-and-slab at $n = 900$, as shown in Figures 3.7 and 3.8. Each row corresponds to a predictor effect pattern (homogeneous and heterogeneous), and the columns are organized by class proportion, each subdivided by predictor intercorrelation. As shown, these three priors showed greater negative relative bias for large effects under the heterogeneous pattern than the homogeneous pattern, with the difference most pronounced under $\rho = 0.8$. This suggests that these priors applied stronger shrinkage when the active

Figure 3.6: Relative Bias for Slope Growth Factor Regression Coefficients Under the Homogeneous Pattern.



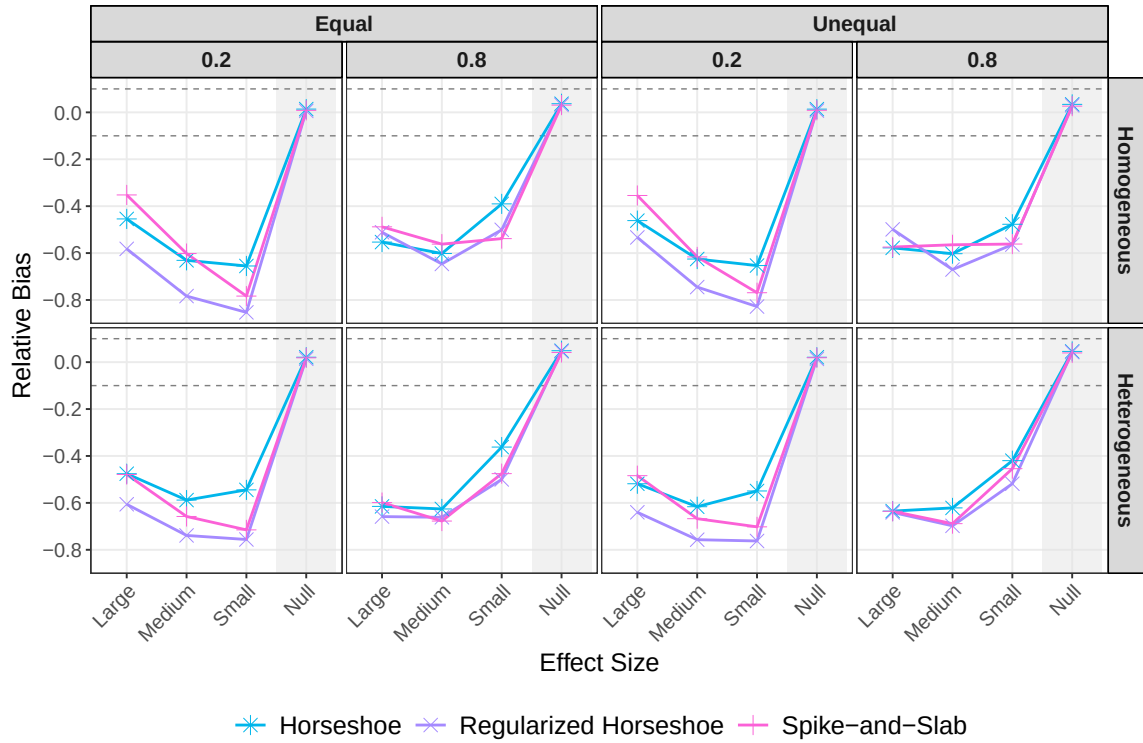
predictor sets differed across classes.

Root Mean Square Error

I report the RMSE for the class membership coefficients in Figure 3.9, which follows the same layout as Figure 3.4.

For $n = 150$ under equal proportions with $\rho = 0.2$, RMSE was highest for large effects and decreased toward null effects across all priors. The ridge and horseshoe produced the highest RMSE for large effects. For the ridge, this was driven by its strong negative bias, and for the horseshoe, by high estimation variability given that its bias for large effects was comparatively small. For null effects, the regularized horseshoe produced the highest RMSE, followed by the ridge. Under $\rho = 0.8$ at $n = 150$, RMSE increased across all priors, with the horseshoe producing the highest RMSE for nonzero effects. Under unequal proportions, the patterns closely resembled those under equal proportions. As sample size increased to $n = 300$ and $n = 900$,

Figure 3.7: Relative Bias Comparison Between Homogeneous and Heterogeneous Patterns at $n = 900$: Intercept Coefficients.



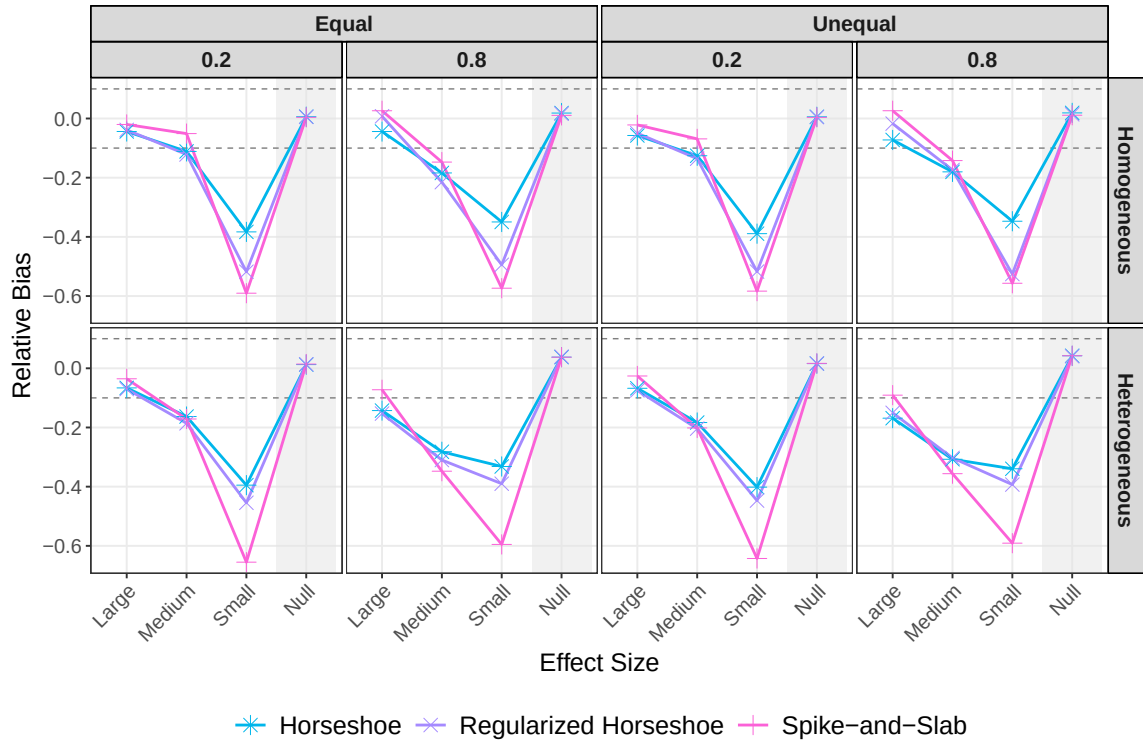
RMSE decreased across all priors and effect sizes.

Figures 3.10 and 3.11 present the RMSE for the intercept and slope growth factor regression coefficients, respectively, under the homogeneous pattern. The figures follow the same layout as Figures 3.5 and 3.6. Results under the heterogeneous pattern are provided in the Supplementary Materials.

In Figure 3.10, at $n = 150$ under equal proportions with $\rho = 0.2$, the ridge produced the lowest RMSE for large effects because its milder shrinkage preserved more of the true signal. The regularized horseshoe and spike-and-slab incurred the highest RMSE for large effects due to stronger shrinkage. For null effects, the ordering reversed. The ridge produced the highest RMSE, and the horseshoe, regularized horseshoe, and spike-and-slab achieved lower RMSE by producing estimates concentrated near the true value of zero. Under $\rho = 0.8$, RMSE increased across all priors. Under unequal proportions, the patterns were similar. As sample size increased to $n = 300$ and $n = 900$, RMSE decreased across all priors, and the ridge continued to show the highest RMSE for null effects.

The slope RMSE (Figure 3.11) showed a similar ordering among priors as the intercept. The ridge produced the lowest RMSE for large effects, and the horseshoe, regularized horseshoe, and spike-and-slab achieved the lowest RMSE for null effects. Under unequal proportions, the slope RMSE was elevated across sample sizes compared to equal proportions, with this difference most pronounced at $n = 900$.

Figure 3.8: Relative Bias Comparison Between Homogeneous and Heterogeneous Patterns at $n = 900$: Slope Coefficients.

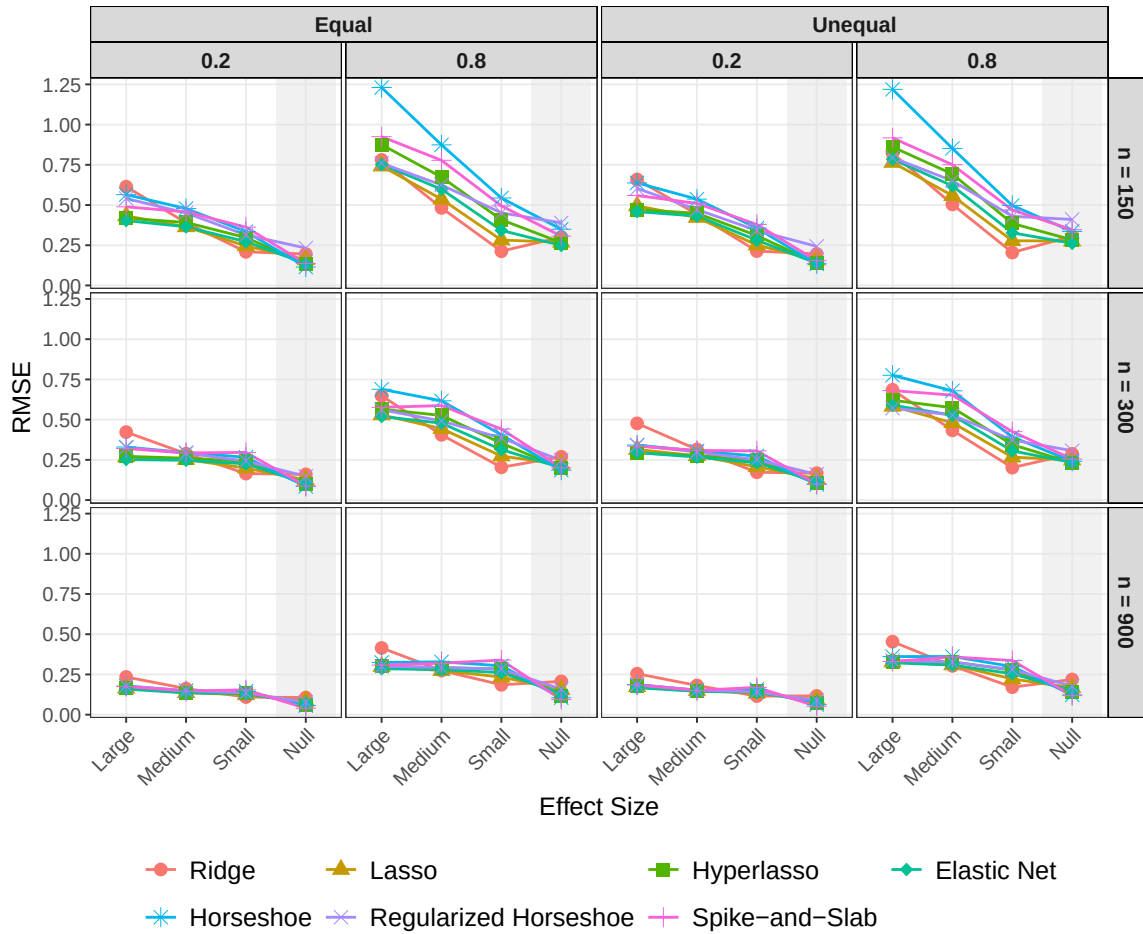


with $\rho = 0.8$. The heterogeneous predictor effect pattern and class-specific results produced similar RMSE patterns (see Supplementary Materials).

3.7.3 Variable Selection Accuracy

Variable selection accuracy concerns the extent to which shrinkage priors appropriately handle predictors of different effect sizes. This includes whether large and medium effects are retained despite partial shrinkage, whether small effects are sufficiently shrunk toward zero, and whether predictors with true zero effects are excluded. Compared to the estimation accuracy results presented previously, the variable selection results provide a more direct assessment of how the shrinkage patterns translate into selection decisions by examining the selection rates across large, medium, small, and null effect sizes. I evaluated variable selection accuracy using the posterior median criterion (i.e., a predictor was deemed selected if the absolute value of the posterior median exceeded 0.1). As with estimation accuracy, results are presented under moderate separation ($MD = 3.0$), with the corresponding results under $MD = 4.0$ provided in the Supplementary Materials. Results based on the posterior mean and 95% credible interval criteria are discussed in the sensitivity analysis below.

Figure 3.9: Root Mean Square Error for Class Membership Coefficients.

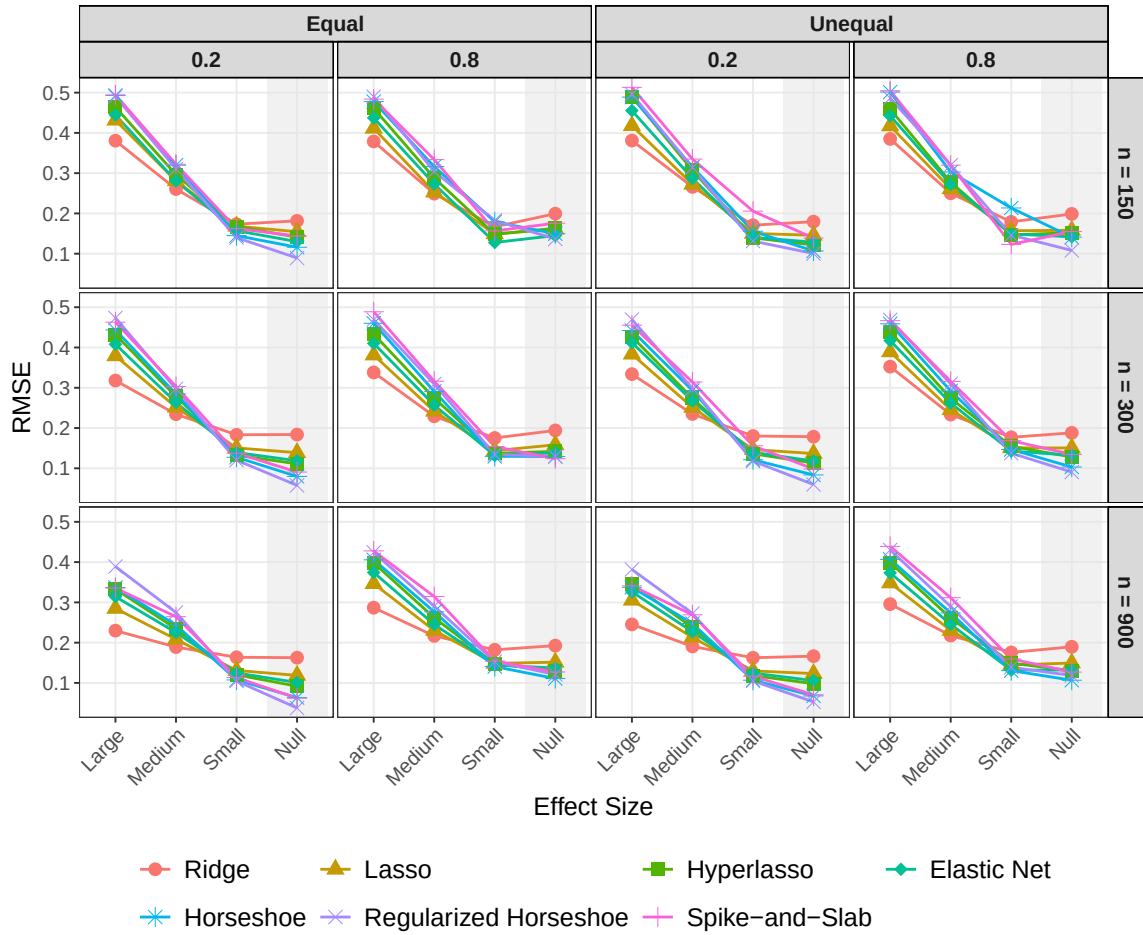


Class Membership Coefficients

Figure 3.12 shows the selection rates for the class membership coefficients under the posterior median criterion. Each row corresponds to a sample size ($n = 150, 300$, and 900), and the columns are organized by class proportion (Equal and Unequal), within which predictor intercorrelation ($\rho = 0.2$ and 0.8) is nested. In each panel, the x -axis is the effect size category (large, medium, small, and null), the y -axis is the selection rate, and each line with a different color and shape represents a different shrinkage prior. Higher selection rates for nonzero effects indicate greater power to detect true predictors, and lower selection rates in the null region (shaded) indicate better control of false selections. Ideally, selection rates should be high for large and most medium effects, although not necessarily 100% due to shrinkage. For small effects, selection rates should be lower because shrinkage pulls these effects toward zero. For null effects, selection rates should be close to zero.

In Figure 3.12, beginning with $n = 150$ under equal proportions with $\rho = 0.2$, all priors achieved near-perfect selection for large effects. Although several priors

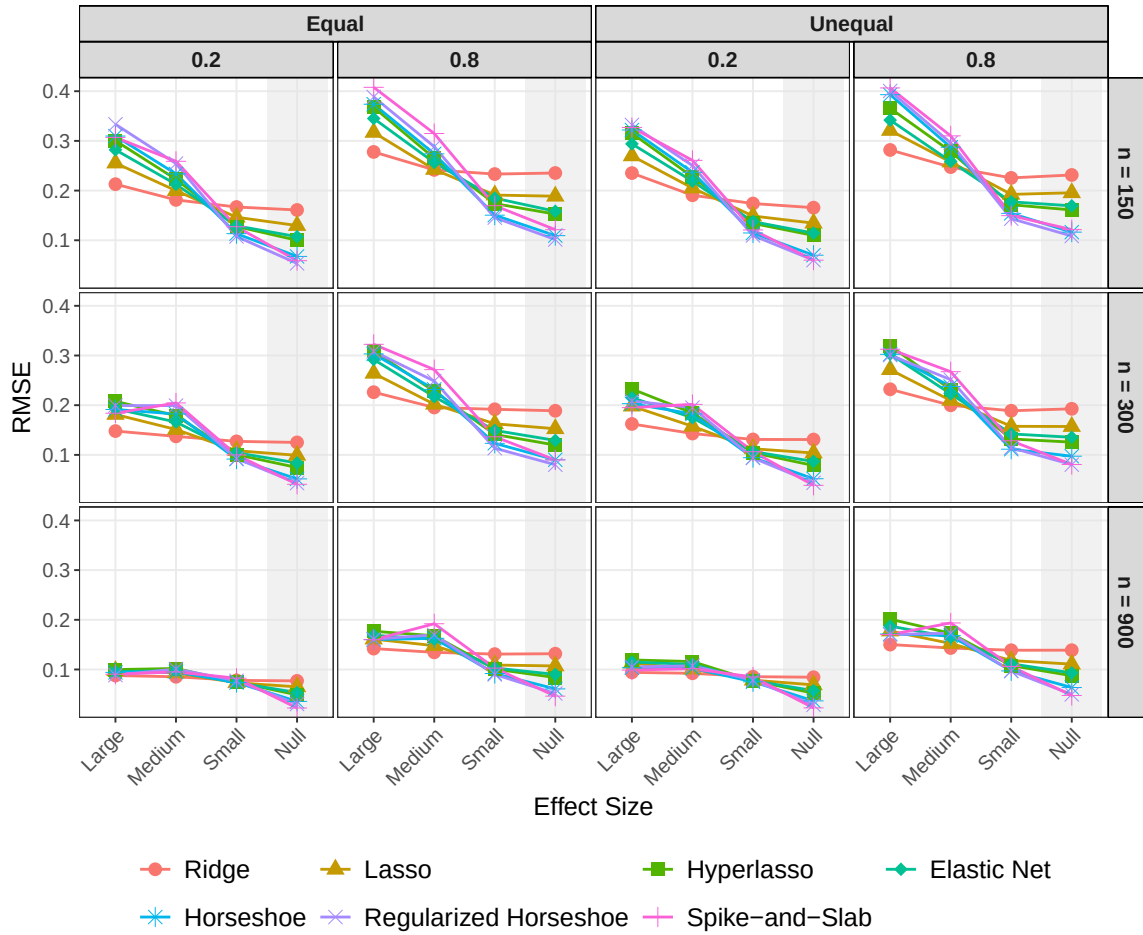
Figure 3.10: Root Mean Square Error for Intercept Growth Factor Regression Coefficients Under the Homogeneous Pattern.



produced negative relative bias for these large effects (Figure 3.4), shrunken posterior estimates of $\gamma = 1.386$ remained well above the 0.1 selection threshold. For medium effects, selection rates remained high across most priors. The priors diverged most for small and null effects. For small effects, the ridge showed the highest selection rate, followed by the lasso, and the spike-and-slab showed the lowest followed by the horseshoe. This indicates that the spike-and-slab and horseshoe imposed much stronger shrinkage on small effects, making these priors more conservative in selecting weak signals. In the null region, the ridge produced the highest false selection rate, and the spike-and-slab and horseshoe the lowest, indicating that the ridge suffered from the most severe false positive problem among all priors.

When predictor intercorrelation increased to $\rho = 0.8$, selection rates for small effects decreased across all priors, and false selection rates for null effects increased. The ridge showed the largest increase in false selection for null effects, reaching approximately 80%. The spike-and-slab and horseshoe showed noticeable decreases in selection rates for large and medium effects under high intercorrelation. Under

Figure 3.11: Root Mean Square Error for Slope Growth Factor Regression Coefficients Under the Homogeneous Pattern.



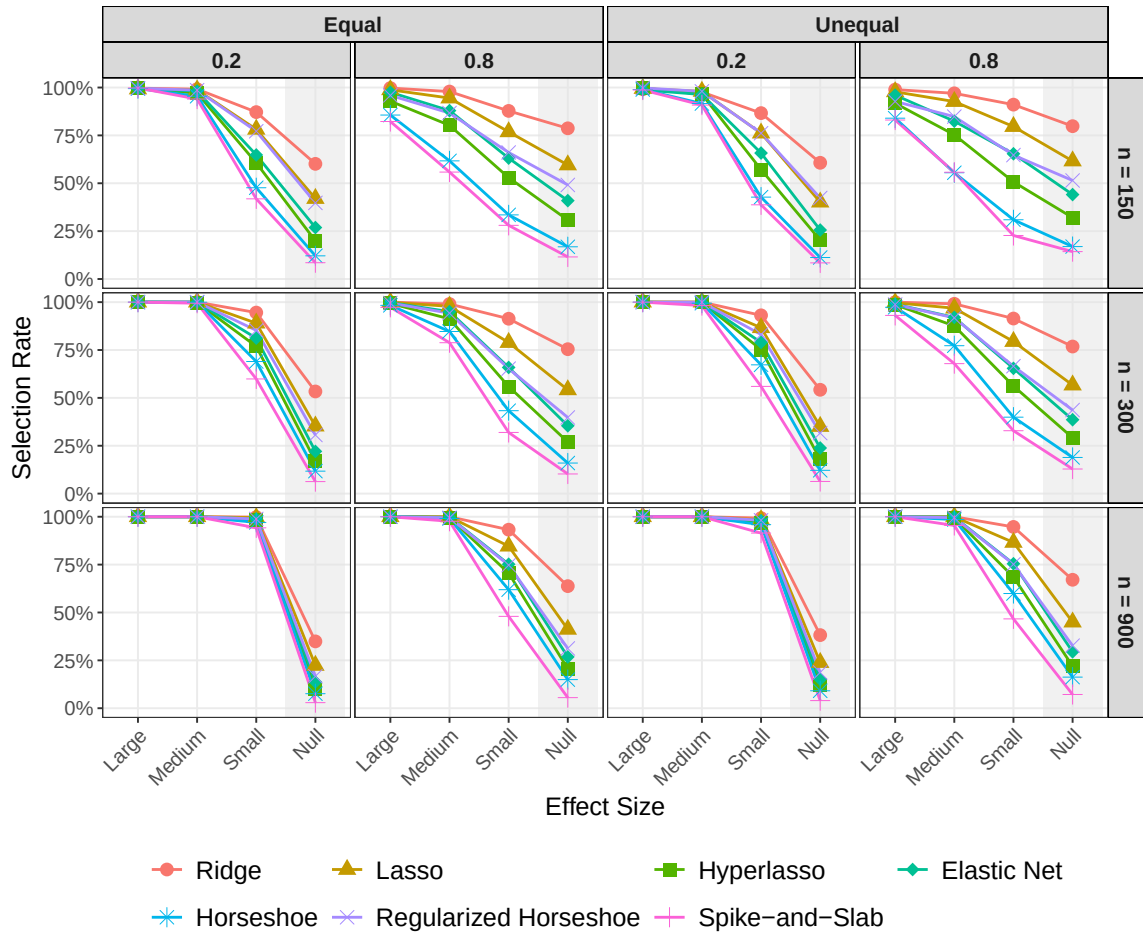
unequal proportions at $n = 150$, the patterns closely resembled those under equal proportions.

As sample size increased to $n = 300$ and $n = 900$, selection rates for large and medium effects reached near 100% for all priors. For small effects, selection rates also improved, with the ordering among priors persisting. This indicates that with larger samples, the shrinkage effects of the priors weakened, allowing more true effects to be detected. For null effects, false selection rates generally decreased with sample size for most priors, but the ridge maintained elevated false selection rates even at $n = 900$.

Growth Factor Regression Coefficients

Figures 3.13 and 3.14 display the selection rates for the intercept and slope growth factor regression coefficients, respectively, under the homogeneous predictor effect pattern. The figures follow the same format as Figure 3.12. Under the homo-

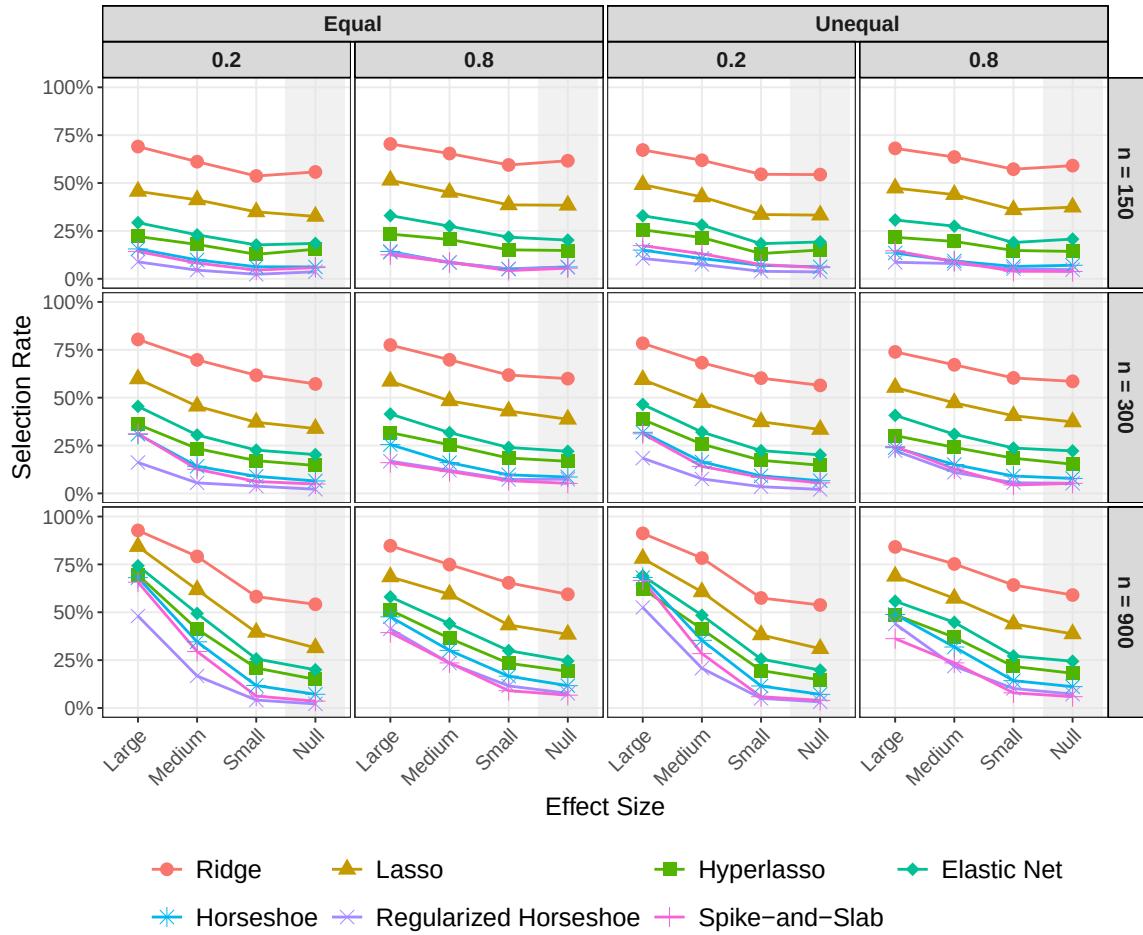
Figure 3.12: Selection Rates for Class Membership Coefficients.



geneous pattern, both classes share the same true coefficient values, and the class-specific selection rates showed negligible differences, so results are aggregated across classes. The full results including the class-specific results and those under the heterogeneous pattern are available in the Supplementary Materials. A focused comparison between the homogeneous and heterogeneous patterns is presented subsequently in Figure 3.15.

In Figure 3.13, selection rates for the intercept coefficients were overall lower than those for the class membership coefficients across all priors and conditions. At $n = 150$, selection rates for each prior remained relatively flat from large to null effects across all class proportions and predictor intercorrelation levels. For example, under equal proportions with $\rho = 0.2$, the ridge maintained selection rates between approximately 55% and 70% regardless of effect size, and the regularized horseshoe and spike-and-slab maintained rates between approximately 2% and 15%. This flat pattern indicates that at this sample size, the priors did not effectively discriminate between true and null effects for the intercept coefficients. As sample size increased to $n = 300$ and $n = 900$, selection rates for large and medium effects increased, and

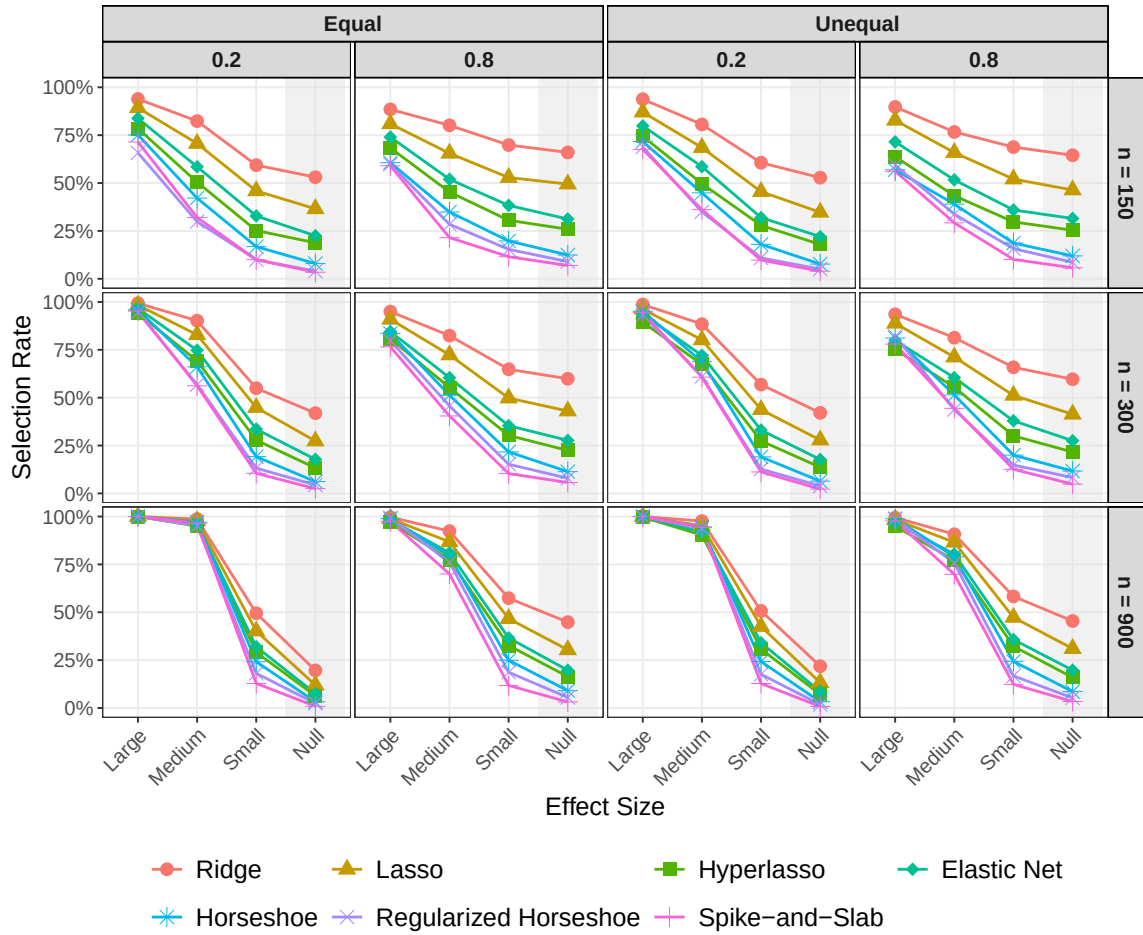
Figure 3.13: Selection Rates for Intercept Growth Factor Regression Coefficients Under the Homogeneous Pattern.



the priors began to show greater separation across effect sizes. This improvement was most visible at $n = 900$ with $\rho = 0.2$ regardless of class proportion, where the regularized horseshoe and spike-and-slab showed increased selection for large effects.

The slope coefficients in Figure 3.14 showed greater differentiation across effect sizes than the intercept coefficients. This indicates that the priors discriminated more effectively between true and null effects for the slope. Beginning with $n = 150$ under equal proportions with $\rho = 0.2$, the priors fell into two groups. The ridge, lasso, hyperlasso, and elastic net showed high selection rates for large effects and slightly lower but still high rates for medium effects. For small effects, these priors maintained relatively high selection rates. Then, for null effects, a substantial proportion of predictors were still selected, which points to a false positive problem. In contrast, the horseshoe, regularized horseshoe, and spike-and-slab showed lower selection rates for large and medium effects, reflecting the stronger shrinkage imposed by these priors. For small effects, selection rates were lower than those of the first group, and for null effects, selection rates approached zero. These priors effectively controlled

Figure 3.14: Selection Rates for Slope Growth Factor Regression Coefficients Under the Homogeneous Pattern.



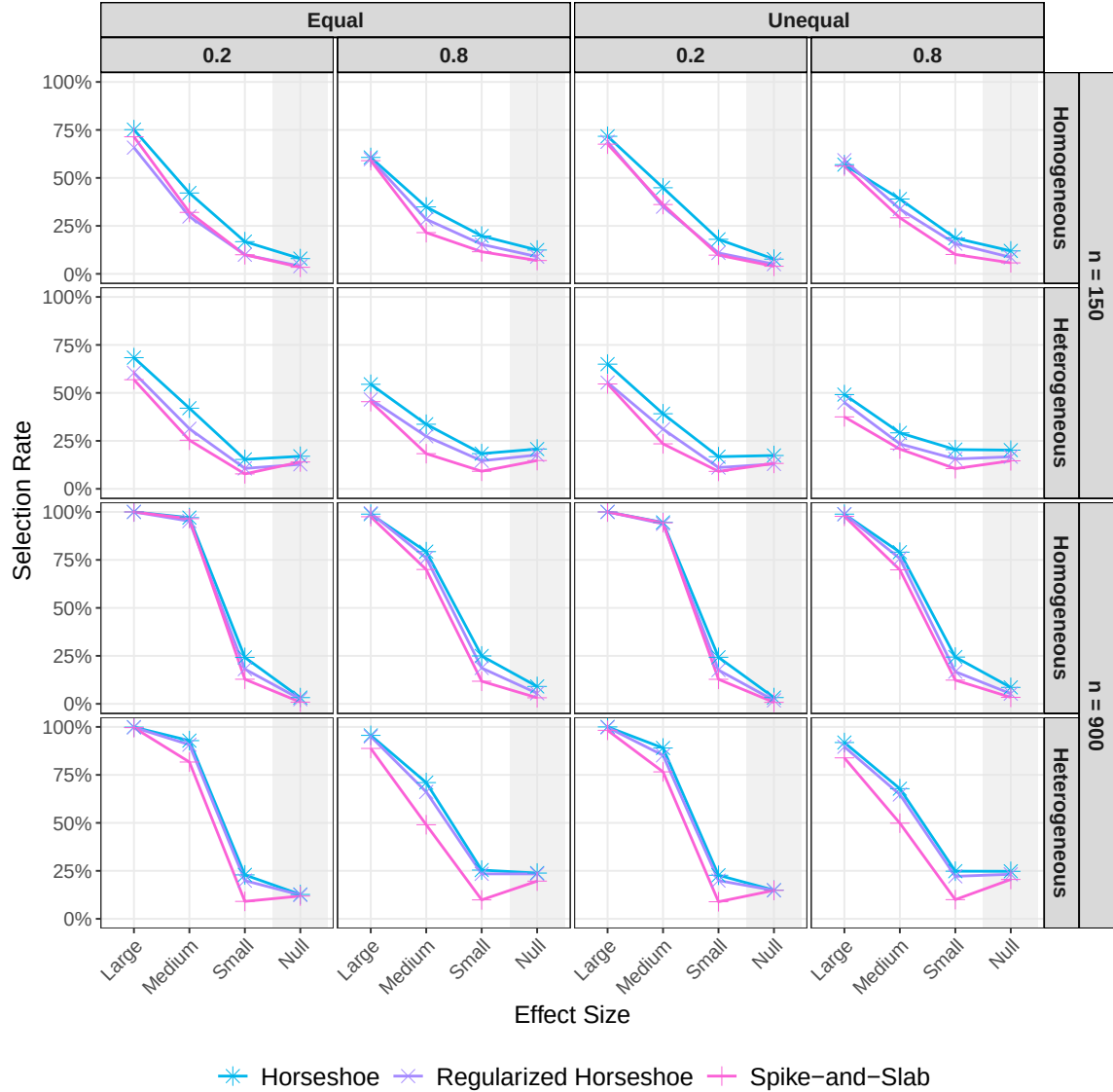
false positives at the cost of reduced power for weak signals. When predictor inter-correlation increased to $\rho = 0.8$, selection rates for medium effects decreased across all priors. Selection rates for small effects increased for the ridge and lasso, and false selection rates for null effects increased for the ridge, lasso, hyperlasso, and elastic net but remained similar for the horseshoe, regularized horseshoe, and spike-and-slab. Under unequal proportions, the patterns were similar to those observed under equal proportions.⁸

For $n = 300$, selection rates increased overall, with improvements most visible for large and medium effects across all priors. At $n = 900$, the results began to approach the ideal selection pattern. Even at this sample size, nuanced differences among priors remained. The ridge, lasso, hyperlasso, and elastic net maintained relatively high selection rates for small effects. The horseshoe, regularized horseshoe, and spike-and-slab continued to impose stronger shrinkage on small effects and showed

⁸When examining class-specific results, the minority class showed slightly lower overall selection rates under unequal proportions, although this difference was minor (see Supplementary Materials).

lower selection rates for weak signals but better control of false selections for null effects.

Figure 3.15: Selection Rates for Slope Regression Coefficients Comparing Homogeneous and Heterogeneous Patterns ($n = 150$ and 900).



As described earlier, the full comparison between homogeneous and heterogeneous patterns is available in the Supplementary Materials, but one important pattern warrants presentation here. In particular, Figure 3.15 compares the selection rates for the horseshoe, regularized horseshoe, and spike-and-slab under the homogeneous and heterogeneous predictor effect patterns at $n = 150$ and $n = 900$. Within each sample size, the homogeneous and heterogeneous patterns are subdivided in the rows. For large and medium effects, selection rates were somewhat lower under the heterogeneous pattern at $n = 150$ but converged across both patterns at $n = 900$. The more

notable difference emerged for null effects. Under the homogeneous pattern (first and third rows), these three priors maintained low false selection rates for null effects, and these rates remained stable or decreased as sample size increased. Under the heterogeneous pattern (second and fourth rows), false selection rates for null effects were elevated relative to the homogeneous pattern, especially under $\rho = 0.8$ where rates increased with sample size. At $n = 900$ with $\rho = 0.8$ across both class proportion conditions, false selection rates for null effects reached approximately 20% to 25% under the heterogeneous pattern, compared to below 10% under the homogeneous pattern. This pattern may reflect the increased difficulty of these three priors in estimating class-specific sparsity structures when the active predictor sets differ across classes.

Sensitivity to Selection Criterion

The variable selection results presented thus far were based on the posterior median criterion, under which a predictor was deemed selected if the absolute value of the posterior median exceeded 0.1. To examine whether the findings were robust to this choice, I conducted sensitivity analyses using two additional criteria. The first applied the same 0.1 threshold to the posterior mean, and the second used the 95% equal-tailed credible interval excluding zero. Figure 3.16 presents a subset of comparisons for the slope growth factor regression coefficients as an illustrative example, and the full results under all three criteria for the class membership, intercept growth factor, and slope growth factor regression coefficients are provided in the Supplementary Materials.

Figure 3.16: Sensitivity of Selection Rates for Slope Growth Factor Regression Coefficients to Selection Criterion ($n = 150$ and 900 , Equal Proportions, Homogeneous Pattern, MD = 3.0).

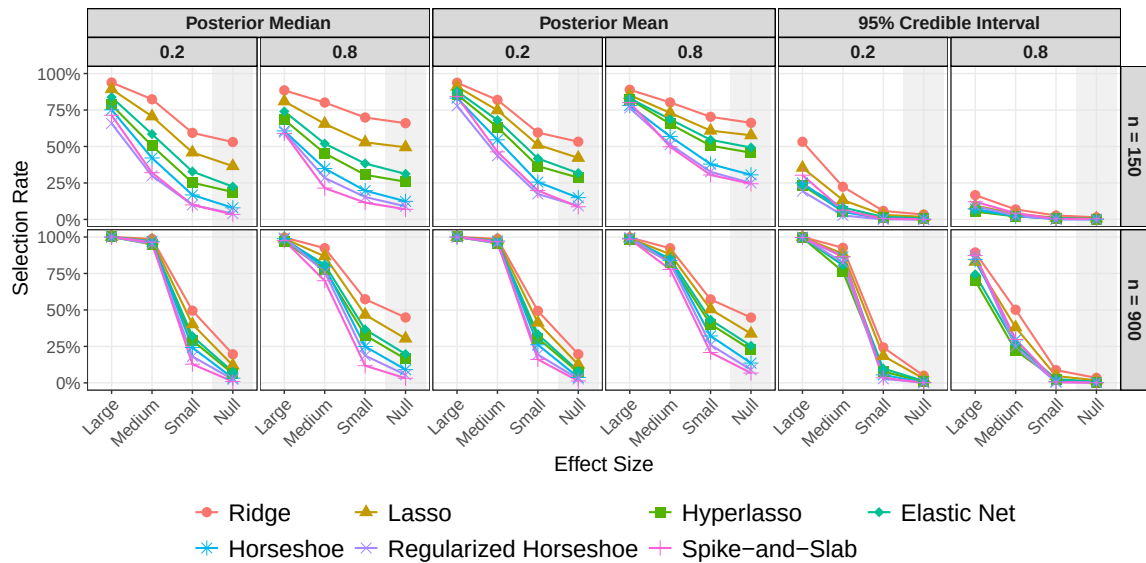


Figure 3.16 compares the three selection criteria under equal proportions with

the homogeneous pattern at $n = 150$ and $n = 900$. Each row corresponds to a sample size, and the columns are organized by criterion (posterior median, posterior mean, and 95% credible interval), within which predictor intercorrelation ($\rho = 0.2$ and 0.8) is nested.

At $n = 150$ (first row) with $\rho = 0.2$, the posterior median criterion (first column) showed the selection patterns described in the preceding paragraphs. The posterior mean criterion (third column) produced higher selection rates overall. Selection rates for large and medium effects increased, but selection rates for small and null effects also increased, indicating that the posterior mean criterion inflated both power and false positives. The 95% credible interval criterion (fifth column) produced substantially lower selection rates overall. Selection rates under this criterion were near zero for most priors across all effect sizes, with the exception of the ridge, which selected the large effect at approximately 55%. These low selection rates arose because the 95% credible interval frequently included zero even when the posterior median was nonzero, and this was more pronounced at smaller sample sizes. When predictor intercorrelation increased to $\rho = 0.8$ (second, fourth, and sixth columns), selection rates for large and medium effects generally decreased across all three criteria, and false selection rates for null effects increased for the posterior median and posterior mean criteria, making the distinction between true and null effects more difficult. For the 95% credible interval criterion, selection rates remained near zero for most priors regardless of effect size.

When $n = 900$ (second row), the posterior median and posterior mean criteria showed more differentiated patterns across effect sizes, approaching the ideal selection pattern. The 95% credible interval criterion improved, but selection rates remained lower than those under the other two criteria. Under $\rho = 0.8$, the credible interval criterion still showed low selection rates for medium effects. The relative ordering of priors was preserved under all three criteria, with the ridge retaining the highest false selection rates and the spike-and-slab and regularized horseshoe the lowest.

3.7.4 Classification Accuracy

Table 3.4 presents the mean adjusted Rand index (ARI) values across priors, sample sizes, and class separation levels. Because ARI values showed negligible differences across levels of class proportion, predictor effect pattern, and predictor intercorrelation, Table 3.4 presents values averaged across these factors. I provide the full results across all conditions in the Supplementary Materials.

Classification accuracy was primarily driven by sample size and class separation rather than the choice of shrinkage prior. At $MD = 3.0$ with $n = 150$, mean ARI values ranged from .813 (ridge) to .828 (hyperlasso), a difference of only .015. At $n = 900$ with $MD = 3.0$, the range narrowed to .846 to .854. Under $MD = 4.0$, all priors achieved ARI values above .930, with differences among priors no more than .005. These results indicate that the choice of shrinkage prior had minimal impact on the recovery of the latent class structure, and the primary determinants of classification

Table 3.4: Mean Adjusted Rand Index Across Shrinkage Priors, Sample Sizes, and Class Separation Levels.

Prior	$n = 150$		$n = 300$		$n = 900$	
	MD = 3.0	MD = 4.0	MD = 3.0	MD = 4.0	MD = 3.0	MD = 4.0
Ridge	.813	.935	.838	.942	.852	.946
Lasso	.820	.936	.841	.943	.854	.946
Hyperlasso	.828	.935	.841	.942	.854	.946
Elastic net	.824	.935	.840	.942	.854	.946
Horseshoe	.820	.935	.839	.942	.854	.946
Reg. horseshoe	.815	.933	.834	.939	.847	.944
Spike-and-slab	.823	.931	.831	.939	.846	.944

Note. Reg. horseshoe = regularized horseshoe.

accuracy were class separation and sample size.

3.7.5 Predictive Performance

For predictive performance evaluated using WAIC and LOOIC, results are displayed in Tables 3.5 and 3.6, where lower values indicate better predictive performance. Because WAIC and LOOIC values showed negligible differences across levels of class proportion and predictor effect pattern, these factors were averaged in the tables. Predictor intercorrelation, which produced larger differences in absolute values, is presented separately. The full results across all conditions are provided in the Supplementary Materials.

As shown in Tables 3.5 and 3.6, predictive performance was mainly driven by the choice of shrinkage prior. The WAIC (Table 3.5) and LOOIC (Table 3.6) values were highly concordant, differing by less than 2 units across all conditions. Overall, considering the full results across all conditions (with those in Supplementary Materials), the regularized horseshoe achieved the best predictive performance, followed by the spike-and-slab and horseshoe. The ridge produced the highest (worst) values across all conditions. Higher predictor intercorrelation ($\rho = 0.8$) increased WAIC and LOOIC values across all priors, but the ordering among priors remained consistent across both levels of predictor intercorrelation and class separation.

3.8 Discussion

Researchers employ GMMs for describing heterogeneous growth trajectories by identifying latent classes with distinct developmental patterns. The practical utility of these models is further enhanced when predictors are introduced to explain and contextualize the identified classes. However, a principled approach for determining which predictors to include has been largely absent from the GMM literature. This chapter addressed this gap by evaluating the performance of seven Bayesian shrinkage

Table 3.5: Mean WAIC Across Shrinkage Priors, Sample Sizes, Predictor Intercorrelations, and Class Separation Levels.

Prior	$n = 150$		$n = 300$		$n = 900$	
	MD = 3.0	MD = 4.0	MD = 3.0	MD = 4.0	MD = 3.0	MD = 4.0
$\rho = 0.2$						
Ridge	2743.5	2758.0	5440.7	5479.6	16263.1	16393.9
Lasso	2738.5	2755.7	5433.8	5475.1	16250.1	16383.7
Hyperlasso	2733.9	2753.2	5429.0	5472.2	16242.6	16381.0
Elastic net	2736.5	2754.5	5432.3	5473.6	16245.2	16381.8
Horseshoe	2727.6	2748.7	5424.1	5469.2	16241.0	16379.3
Reg. horseshoe	2723.3	2745.2	5418.6	5465.9	16236.4	16377.0
Spike-and-slab	2727.0	2750.0	5425.4	5469.9	16244.7	16381.4
$\rho = 0.8$						
Ridge	2848.1	2854.4	5610.9	5635.1	16705.4	16798.0
Lasso	2843.7	2852.3	5606.2	5632.9	16697.6	16791.1
Hyperlasso	2837.5	2849.4	5602.1	5628.3	16691.4	16785.8
Elastic net	2840.6	2849.5	5604.0	5630.6	16692.9	16789.5
Horseshoe	2829.5	2842.9	5595.6	5624.0	16685.5	16784.1
Reg. horseshoe	2826.3	2839.3	5582.7	5618.0	16683.1	16783.6
Spike-and-slab	2835.2	2843.8	5590.4	5626.8	16691.4	16782.2

Note. Reg. horseshoe = regularized horseshoe. WAIC = Watanabe-Akaike information criterion.

priors for variable selection in GMMs through a comprehensive Monte Carlo simulation that provides a detailed characterization of how these priors behave across a wide range of conditions.

Broadly speaking, the shrinkage priors differed substantially in their bias and variable selection profiles. The horseshoe, regularized horseshoe, and spike-and-slab offered the best balance between detecting true predictors and controlling false selections, although the regularized horseshoe and spike-and-slab exhibited lower convergence rates than the remaining priors. These differences among priors were most pronounced at smaller sample sizes, where the shrinkage behavior of each prior had a stronger impact on selection outcomes. The type of coefficient being predicted also mattered. Variable selection for the intercept growth factor regression coefficients was poor across all priors. In addition, the choice of selection criterion influenced the results, with the posterior mean producing higher selection rates and the 95% credible interval producing overly conservative ones. Classification accuracy showed minimal differences across priors and was primarily driven by class separation and sample size. On the other hand, predictive performance was influenced by the choice of prior, with the regularized horseshoe achieving the best predictive accuracy followed by the spike-and-slab and horseshoe. I discuss these findings and their implications below.

Table 3.6: Mean LOOIC Across Shrinkage Priors, Sample Sizes, Predictor Intercorrelations, and Class Separation Levels.

Prior	$n = 150$		$n = 300$		$n = 900$	
	MD = 3.0	MD = 4.0	MD = 3.0	MD = 4.0	MD = 3.0	MD = 4.0
$\rho = 0.2$						
Ridge	2745.3	2759.4	5441.2	5480.0	16263.2	16394.0
Lasso	2740.3	2757.2	5434.3	5475.5	16250.2	16383.8
Hyperlasso	2735.5	2754.6	5429.4	5472.6	16242.7	16381.1
Elastic net	2738.2	2756.0	5432.8	5473.9	16245.2	16381.8
Horseshoe	2729.2	2750.2	5424.6	5469.5	16241.1	16379.4
Reg. horseshoe	2724.8	2746.5	5419.0	5466.2	16236.5	16377.0
Spike-and-slab	2728.5	2751.4	5425.7	5470.2	16244.8	16381.4
$\rho = 0.8$						
Ridge	2849.8	2856.0	5611.4	5635.5	16705.5	16798.1
Lasso	2845.3	2854.0	5606.7	5633.3	16697.7	16791.2
Hyperlasso	2839.2	2851.0	5602.5	5628.7	16691.5	16785.9
Elastic net	2842.3	2851.0	5604.5	5631.0	16693.0	16789.6
Horseshoe	2830.9	2844.5	5596.0	5624.3	16685.6	16784.2
Reg. horseshoe	2827.8	2840.8	5583.1	5618.3	16683.2	16783.7
Spike-and-slab	2836.6	2845.1	5590.7	5627.1	16691.5	16782.3

Note. Reg. horseshoe = regularized horseshoe. LOOIC = leave-one-out information criterion.

3.8.1 Shrinkage Behavior and the Role of Effect Magnitude

A central finding of this chapter is that the degree of shrinkage applied by each prior varied with the magnitude of the true effects. For the class membership coefficients, which had larger true values ($\gamma = 1.386, 0.916, 0.405$), the ridge exhibited the strongest negative relative bias for large and medium effects. This reflects uniform shrinkage that does not differentiate between effect sizes. The horseshoe, regularized horseshoe, and spike-and-slab, in contrast, produced less negative or even positive relative bias for large and medium effects, consistent with their theoretical design to preserve large signals (Bhadra et al., 2019; van Erp et al., 2019). Priors with only global penalty parameters (ridge, lasso) apply uniform shrinkage, whereas priors with local penalty parameters (horseshoe, regularized horseshoe, spike-and-slab) can adapt the degree of shrinkage to each coefficient.

The hyperlasso and elastic net fell between these two groups, with the hyperlasso producing relative bias within the ± 0.1 acceptability band for large class membership effects and the elastic net behaving more similarly to the lasso. Although the hyperlasso incorporates predictor-specific local variance parameters and the elastic net combines the lasso and ridge penalties, neither achieved the degree of adaptive shrinkage exhibited by the horseshoe, regularized horseshoe, and spike-and-slab. The hyperlasso's gamma hyperprior on the local variance may not have provided sufficient tail mass to fully protect large coefficients from shrinkage in the GMM setting. This is because the thinner tails of the gamma distribution, compared

to the half-Cauchy used in the horseshoe, limit how large the local variance can grow for any given coefficient. For the elastic net, the ridge component introduces uniform shrinkage that partially counteracts the selective shrinkage of the lasso component, limiting its adaptive capacity.

For the growth factor regression coefficients, which had smaller true values ($\beta = 0.5, 0.3, 0.1$), the prior ordering reversed. The ridge exhibited the least negative relative bias, and the regularized horseshoe the most. This reversal reflects the adaptive shrinkage mechanism. When true effects are small, the horseshoe, regularized horseshoe, and spike-and-slab treated these effects as closer to the noise level and shrank them more heavily. A similar pattern was reported by Zhang and Liang (2024) in MIMIC models, where the horseshoe showed strong shrinkage for small structural coefficients. This reversal is consistent with the theoretical design of adaptive shrinkage priors, which distinguish signal from noise based on effect magnitude. In the variable selection framework, this behavior is desirable because large true effects are preserved for detection and small effects that are difficult to distinguish from zero are regularized toward the null.

The intercept growth factor coefficients were more heavily shrunk than the slope coefficients across all conditions. In the present design, the intercept growth factor had a variance of 18, whereas the slope had a variance of 2. Because predictors were generated with unit variance, a regression coefficient of $\beta = 0.5$ explained approximately 1.4% of the intercept growth factor variance ($0.5^2/18$) but 12.5% of the slope growth factor variance ($0.5^2/2$). The same coefficient value thus carried different implications across the two factors because they were on a raw metric rather than standardized. The intercept coefficients represented a smaller share of explained variance and were therefore harder to distinguish from zero, leading to greater shrinkage and poorer variable selection. This outcome follows from applying identical coefficient values across growth factors with different variances, and subsequent work may consider a wider range of coefficient values to further characterize how shrinkage priors behave across different growth factor metrics. Still, growth factor regression coefficients in applied GMM research are often of modest magnitude, and intercept factors tend to have larger variance than slope factors (see, e.g., Heo, Depaoli, et al., 2024; Kaplan, 2002).

Predictor intercorrelation intensified shrinkage across most priors. Under $\rho = 0.8$, relative bias generally became more negative for nonzero effects and positive bias increased for null effects, narrowing the gap between true and null estimates and making it harder for the priors to discriminate between signal and noise. This is consistent with findings from Bainter et al. (2023), who reported that correlated predictors increased variability in coefficient estimates and reduced the accuracy of variable selection in a regression setting. The spike-and-slab and regularized horseshoe showed weaker shrinkage for slope coefficients at larger sample sizes and eventually produced less bias than most of the remaining priors in recovering the true effects. This suggests that with sufficient data, the adaptive shrinkage mechanism can overcome initial overshrinkage for coefficients whose effects are sufficiently large relative

to the growth factor variance.

The RMSE results complemented the relative bias findings by reflecting the bias-variance tradeoff. For the growth factor regression coefficients, the ridge achieved the lowest RMSE for large effects, as its milder shrinkage both preserved estimates closer to the true values and reduced estimation variance. For the class membership coefficients, however, the ridge and horseshoe produced the highest RMSE for large effects, driven by strong negative bias and high estimation variability, respectively. Conversely, the horseshoe, regularized horseshoe, and spike-and-slab showed higher RMSE for large growth factor effects due to stronger shrinkage, which produced larger bias for the regularized horseshoe and greater estimation variability for the horseshoe and spike-and-slab, but lower RMSE for null effects where their strong shrinkage reduced both bias and variance simultaneously.

3.8.2 Variable Selection: Power, False Positives, and the Heterogeneous Pattern

The variable selection results demonstrated a tradeoff between the power to detect true predictors and the control of false selections. For the class membership coefficients, the ridge achieved the highest selection rates across all effect sizes but suffered from the highest false positive rates, approaching 80% for null effects under high intercorrelation at $n = 150$. The horseshoe and spike-and-slab achieved the lowest false positive rates, consistent with findings from Bainter et al. (2023), who reported that stochastic search variable selection (a spike-and-slab variant) maintained the lowest false inclusion rates among the methods they compared in a linear regression setting.

The intercept coefficients showed a flat selection curve at $n = 150$, indicating that no prior effectively discriminated between true and null effects at this sample size. The slope coefficients fared better, and two groups of priors emerged. The ridge, lasso, hyperlasso, and elastic net showed higher selection rates but also higher false positives, and the horseshoe, regularized horseshoe, and spike-and-slab showed lower selection rates but better false positive control. Although the hyperlasso and elastic net incorporate local shrinkage parameters in their penalty structures, they behaved more similarly to the ridge and lasso in terms of variable selection and produced higher false positive rates than the horseshoe, regularized horseshoe, and spike-and-slab. This grouping extends the global-versus-local penalty distinction described by van Erp et al. (2019), suggesting that the presence of local shrinkage parameters alone is not sufficient and that the degree of local adaptivity determines variable selection performance. A similar pattern was observed by Harra and Kaplan (2024), who found that the horseshoe and regularized horseshoe outperformed the ridge and lasso in structural equation models.

One noteworthy finding involved the heterogeneous predictor effect pattern. Under this pattern, where the active predictor sets differed across classes, the horseshoe, regularized horseshoe, and spike-and-slab showed elevated false selection rates

for null slope coefficients, especially under $\rho = 0.8$ where these rates increased with sample size. No such pattern was observed under the homogeneous condition, suggesting that these priors may face additional challenges in recovering class-specific sparsity structures when the true sparsity patterns differ across latent classes. The finding connects to the broader challenge of variable selection in finite mixture of regression models (Khalili & Chen, 2007), where selecting component-specific predictor sets becomes more difficult under unequal mixing proportions and differing coefficient structures across components. Researchers working with GMMs in which predictors are expected to operate differently across classes should examine heterogeneous pattern results carefully.

The sensitivity analysis revealed that the posterior mean criterion generally produced higher selection rates than the posterior median criterion used in this chapter. This is because the posterior mean is more sensitive to extreme draws in the tails of the posterior distribution, which can inflate the point estimate above the selection threshold even when the majority of posterior mass is near zero. Compared to these threshold-based criteria, the 95% credible interval criterion was overly conservative and produced near-zero selection rates at small sample sizes. Consistent with these patterns, Zhang and Liang (2024) reported that the threshold criterion was more robust to sample size than the credible interval criterion in MIMIC models. Relatedly, van Erp et al. (2019) showed that the optimal credibility level varied across conditions and cautioned that applying a single credibility level uniformly across all coefficients may not be appropriate for priors that induce shrinkage jointly.

3.8.3 Convergence, Classification, and Predictive Performance

The regularized horseshoe and spike-and-slab exhibited the lowest convergence rates, especially under high predictor intercorrelation (as detailed in the Supplementary Materials). The ridge, lasso, hyperlasso, and horseshoe converged at substantially higher rates, and the elastic net fell slightly below these four but well above the regularized horseshoe and spike-and-slab. The convergence challenges for the latter two priors likely stem from the increased complexity of the posterior landscape introduced by the local shrinkage parameters, which can slow MCMC mixing and delay convergence. Convergence and estimation challenges for horseshoe-type priors have been observed in other modeling contexts as well, with van Erp et al. (2019) reporting convergence difficulties for the horseshoe and regularized horseshoe in regression models and Zhang and Liang (2024) finding lower convergence rates for the horseshoe than the ridge and lasso in MIMIC models under high collinearity conditions. An additional observation is that convergence rates at $n = 900$ did not improve beyond those at $n = 300$ and were slightly lower for several priors. One possible explanation is that as sample size grows, the likelihood becomes more concentrated and can create tension with the shrinkage prior. This tension can result in a posterior geometry that is more difficult for the MCMC sampler to navigate efficiently within the fixed number of iterations used in this chapter.

Classification accuracy, as measured by the ARI, showed negligible differences across priors. This indicates that the choice of shrinkage prior for predictor selection does not compromise the recovery of the latent class structure. The primary determinants of classification accuracy were class separation and sample size.

Predictive performance did differ across priors. The regularized horseshoe generally achieved the lowest WAIC and LOOIC values, followed by the spike-and-slab and horseshoe. This indicates superior out-of-sample prediction. This finding parallels the results of Harra and Kaplan (2024), who found that the horseshoe priors outperformed the ridge and lasso in predictive accuracy for path analysis and growth curve models. The hyperlasso and elastic net produced intermediate predictive performance that fell between the ridge and lasso on one end and the horseshoe, regularized horseshoe, and spike-and-slab on the other. The superior predictive performance of the latter group is consistent with their variable selection properties. By more effectively shrinking noise coefficients toward zero, these priors produce models with lower effective complexity and better generalization.

3.8.4 Advice for Applied Researchers

One of the goals of this chapter is to provide practical guidelines for applied researchers. Drawing on the simulation results, I synthesize the findings into a performance summary (Table 3.7), a decision framework (Figure 3.17), condition-specific recommendations, and a set of caveats that researchers should bear in mind. Throughout this workflow, variable selection decisions should be grounded in both statistical criteria and substantive context.

Prior Performance at a Glance

Table 3.7 summarizes each prior’s performance across three key dimensions, namely selection accuracy (the joint balance of power for true effects and control of false selections), predictive accuracy (WAIC and LOOIC), and convergence. The regularized horseshoe and spike-and-slab achieved the strongest selection accuracy and, together with the horseshoe, showed the desirable predictive accuracy. However, the regularized horseshoe and spike-and-slab required substantially more MCMC effort to achieve convergence. The horseshoe achieved the best all-around balance, combining strong selection and predictive performance with reliable convergence. Ridge is included for completeness but is not recommended for variable selection because it cannot shrink coefficients to exactly zero and produced the highest false-selection rates across conditions.

Decision Framework

Figure 3.17 organizes these recommendations into a four-step workflow. Under Step 1, the prior choice branches on sample size, with the regularized horseshoe or spike-and-slab recommended when $n \geq 300$ (with longer MCMC chains) and the

Table 3.7: Summary of Shrinkage Prior Performance Across Key Dimensions.

Prior	Selection Accuracy	Predictive Accuracy	Convergence	Key Use Case
Ridge	○	○	●●●	Regularization only; not suited for discrete selection
Lasso	●	●	●●●	Reliable convergence with moderate selection
Hyperlasso	●	●	●●	Middle-tier general-purpose selection
Elastic net	●	●	●●	When grouping correlated predictors matters
Horseshoe	●●	●●	●●	Balanced across selection, prediction, and convergence
Reg. horseshoe	●●●	●●●	○	Best selection and predictive performance (needs longer chains)
Spike-and-slab	●●●	●●●	○	Lowest false-selection rate (needs longer chains)

Note. ●●● = strong; ●● = good; ● = moderate; ○ = weak. Reg. horseshoe = regularized horseshoe. Selection Accuracy integrates power for true effects and control of false selections for null effects.

horseshoe recommended when $n < 300$ (for reliable convergence with strong selection).

Condition-Specific Recommendations

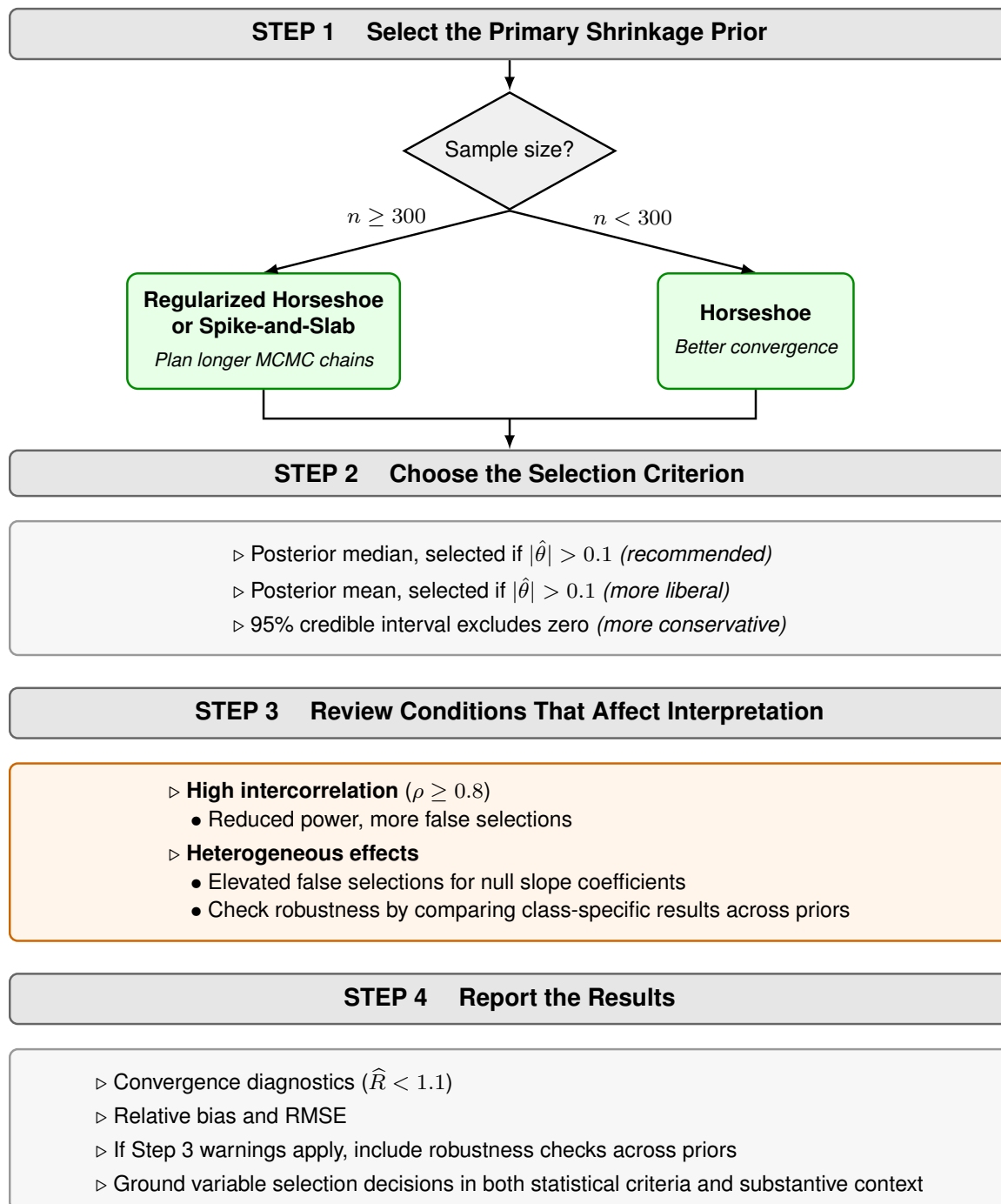
I consolidate the findings into six condition-specific recommendations.

When variable selection is the primary goal with adequate sample size ($n \geq 300$). The regularized horseshoe and spike-and-slab offer the strongest balance between detecting true predictors and controlling false selections. The spike-and-slab achieved the lowest false-selection rates, and the regularized horseshoe provided the best predictive accuracy. Researchers using either prior should plan longer MCMC chains and conduct thorough convergence diagnostics, because convergence rates for these two priors remained substantially below those of the other priors across conditions.

When variable selection is the primary goal with limited sample size ($n < 300$). The horseshoe offers comparable variable selection accuracy with better convergence than the regularized horseshoe and spike-and-slab. Researchers working with small to moderate samples should prefer the horseshoe.

When choosing the selection criterion. Among the selection criteria examined, the posterior median provided the most balanced performance, avoiding the inflated selection rates of the posterior mean and the overly restrictive selection of the 95%

Figure 3.17: Workflow for Selecting a Shrinkage Prior for Variable Selection in Growth Mixture Models.



credible interval at small sample sizes. The choice among these criteria should depend on the researcher's substantive goals and the degree of conservatism desired. Rather than relying on a single criterion, researchers are encouraged to report results under multiple criteria to demonstrate the sensitivity of their findings. The optimal selection criterion for Bayesian variable selection in GMMs remains an open question, as discussed in the Future Research Avenues section.

When predictive accuracy is the primary goal. The regularized horseshoe, spike-and-slab, and horseshoe all achieved superior predictive accuracy, with the regularized horseshoe consistently producing the lowest WAIC and LOOIC values. Researchers prioritizing out-of-sample prediction should adopt the regularized horseshoe when computation allows, or the horseshoe when MCMC convergence is also a concern.

When predictors are highly intercorrelated. Under $\rho = 0.8$, all priors showed reduced power for true effects and inflated false-selection rates for null effects. Convergence rates for the regularized horseshoe and spike-and-slab dropped by roughly 10 to 45 percentage points compared with $\rho = 0.2$. Researchers facing multicollinearity should plan for larger samples, consider reporting results under multiple priors, and anticipate reduced selection accuracy overall.

When class-specific predictor effects are expected. The results showed that the horseshoe, regularized horseshoe, and spike-and-slab exhibited increased false selection for null slope coefficients under the heterogeneous predictor effect pattern. Applied mixture solutions may also involve unequal or extremely small classes (Henson et al., 2007; Tueller & Lubke, 2010), which could further complicate variable selection accuracy. Researchers should compare results across multiple priors and report the full set of selection outcomes to assess the robustness of the findings.

Caveats and Cautions

Applied researchers should bear three caveats in mind when applying shrinkage priors to predictor selection in GMMs.

Overshrinkage of moderate effects. Under shrinkage priors, posterior estimates of substantively large effects can be pulled below the 0.1 selection threshold. This was most pronounced for the intercept growth factor coefficients in this design, where $\beta = 0.3$ (medium effect) routinely showed relative bias exceeding 60% shrinkage at $n = 150$. Researchers should interpret null findings for moderate effects cautiously, especially under small sample sizes, and consider reporting selection outcomes under multiple criteria to avoid dismissing genuine effects as noise.

Sparsity assumption. Shrinkage priors assume sparsity, meaning that only a subset of candidate predictors have truly nonzero effects. This simulation specified 50% sparsity (five of ten predictors truly nonzero). When the underlying model is non-sparse, that is, when most candidate predictors have substantial effects, adaptive shrinkage priors (e.g., the horseshoe family and spike-and-slab) may over-regularize by treating moderate effects as noise and underestimating genuine effects. Applied researchers should consider whether their substantive context supports a sparsity assumption. In non-sparse settings, uniform shrinkage priors such as the ridge can still provide useful regularization by reducing estimation variance, though they do not perform discrete variable selection.

Heterogeneous predictor effects. When active predictor sets differ across latent classes, the horseshoe, regularized horseshoe, and spike-and-slab showed elevated false-selection rates for null slope coefficients, especially at high predictor intercorrelation ($\rho = 0.8$) where these rates increased with sample size rather than decreasing. Applied researchers anticipating class-specific predictor structures should compare results across multiple priors and report the full set of selection outcomes, rather than relying on a single prior whose class-specific sparsity recovery may be unreliable.

3.8.5 Future Research Avenues

This chapter opens several directions for future methodological investigations. One meaningful extension is to investigate the third predictor pathway in which predictors influence the repeated measures directly as time-varying covariates. The regression form for this pathway is determined by the type of outcome variable, making it methodologically distinct from the within-class growth factor regression examined in this chapter. Beyond this pathway, applying the shrinkage prior framework to a broader range of mixture and latent class models would further extend the present work. This could include GMMs with nonlinear growth trajectories, latent class models with covariates, mixture mediation models, and mixture multilevel vector-autoregressive models (e.g., Bibriescas & Whittaker, 2023; Depaoli, 2013; Ernst et al., 2024; Heo, Depaoli, et al., 2024; Vermunt, 2010). Investigating whether the patterns observed here generalize to these more complex specifications would strengthen the practical relevance of the findings.

Note that I reported variable selection results under multiple criteria, and the results were sensitive to this choice. In particular, a threshold of 0.1 (Feng et al., 2017) or the credibility level of intervals (van Erp et al., 2019) may produce different selection outcomes. However, a systematic evaluation of how these cutoff values perform across different model types and data conditions remains lacking. Future research should examine the sensitivity of selection criteria across a range of modeling contexts and develop practical guidelines for how researchers can choose among criteria based on their substantive goals, report results transparently, and integrate variable selection into a broader analytic workflow.

When GMMs or other latent variable models are regularized, the evaluation of model fit and selection becomes an open question. Bayesian model fit and selection indices have recently been an active area of research for latent variable and mixture models (e.g., Asparouhov & Muthén, 2021; Depaoli, Heo, et al., 2026; Depaoli et al., 2023; Garnier-Villarreal & Jorgensen, 2020; Heo, Jia, & Depaoli, 2024; H. Liu, Heo, et al., 2025; Winter & Depaoli, 2022), but these developments have focused on non-regularized settings. Regularization introduces a deliberate tradeoff between bias and variance, and existing fit indices were not designed to account for this tradeoff. Developing and evaluating model fit tools that reflect the tradeoff in regularized models represents a needed direction.

GMMs in applied research frequently involve incomplete data (Bauer, 2007), yet the interaction between shrinkage priors and missing data handling strategies has not been investigated. Scarce but related work by Bain and Shi (2024) demonstrated that the performance of frequentist variable selection techniques such as the lasso and elastic net can be affected by missing data conditions. In their simulation, predictive accuracy was most sensitive to the missing data mechanism, with stronger effects observed under missing at random than missing completely at random conditions. Whether Bayesian shrinkage priors exhibit similar sensitivity to missingness in terms of predictive accuracy and variable selection, and how they interact with existing missing data handling techniques, remain unanswered.

On a broader philosophical level, this chapter adopted a single-model inference perspective that selects predictors given a single model specification. An alternative approach is Bayesian model averaging (BMA). BMA is a multi-model inferential framework that accounts for model uncertainty by combining inferences across all plausible models weighted by their posterior probabilities (Hoeting et al., 1999; Raftery et al., 1997) and has been gaining traction in applied research (e.g., Heo, Simons, & Liu, 2025; Hinne et al., 2020). These two approaches have traditionally been developed and studied separately, but recent work has begun to compare them directly, as in Sekulovski et al. (2025). These endeavors suggest that shrinkage-based and averaging-based strategies address different aspects of model uncertainty and may complement each other for identifying important variables. Investigating how these two frameworks can be integrated or compared within the mixture modeling setting represents a promising direction.

3.8.6 Concluding Remarks

The absence of systematic guidelines for predictor selection in GMMs is not a recent observation, yet it has persisted as an unresolved methodological gap. I proposed Bayesian variable selection via shrinkage priors as a principled approach, one that retains substantively important predictors and regularizes small or null effects toward zero. The work in this chapter contributes to addressing this longstanding gap and offers concrete solutions to enhance the practical utility of GMMs. Beyond the GMM setting, this approach can be extended to other complex latent variable

and mixture models where predictor selection can further enrich substantive interpretation. I am hopeful that Bayesian variable selection will lead to improved modeling practices and encourage informed decisions in both methodological and applied research.

Chapter 4

Bayesian Nonparametric Multiple Imputation via the Chinese Restaurant Process for Growth Mixture Models

4.1 Introduction

4.1.1 Heterogeneity in Longitudinal Change and Growth Mixture Models

One central goal in social and behavioral research is to describe interindividual differences, or similarities, in patterns of intraindividual change (Grimm et al., 2016; Meredith & Tisak, 1990). This goal constitutes one of the fundamental rationales for collecting and analyzing longitudinal data: to study how individuals change over time and whether their change patterns qualitatively differ (Baltes & Nesselroade, 1979). To illustrate, consider children’s academic growth in mathematics. One common approach to modeling intraindividual change is to assume that a single growth trajectory applies to all students in the sample (DiPerna et al., 2007; Heo, Jia, & Depaoli, 2025; Marvin et al., 2023; Wei et al., 2015). However, research has revealed heterogeneity in mathematics achievement trajectories, indicating that students do not uniformly follow a single developmental pattern but instead exhibit qualitatively distinct growth profiles (i.e., subgroups differing in initial levels and rates of change; see, e.g., Heo, Depaoli, et al., 2024; Hong & You, 2012; Kohli & Sullivan, 2019; Mattison et al., 2023). Recognizing such heterogeneity is of both theoretical and practical importance. It enables researchers to characterize the range of developmental outcomes, evaluate whether interventions operate differently across subgroups, and investigate the determinants and consequences of subgroup membership (Henson et al., 2015; Hodis et al., 2011; L. C. Liu et al., 2010; B. Muthén et al., 2002; Petras & Masyn, 2010).

The methodological implication of this heterogeneity is clear. Standard latent growth curve models estimate a single set of growth parameters for the entire sample, resting on the assumption that all individuals are drawn from a single population (Bollen & Curran, 2006; McArdle & Epstein, 1987; Meredith & Tisak, 1990; Ram & Grimm, 2009). When the population comprises qualitatively distinct subgroups with different developmental trajectories, growth parameter estimates are averaged across subgroups, obscuring the very differences that researchers seek to understand (Jung & Wickrama, 2008). Growth mixture models (GMMs; B. Muthén, 2004; B. Muthén & Shedden, 1999) address this limitation by relaxing the single-population assumption and instead positing that the observed sample is drawn from a finite mixture of latent classes, each characterized by a distinct developmental trajectory. Here, each of the identified latent classes is characterized by its own intercept (i.e., initial level), slope (i.e., rate of change), and within-class covariance structure. Since their introduction, GMMs have been widely adopted to identify qualitatively different developmental subgroups across substantive domains (e.g., see a citation search in Bauer, 2007).

4.1.2 Missing Data in Growth Mixture Models

Because virtually all longitudinal data contain missing values (Ibrahim & Molenberghs, 2009; Laird, 1988), longitudinal studies need to address this issue, and GMMs are no exception. In the context of GMMs, Bauer (2007) found that missing data proportions in the applied studies they reviewed ranged from 3% to 44%, with an average of 20%, showing the pervasiveness of missing data in GMM research. Beyond their prevalence, missing data vary in how they arise and manifest in multiple patterns (e.g., intermittent missingness, planned missingness). Among these, attrition, the loss of participants from a study over time (Nicholson et al., 2017; Twisk & de Vente, 2002), is among the most common in longitudinal research. It occurs when participants drop out partway through the study and provide no further data from that point onward. This type of missingness is by no means unusual in longitudinal research and thus has been a target of investigation in methodological studies (Heo, Jia, & Depaoli, 2024, 2025; McNeish & Harring, 2021; B. Muthén et al., 2011; Shi et al., 2021; Vallejo et al., 2019; Winter & Depaoli, 2022).

The presence of missing data, particularly attrition, may compromise the generalizability of findings to the population of interest (Graham, 2009). In the context of GMMs, this means that the observed data may no longer represent the true population heterogeneity, carrying two interrelated ramifications. First, growth factor estimates can be biased because the identified latent classes and the parameter estimates obtained therein may not represent the full population. The results can deliver only a partial picture, or even a misleading understanding, of the phenomena under study. Incomplete trajectories distort within-class growth, shifting estimates of class-specific intercepts and slopes (Bauer, 2007; D. Y. Lee & Harring, 2023). Second, the latent class structure itself can be distorted. In GMM applications, deciding the number of latent classes and interpreting the identified classes is a core task (Heo, Depaoli, et al.,

2024; Jung & Wickrama, 2008; Peugh & Fan, 2012; Ram & Grimm, 2009; Tofighi & Enders, 2007). Attrition obscures and erodes the latent class structure that GMMs are designed to capture, reducing classification accuracy. These consequences underscore that handling missing data in GMMs is a nontrivial methodological concern with direct implications for the validity of substantive conclusions. The importance of addressing missing data in GMMs has been raised repeatedly in previous studies (Bauer, 2007; Depaoli, Heo, et al., 2026; Heo, Depaoli, et al., 2024; D. Y. Lee & Harring, 2023; Lu et al., 2011; B. Muthén, 2002).

Although this importance has been widely recognized, evaluating and comparing strategies for handling missing data in GMMs has received relatively limited attention (but see D. Y. Lee & Harring, 2023; Lu & Zhang, 2014; Lu et al., 2011). The few existing studies have examined methods including listwise deletion, full information maximum likelihood (FIML), Bayesian estimation via Markov chain Monte Carlo (MCMC), and multiple imputation (MI). These methods have also been evaluated in the broader mixture modeling context (see, e.g., Cai et al., 2010; Depaoli et al., 2025; Sterba, 2014, 2016). Despite these contributions, the evaluation of missing data methods for GMMs remains incomplete. In what follows, I situate the existing approaches within a framework of three methodological considerations that together define the requirements for an adequate solution.

The first consideration concerns preserving observed data as much as possible. Researchers would prefer to retain all available information rather than discard incomplete cases. Observed data, even when partially missing, still carry valuable information about individuals' trajectories and their latent class membership. Listwise deletion falls short of this consideration because it removes an entire case as soon as any value is missing. This practice has been shown to produce biased parameter estimates and decreased statistical power (Little, 1992; Little & Zhang, 2011), and in the context of GMMs, to compromise parameter recovery and classification accuracy (D. Y. Lee & Harring, 2023). The remaining methods, FIML, Bayesian estimation, and MI, all satisfy this first requirement by using all available data in the estimation or imputation process.

The second consideration concerns the specification of the number of latent classes when handling missing data, even though the true number of classes is unknown. When missing data are handled directly within the model-fitting process, as in FIML and Bayesian estimation via MCMC, researchers must pre-specify the number of latent classes C before the model is fit. In practice, this number is determined empirically through model comparison (Heo, Depaoli, et al., 2024; Jung & Wickrama, 2008; Ram & Grimm, 2009). Although FIML and Bayesian estimation via MCMC are known to produce valid inferences when their assumptions are met, the extent to which missing data patterns influence the determination of C through model comparison has not been fully established. This potential circularity, wherein incomplete data inform the selection of C , which in turn determines how those incomplete data are modeled, warrants careful attention. Indeed, how missing data affect the determination of the number of latent classes has been called for further

investigation (e.g., Depaoli, Heo, et al., 2026). While FIML and Bayesian estimation via MCMC are powerful tools for handling missing data (Enders, 2022), they may still be sensitive to how missing data affect the class structure to be specified, which remains a potential vulnerability.

The third consideration involves how missing values are imputed and whether the imputation process is compatible with the latent class structure underlying GMMs. Ideally, researchers want to replace missing values with multiple plausible values, rather than a single imputed value, so that uncertainty about missing values is properly propagated into downstream analyses. This motivates the use of multiple imputation (MI; Rubin, 1987), which additionally yields completed data sets that can support follow-up analyses requiring complete data. When MI is applied in the context of GMMs, the imputation model should reflect population heterogeneity. In other words, the imputation process should be compatible with the latent class structure that GMMs are designed to capture. Yet, meeting this requirement is notoriously difficult under standard MI frameworks. As Enders and Gottschall (2011) cautioned, the latent class structure is typically unknown to the imputer and must itself be inferred from the data, and as such preserving group differences in the mean and covariance structure is not possible. Consequently, commonly used approaches such as MI via joint modeling (Asparouhov & Muthén, 2010; Schafer & Yucel, 2002) and MI via fully conditional specification (Raghunathan et al., 2001; van Buuren, 2018; van Buuren & Groothuis-Oudshoorn, 2011) may not always align well with the mixture structure of GMMs.

4.1.3 Bayesian Nonparametric Multiple Imputation and the Present Study

In summary, existing methods for handling missing data in GMMs face a dilemma. Approaches that preserve the full data either require pre-specifying the number of latent classes (FIML, Bayesian estimation via MCMC) or may not fully accommodate the mixture structure in the imputation process (MI via joint modeling, MI via fully conditional specification), while listwise deletion forfeits information entirely. What is needed is a flexible framework that (a) retains all observed data, (b) separates the treatment of missing data from the specification of the latent class structure, mitigating the potential circularity that arises when incomplete data are modeled directly, and (c) replaces missing values with multiple plausible draws from an imputation model that respects the population heterogeneity inherent in mixture models.

Bayesian nonparametric methods provide a principled foundation for such a framework. At the core of these methods is the Dirichlet process (DP; Ferguson, 1973). Unlike standard parametric priors, whose draws are parameter values, draws from the DP are random probability distributions, and these distributions are discrete with probability one. This discreteness implies that multiple observations share the same underlying parameters, and such shared values form natural clusters in the data.

Building on this foundation, Dirichlet process mixture models (DPMMs; Antoniak, 1974) use the DP as a mixing distribution, allowing the number of clusters to be inferred from the data rather than fixed a priori. I employ DPMMs as the basis for multiple imputation in this chapter. Under this approach, latent clusters are first identified in the observed data, and missing values are then drawn from each cluster’s posterior predictive distribution. This directly addresses the dilemma outlined above. The approach preserves all observed data, separates the treatment of missing data from the specification of the latent class structure, and generates imputed values from a model that captures population heterogeneity by allowing the number of mixture components to be inferred from the data.

The DP admits several constructions. In this chapter, I use the Chinese restaurant process construction (CRP; Aldous, 1985). In the CRP, each observation is sequentially assigned either to an existing cluster with probability proportional to its current size or to a new cluster with probability proportional to the concentration parameter. This process is conceptually analogous to latent class assignment in GMMs, where each individual belongs to one of several trajectory subgroups. Crucially, the number of clusters emerges from the data during posterior inference rather than being fixed in advance. I refer to the resulting imputation approach as MI via the Chinese restaurant process.

To address the missing data challenges in GMMs, the goals of this chapter are twofold. First, I propose a Bayesian nonparametric multiple imputation approach using the Chinese restaurant process for handling missing data in GMMs. Second, I conduct a Monte Carlo simulation study to compare the performance of this proposed approach with existing methods under systematically varied simulation conditions. Through this investigation, I offer practical implications for how Bayesian nonparametric multiple imputation can improve inference in growth mixture modeling with incomplete data.

The remainder of this chapter is organized as follows. I begin by delineating the missing data problem within the GMM context. I then review existing approaches to handle missing data in GMM research and motivate the introduction of Bayesian nonparametric methods. After presenting the CRP and describing how it is integrated into the multiple imputation scheme, I detail the design and results of the simulation study. I conclude with a discussion of the findings, practical implications, and suggestions for future research.

4.2 Missing Data Issues in Growth Mixture Models

The ability of GMMs to estimate growth parameters and recover latent classes of growth trajectories depends on having reasonably complete observed data. In reality, missing data are nearly unavoidable in longitudinal research, and likewise in GMM research (e.g., Bauer, 2007). Several methods have been evaluated for han-

dling missing data in the mixture modeling context, and their performance should be understood in light of the missing data mechanism assumed. I therefore begin by reviewing missing data mechanisms before turning to the literature on existing approaches.

4.2.1 Missing Data Mechanisms

The statistical properties of any missing data method depend on the mechanism that generated the missing values. Rubin (1976) classified the probability of each data point being missing into three categories, a framework that has since become the de facto standard in the missing data literature (see also Little & Rubin, 1987). This framework characterizes the relationship between the probability of missingness and the data themselves by partitioning the hypothetically complete data into observed components, $\mathbf{Y}_{\text{obs}}$, and missing components, $\mathbf{Y}_{\text{mis}}$. A missing data indicator $\mathbf{R}$ is then defined, where $R = 1$ if a value is missing and $R = 0$ otherwise.

Data are said to be missing completely at random (MCAR) when the probability of a value being missing is unrelated to both the observed and missing parts of the data:

$$\Pr(\mathbf{R} = 1 \mid \mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}}, \boldsymbol{\phi}) = \Pr(\mathbf{R} = 1 \mid \boldsymbol{\phi}), \quad (4.1)$$

where $\boldsymbol{\phi}$ denotes the parameters governing the missingness process (also known as the missing data model; see van Buuren, 2018). Under MCAR, the observed data constitute a simple random sample of the complete data, and all participants share the same probability of missing values regardless of their scores (Enders, 2022; van Buuren, 2018). Although convenient, this assumption is rarely plausible because it rules out any systematic relationship between missingness and the variables under study.

A more realistic assumption is missing at random (MAR), which allows the probability of missingness to depend on the observed data but not on the missing values themselves:

$$\Pr(\mathbf{R} = 1 \mid \mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}}, \boldsymbol{\phi}) = \Pr(\mathbf{R} = 1 \mid \mathbf{Y}_{\text{obs}}, \boldsymbol{\phi}). \quad (4.2)$$

Despite its name, MAR does not imply that missingness is haphazard. Rather, it means that after conditioning on the observed data, the missing values carry no additional information about the missingness process (Enders, 2022; Schafer & Graham, 2002). In a longitudinal setting, MAR implies that an individual's propensity to drop out at a given wave can be predicted from their observed data, such as previously measured scores or other covariates, but not from the unobserved scores they would have provided had they remained in the study (see Peeters et al., 2015, with real data example and justification). MCAR and MAR are collectively referred to as ignorable missingness because modern missing data handling methods such as FIML and MI do not require modeling the missingness process governed by $\boldsymbol{\phi}$ (Enders, 2022; Rubin, 1976; Schafer & Graham, 2002; van Buuren, 2018). That is, these methods can

handle ignorable missing data without explicitly specifying why values are missing (Collins & Lanza, 2010).

When the probability of missingness depends on the unobserved values themselves, even after conditioning on the observed data, the mechanism is classified as missing not at random (MNAR):

$$\Pr(\mathbf{R} = 1 \mid \mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}}, \boldsymbol{\phi}). \quad (4.3)$$

MNAR is the most problematic mechanism because there is a systematic relationship between the unobserved values and the probability of being missing, making it particularly difficult to adjust for this dependency. As a result, standard analytic methods may produce biased parameter estimates (Collins & Lanza, 2010; Enders, 2022; Graham, 2009; Kristman et al., 2004). For this reason, MNAR is also referred to as nonignorable missingness.

Among the three mechanisms, MAR is widely considered a reasonable assumption for longitudinal studies and has been routinely adopted in methodological studies examining longitudinal missing data (e.g., Heo, Jia, & Depaoli, 2024, 2025; D. Y. Lee & Harring, 2023; Shi et al., 2021; Winter & Depaoli, 2022).

4.2.2 Review of Existing Approaches

Several studies have evaluated the performance of missing data methods in the context of GMMs and related mixture models. The methods examined include listwise deletion, FIML, Bayesian estimation via MCMC, and MI, assessed under both ignorable and nonignorable missing data conditions. Below, I review these approaches and summarize the key findings from the available evidence.

Listwise Deletion

The most straightforward response to missing data is to exclude any case with one or more missing values from the analysis. This approach has the advantage of simplicity, as standard estimation procedures can be applied directly once complete cases are selected. However, listwise deletion discards observed measurements, resulting in reduced statistical power and estimation efficiency (Little, 1992; Little & Zhang, 2011). Although listwise deletion produces unbiased parameter estimates under MCAR, the resulting standard errors and significance levels, while correct for the reduced subset of data, tend to be larger than those that would be obtained using all available data (van Buuren, 2018). Under MAR, the complete cases are no longer representative of the original sample, and parameter estimates can be systematically biased (Collins & Lanza, 2010; Enders, 2022; Little & Rubin, 1987).

The empirical evidence for listwise deletion in GMMs is correspondingly unfavorable. In the simulation study by D. Y. Lee and Harring (2023), which evaluated missing data methods for a two-class linear GMM, listwise deletion produced the lowest convergence rates among all methods, falling as low as 28% when the sample size

was 100. Even at the largest sample size of 1,000, convergence for listwise deletion reached only 67%, remaining below that of competing approaches. The overall absolute relative bias was also the largest for the listwise deletion method, incurring bias for all parameters examined (including slope means and growth factor variances and covariance). Moreover, listwise deletion had a minimal positive effect on classification accuracy, offering no compensating advantage for the information it discarded. In the broader mixture modeling context, Depaoli et al. (2025) found that listwise deletion consistently biased the intercept of the structural model under MAR conditions in latent class analysis with missing covariates, with the bias increasing as the percentage of missing data grew. These findings collectively underscore the limitations of listwise deletion for mixture models.

Full Information Maximum Likelihood

Full information maximum likelihood (FIML) addresses missing data by maximizing the likelihood function computed over each individual's available data rather than restricting estimation to complete cases (Arbuckle, 1996). Rather than discarding incomplete cases, FIML evaluates each individual's contribution to the likelihood using only the observed variables for that individual, and the overall log-likelihood is the sum of these case-specific contributions. The maximization is typically carried out via the expectation-maximization (EM) algorithm (Dempster et al., 1977). Under the MAR and MCAR assumptions, this approach yields consistent and asymptotically efficient estimates without the need to explicitly model the missingness mechanism (Enders, 2022; Little & Rubin, 1987).

Empirical evidence on the performance of FIML in GMMs and related mixture models has varied across study designs and parameter types. D. Y. Lee and Harring (2023) examined this performance in a simulation using a two-class linear GMM with four time points, equal class proportions, and sample sizes ranging from 100 to 1,000. They found that FIML produced relatively low convergence rates compared to Bayesian and imputation-based methods, with rates as low as 39% at a sample size of 100 and reaching only 75% at 1,000. When FIML did converge, the variance and covariance parameters were underestimated, and the slope means showed the most severe bias among methods examined. Classification accuracy also remained limited, reaching only 60% even under high class separation. In contrast, Depaoli et al. (2025) found that FIML consistently produced unbiased parameter estimates in latent class analysis with missing covariates. This discrepancy suggests that FIML may perform differently depending on the complexity of the mixture model and the type of parameters being estimated.

Bayesian Estimation via Markov Chain Monte Carlo

Bayesian estimation provides an alternative framework for handling missing data. As introduced in Chapter 2, Bayesian inference combines prior distributions with a likelihood to obtain a posterior distribution, which is typically approximated

via Markov chain Monte Carlo (MCMC) algorithms. In the context of missing data, Bayesian estimation treats the missing values as additional unknown quantities to be estimated alongside the model parameters. Specifically, the observed data are supplemented with draws of the missing values, which are treated as unknown parameters, a concept referred to as data augmentation (Tanner & Wong, 1987). As Enders (2022) described, Bayesian estimation can be viewed as a bridge between FIML and MI. Like FIML, the primary focus is on estimating parameters of interest, but like MI, the procedure generates plausible values for the missing data as part of the estimation process. To avoid confusion with the Bayesian nonparametric approach introduced in the next section, I explicitly refer to the method discussed here as Bayesian estimation via MCMC.

In handling missing data in GMMs, Bayesian estimation via MCMC has demonstrated the most favorable performance profile. In the same simulation by D. Y. Lee and Harring (2023), noninformative priors were used for both the growth factor means and the growth factor covariance matrix. Under ignorable missing data conditions, Bayesian estimation achieved the highest convergence rates across all sample sizes, including at the smallest sample size of 100 (reaching 90%). The method also produced the least biased parameter estimates overall and maintained classification accuracy rates higher than competing methods. Beyond the ignorable missing data setting, Lu et al. (2011) proposed a Bayesian method for GMMs with latent class dependent missingness, a form of MNAR in which the probability of dropout depends on latent class membership. Their method demonstrated accurate parameter recovery under both MCAR and MNAR conditions. Lu and Zhang (2014) extended this work by introducing robust models using t -distributions to handle outliers alongside four types of non-ignorable missingness, showing that the Bayesian method recovers parameters well when both the distributional and missingness models are correctly specified.

In the broader mixture modeling context, Cai et al. (2010) developed a Bayesian approach for mixture structural equation models with non-ignorable missing responses and covariates. Their simulation results showed that the Bayesian method produced accurate parameter estimates with small bias and that model selection correctly identified the true number of mixture components. In a related context, Depaoli et al. (2025) found that when covariates contain missing values in latent class analysis, Bayesian estimation could help but was sensitive to prior specification. According to Depaoli et al. (2025), informative priors centered on the correct population values produced the most favorable parameter recovery results, whereas priors centered on incorrect values produced the most biased results.

Multiple Imputation

Multiple imputation (MI; Rubin, 1987) takes a fundamentally different approach to missing data. Whereas FIML and Bayesian estimation via MCMC handle missing values behind the scenes en route to estimating model parameters, MI fills in the incomplete data first and creates multiple completed data sets suitable for

later analysis (Enders, 2022). The standard MI procedure consists of three stages. In the imputation stage, an imputation model is used to generate M completed data sets, each containing different plausible draws for the missing values. In the analysis stage, the substantive model is fit separately to each of the M completed data sets, yielding M sets of parameter estimates and standard errors. In the pooling stage, these results are combined using Rubin’s rules (Little & Rubin, 1987; Rubin, 1987) to produce a single set of point estimates and standard errors that appropriately reflect both within-imputation sampling variability and the between-imputation uncertainty due to missing data.

Applying MI to mixture models raises particular concerns. Because the number of latent classes and the nature of class-varying relations are unknown a priori, predictor–outcome relations can differ across latent classes, and an imputation model that ignores this heterogeneity can produce biased parameter estimates (Enders & Gottschall, 2011; Sterba, 2014). Beyond parameter recovery, Sterba (2016) demonstrated that the choice of missing data handling method can also affect model selection. Under conditions of moderate class separation and larger proportions of missingness, imputing from a saturated multivariate normal model reduced the probability of correctly identifying the true latent class model relative to FIML. These findings suggest that standard MI schemes may fail to preserve the between-class differences that mixture models are designed to detect. These concerns manifest differently depending on how the imputation model is specified. Below, I review the two major MI frameworks and a two-stage extension.

Joint Model Imputation. Multiple imputation via joint modeling (MI-JM; Schafer, 1997) specifies a single multivariate distribution, typically multivariate normal, for the incomplete data and draws imputations from the corresponding posterior predictive distribution. The core algorithm is the Gibbs sampler, which iterates between two steps: a posterior step (P-step) that draws new values of the imputation model parameters given the current complete data, and an imputation step (I-step) that draws new values of the missing data from the predictive distribution given the current parameter estimates (Enders, 2022; Schafer, 1997). After a sufficient number of iterations, completed data sets are saved at regular intervals to form the M multiply imputed data sets (Asparouhov & Muthén, 2010). In handling missing data in GMMs, D. Y. Lee and Haring (2023) found that MI using joint modeling produced parameter estimates with relative bias exceeding the 10% acceptability threshold for slope means and variances, performing comparably to listwise deletion and FIML. Correspondingly, classification accuracy remained limited.

Fully Conditional Specification. Multiple imputation via fully conditional specification (MI-FCS; Raghunathan et al., 2001; van Buuren, 2018; van Buuren & Groothuis-Oudshoorn, 2011), also known as chained equations imputation, takes a different approach by imputing missing data on a variable-by-variable basis. Rather than defining a single joint distribution, MI-FCS decomposes the multivariate missing

data problem into a series of univariate conditional models, one for each incomplete variable given all other variables. The same P-step and I-step logic described for joint modeling applies to each variable in turn, with the algorithm iteratively cycling through the incomplete variables until the imputations stabilize (van Buuren, 2018). This approach accommodates mixed data types, as each variable can be imputed via an appropriate regression model (e.g., linear regression for continuous variables, logistic regression for binary variables). Among the available imputation models, predictive mean matching (Little, 1988) is one of the most commonly used. Because it draws imputed values from observed donors rather than from a parametric distribution, predictive mean matching preserves the distributional characteristics of the observed data (van Buuren, 2018). MI-FCS has not yet been directly evaluated in the context of GMMs or, more broadly, mixture modeling with missing data. This chapter includes MI-FCS as one of the methods compared.

Two-Stage Multiple Imputation. Recognizing the challenge of applying standard MI to models with latent grouping variables, Harel (2007) proposed a two-stage MI approach that decomposes the missing data problem into two steps. In the first stage, M sets of the latent class indicators are drawn from the posterior predictive distribution of the class memberships given the observed data. In the second stage, conditional on each set of imputed class memberships, additional sets of the remaining missing data are filled in. The resulting completed data sets are analyzed separately, and the estimates are combined using a modified version of Rubin’s rules that accounts for the two-stage structure (Harel, 2007; Shen, 2000). In D. Y. Lee and Harring (2023), the two-stage approach produced parameter estimates that were among the least biased, approaching the accuracy of Bayesian estimation via MCMC. However, classification accuracy remained low even under high class separation. Although this approach represents an attempt to incorporate latent class structure into the imputation process, its practical implementation is more involved than standard MI (e.g., MI-JM and MI-FCS) and requires additional steps outside standard software workflows (D. Y. Lee & Harring, 2023).

4.2.3 Summary of the Current Landscape

Across the reviewed methods, each addresses some but not all of the three considerations outlined in the Introduction. Listwise deletion discards observed data. FIML and Bayesian estimation via MCMC require C to be specified before model fitting. Standard MI (MI-JM and MI-FCS) preserves data and separates missing data handling from class enumeration but does not account for the latent class structure during imputation. Among MI approaches, the two-stage MI method comes closest by attempting to incorporate latent classes, yet it still requires C to be specified before class membership draws. These limitations motivate the Bayesian nonparametric approach developed in the following section.

4.3 Bayesian Nonparametric Multiple Imputation

I now introduce a Bayesian nonparametric approach to multiple imputation that simultaneously addresses the limitations identified above. By placing a prior over an unbounded collection of mixture components, the approach allows missing data to be handled independently of the latent class structure while generating multiply imputed data sets from an imputation model that accounts for population heterogeneity.

4.3.1 From the Dirichlet Distribution to the Dirichlet Process

To motivate the Bayesian nonparametric approach, it is helpful to begin with the finite-dimensional Dirichlet distribution and then describe how it generalizes to an infinite-dimensional prior over probability distributions. In a standard finite mixture model with K components, the mixing proportions $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)$ are commonly assigned a symmetric Dirichlet prior,

$$\boldsymbol{\pi} \sim \text{Dirichlet}\left(\frac{\alpha}{K}, \dots, \frac{\alpha}{K}\right), \quad (4.4)$$

where $\alpha > 0$ is a concentration parameter that controls how uniformly the mixing weights are distributed across components. When α/K is large, the prior favors roughly equal weights across components; when it is small, the prior concentrates mass on a few dominant components, with the rest remaining nearly empty. Researchers often set K to a large value, expecting this to make the prior more flexible about the number of components. However, this expectation does not always hold. The prior's behavior depends not on K alone but on the ratio α/K . When K is set to a large value without a corresponding adjustment of α , the resulting model can become unstable, with some components absorbing most of the data and others remaining nearly empty (Frühwirth-Schnatter & Malsiner-Walli, 2019; Green & Richardson, 2001; Rousseau & Mengersen, 2011). Moreover, this formulation still requires researchers to fix K in advance, a constraint that the Dirichlet process relaxes by extending the mixture framework to an unbounded number of components.

The Dirichlet process (DP; Ferguson, 1973) extends the Dirichlet distribution to a prior over the space of probability distributions. Conceptually, the DP can be understood as the limiting case of the finite Dirichlet-multinomial mixture model as $K \rightarrow \infty$ (Gershman & Blei, 2012; Hjort et al., 2010). Rather than allocating probability mass across a fixed number of components, the DP places a prior over an unbounded collection of potential components, allowing the number of occupied clusters to be determined by the data during posterior inference. A random distribution G is said to follow a Dirichlet process, written $G \sim \text{DP}(\alpha, G_0)$, if for every finite measurable partition $(A_1, \dots, A_r)$ of the sample space, the vector $(G(A_1), \dots, G(A_r))$ follows a Dirichlet distribution with parameters $(\alpha G_0(A_1), \dots, \alpha G_0(A_r))$ (Ferguson, 1973). Here, $\alpha > 0$ is the concentration parameter and G_0 is the base distribution,

which represents the prior expectation of G in the sense that $E[G(\cdot)] = G_0(\cdot)$. A key property of the DP is that it generates a discrete distribution with probability one (Ferguson, 1973), meaning that draws from G will take repeated values even when G_0 is continuous. This discreteness is what gives rise to the clustering behavior that makes the DP useful as a mixture model prior.

4.3.2 Dirichlet Process Mixture Models

A Dirichlet process mixture model (DPMM; Antoniak, 1974) uses the DP as a prior on the mixing distribution of a mixture model. Because the DP does not restrict the number of mixture components in advance, the number and nature of components are inferred from the data rather than fixed by the researcher. Consider a data set $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)$ where each observation $\mathbf{y}_i$ is associated with a cluster-specific parameter $\boldsymbol{\theta}_i$ drawn from a random mixing distribution G :

$$\mathbf{y}_i \mid \boldsymbol{\theta}_i \sim F(\boldsymbol{\theta}_i), \quad \boldsymbol{\theta}_i \mid G \sim G, \quad G \sim \text{DP}(\alpha, G_0), \quad (4.5)$$

where $F(\boldsymbol{\theta}_i)$ is a parametric kernel (e.g., a multivariate normal distribution) and G_0 is the base distribution that serves as the prior for the parameters of each cluster (Gershman & Blei, 2012; Hjort et al., 2010). Because G is discrete with probability one, multiple observations will share the same parameter value $\boldsymbol{\theta}_i$, naturally inducing a partition of the data into clusters. Unlike a finite mixture model in which K must be fixed before estimation, the DPMM allows new clusters to emerge as the posterior is updated with data.

The concentration parameter α governs model complexity. Small values of α favor partitions with few large clusters, whereas large values encourage many small clusters. Importantly, the expected number of distinct clusters grows logarithmically with the sample size, $E[K] \approx \alpha \log(1 + N/\alpha)$, providing an automatic form of complexity control that adapts to the amount of available data (Antoniak, 1974). The base distribution G_0 specifies the prior for the parameters of each newly created cluster. A common specification is to place independent priors on each cluster-specific parameter. For example, a normal prior for the mean and an inverse-gamma prior for the variance (see, e.g., Qiu et al., 2025).

4.3.3 The Chinese Restaurant Process

The DP admits several equivalent representations, each based on a different construction. One widely used construction is the Chinese restaurant process (CRP; Aldous, 1985), which describes how observations are sequentially assigned to clusters. The CRP metaphor envisions a restaurant with an infinite number of tables. Customers (observations) enter one at a time and either sit at an occupied table (join an existing cluster) or start a new table (create a new cluster). The seating probabilities for the $(n+1)$ th customer, given the seating arrangement of the previous n customers,

are:

$$\Pr(c_{n+1} = k \mid c_1, \dots, c_n) = \begin{cases} \frac{n_k}{n + \alpha} & \text{for an existing cluster } k \text{ with } n_k \text{ members,} \\ \frac{\alpha}{n + \alpha} & \text{for a new cluster,} \end{cases} \quad (4.6)$$

where c_{n+1} denotes the cluster assignment for observation $n + 1$ and n_k is the number of observations currently in cluster k . This formulation exhibits a “rich get richer” property. Larger clusters attract new members with higher probability, which naturally produces a small number of well-populated clusters alongside a tail of smaller ones.

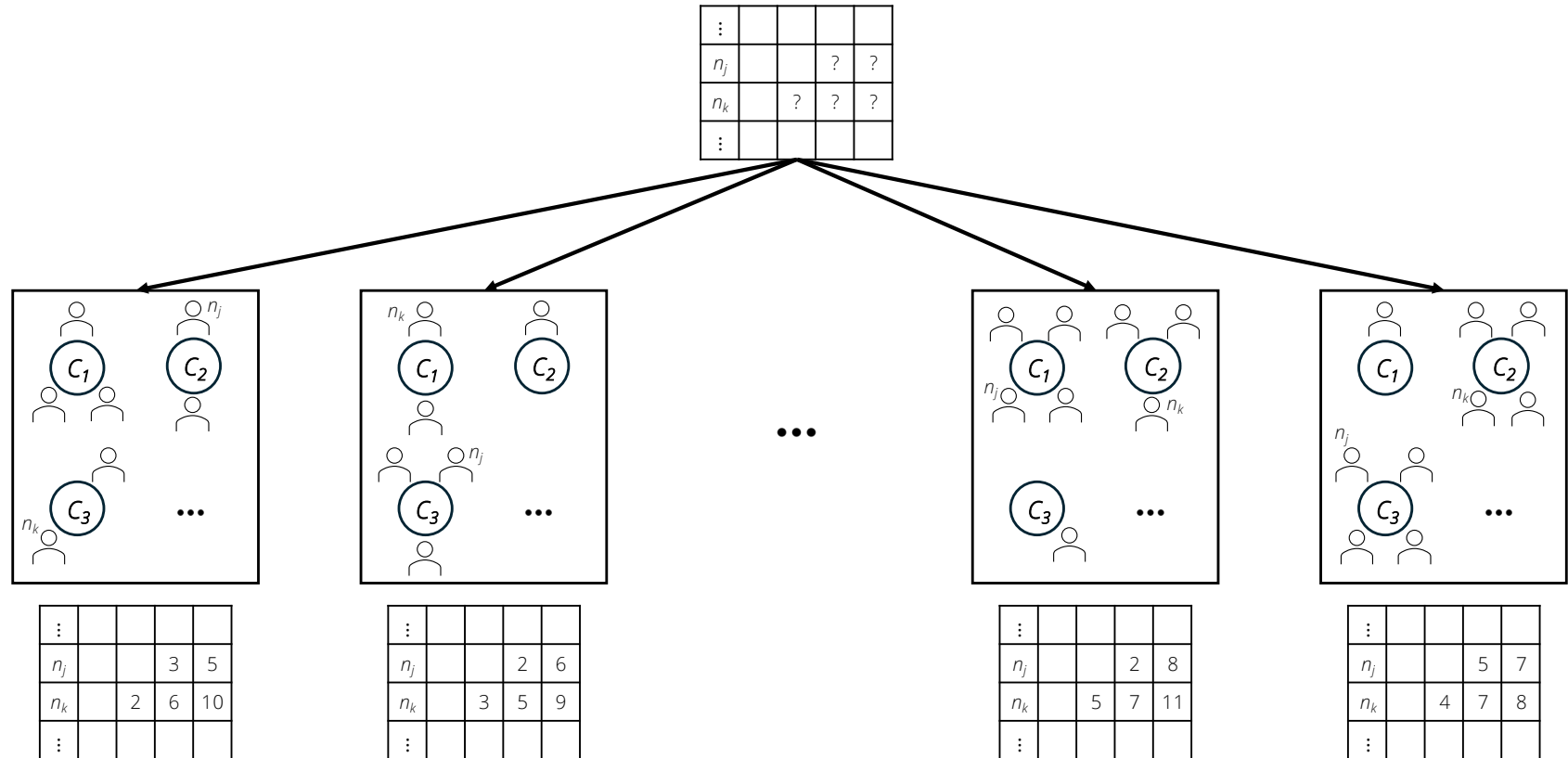
The CRP representation is particularly well suited as the basis for a missing data imputation engine in the GMM context for two reasons. First, the sequential cluster assignment in the CRP is conceptually parallel to latent class assignment in GMMs. Each individual is allocated to one of several subgroups, and the number and composition of those subgroups are determined by the data. Second, the number of clusters is not fixed in advance but emerges during posterior inference, directly addressing the concern that pre-specifying K in the presence of missing data may lead to a class structure that reflects artifacts of the missingness pattern rather than genuine population heterogeneity. By allowing the imputation model to discover the latent cluster structure in the data, the CRP-based approach can generate imputed values that respect the boundaries between subgroups. The resulting completed data sets are therefore more likely to preserve the heterogeneity that the downstream GMM analysis is designed to detect.

4.3.4 Multiple Imputation via the Chinese Restaurant Process

Combining the CRP with the MI framework yields an imputation procedure that I refer to as MI via the Chinese restaurant process (MI-CRP). At a conceptual level, in each imputation round, a DPMM based on the CRP is fitted to the observed data. Through this process, individuals are probabilistically assigned to clusters that represent distinct growth trajectories. The number of clusters is not fixed in advance and can grow as needed to reflect population heterogeneity. Given these cluster assignments, missing values are imputed using predictive draws from the cluster-specific growth trajectories learned from the data. This results in a completed data set, and the entire process is repeated multiple times to generate multiple imputed data sets. A conceptual illustration of this procedure is presented in Figure 4.1.

An important feature of this procedure is that cluster assignments need not be consistent across imputation rounds. What matters is not that a given individual is assigned to the same cluster label in every round, but that within each round, individuals with similar observed trajectories are grouped together. Because imputed values are drawn from the cluster-specific distributions learned within each round,

Figure 4.1: Conceptual illustration of the MI-CRP procedure. The top panel shows an incomplete data set in which individuals n_j and n_k have missing values (denoted by “?”) at later time points due to attrition. The middle panel depicts multiple imputation rounds, where the CRP assigns individuals to latent clusters ($C_1, C_2, C_3, \dots$). The number of clusters is not fixed in advance and may vary across rounds, as indicated by the ellipsis. Cluster assignments differ across imputation rounds, reflecting the stochastic nature of the CRP. The bottom panel shows the resulting completed data sets, in which missing values have been replaced by draws from the cluster-specific posterior predictive distributions. Each completed data set contains different imputed values, propagating the uncertainty due to missing data into downstream analyses.



between-group differences are preserved in each completed data set. The issue of aligning class labels across the M completed data sets for downstream GMM analysis is a separate concern and is addressed at the analysis stage.

In detail, the algorithm proceeds by fitting a DPMM to the observed data via MCMC, using the CRP to perform cluster assignments at each iteration. At each MCMC step, the algorithm cycles through two operations: (1) reassign each observation to an existing or new cluster according to the CRP probabilities in Equation 4.6, updated to reflect the likelihood of the observation under each cluster’s parameters; and (2) draw missing values for each observation from the posterior predictive distribution of its assigned cluster.

More concretely, at MCMC iteration s , the cluster assignment for observation i with observed data $\mathbf{y}_{i,\text{obs}}$ is updated according to:

$$p(c_i = k \mid \mathbf{y}_{i,\text{obs}}, \boldsymbol{\theta}_k^{(s)}, \dots) \propto \begin{cases} n_{-i,k} \cdot f(\mathbf{y}_{i,\text{obs}} \mid \boldsymbol{\theta}_k^{(s)}) & \text{for existing cluster } k, \\ \alpha \cdot \int f(\mathbf{y}_{i,\text{obs}} \mid \boldsymbol{\theta}) dG_0(\boldsymbol{\theta}) & \text{for a new cluster,} \end{cases} \quad (4.7)$$

where $n_{-i,k}$ is the number of observations in cluster k excluding observation i , and $\boldsymbol{\theta}_k^{(s)}$ denotes the current parameter estimates for cluster k . Once all cluster assignments and parameters have been updated, missing values for observation i assigned to cluster k are drawn from the cluster-specific posterior predictive distribution:

$$\mathbf{y}_{i,\text{mis}}^{(s)} \sim p(\mathbf{y}_{i,\text{mis}} \mid \mathbf{y}_{i,\text{obs}}, \boldsymbol{\theta}_k^{(s)}). \quad (4.8)$$

After a suitable burn-in period, M completed data sets are saved at regular intervals from the MCMC chain. Each completed data set is then submitted to the substantive GMM analysis, and the resulting estimates are pooled using Rubin’s rules (Little & Rubin, 1987; Rubin, 1987).

Bayesian nonparametric imputation models have shown promise in related settings. Si and Reiter (2013) used a Dirichlet process mixture of products of multinomials to multiply impute incomplete categorical variables in large-scale assessment surveys. Their approach did not require the number of mixture components to be fixed and allowed the model to capture complex dependencies among variables. In simulation, the approach produced point estimates closer to the true values and confidence interval coverage rates closer to the nominal level than standard chained equations imputation. More recently, Qiu et al. (2025) introduced a Bayesian nonparametric approach to latent class analysis using the Dirichlet process, demonstrating that the CRP-based model correctly identified the true number of classes across a range of conditions with accurate parameter recovery. In both settings, Bayesian nonparametric models accommodate latent heterogeneity without requiring the number of subgroups to be pre-specified. In this chapter, I extend this line of work to the problem of missing data in GMMs, using the CRP as the basis for generating multiply imputed data sets that respect the mixture structure underlying the substantive analysis model.

The proposed MI-CRP approach has the potential to address the limitations of existing methods by decoupling the treatment of missing data from the specification of the latent class structure while generating imputed values that respect

population heterogeneity. At the same time, no study has examined how a Bayesian nonparametric imputation approach performs in the context of GMMs or, more broadly, mixture modeling with missing data. My goal is to evaluate this capability through a Monte Carlo simulation study comparing MI-CRP with FIML, Bayesian estimation via MCMC, and standard MI methods (MI-JM and MI-FCS). Details of the simulation design are provided next.

4.4 Simulation Design

I conducted a Monte Carlo simulation study to evaluate the performance of the proposed MI-CRP approach relative to existing missing data methods for GMMs. The simulation design was a fully crossed factorial with four data-generation factors: sample size ($n = 150, 300, 900$), class separation (Mahalanobis distance of 2.0, 3.0, and 4.0), class proportion (equal at 50:50 and unequal at 70:30), and missing data proportion (10% and 40% attrition). This yielded $3 \times 3 \times 2 \times 2 = 36$ conditions. Each condition included 500 replications, a number deemed sufficient for this simulation, resulting in 18,000 simulated data sets. Each simulated data set was then analyzed with five missing data methods: FIML, Bayesian estimation via MCMC (hereafter Bayesian-MCMC), MI-JM, MI-FCS, and MI-CRP, yielding 90,000 total analyses. Listwise deletion was not considered because it has consistently been shown to perform poorly in the GMM context (D. Y. Lee & Harring, 2023). A summary of the simulation design factors and their levels is presented in Table 4.1. Details of the population model, simulation conditions, analytic procedures, and outcome measures are described below.

Table 4.1: Summary of Simulation Design.

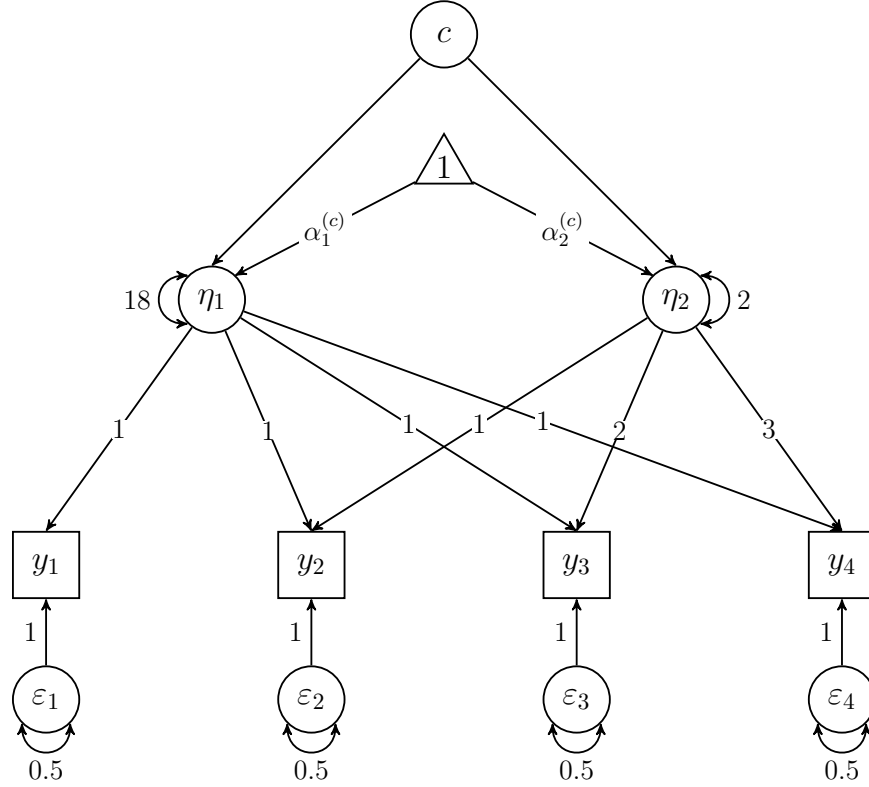
Factor	Levels
Sample size	$n = 150$ (small), 300 (medium), 900 (large)
Class proportion	Equal (50:50), Unequal (70:30)
Class separation	MD = 2.0 (low), 3.0 (moderate), 4.0 (high) Class 1: $\alpha_1^{(1)} = 48$, $\alpha_2^{(1)} = 3$ (fixed across conditions) Class 2: $\alpha_1^{(2)} = 47$, $\alpha_2^{(2)} = 0.191$ (MD = 2.0) $\alpha_1^{(2)} = 39$, $\alpha_2^{(2)} = 0$ (MD = 3.0) $\alpha_1^{(2)} = 33.613$, $\alpha_2^{(2)} = 0$ (MD = 4.0)
Missing data proportion	10% (small), 40% (large)

Note. MD = Mahalanobis distance.

4.4.1 Population Model

The population model was a two-class ($C = 2$) linear GMM with four equally spaced time points ($T = 4$) and two growth factors ($m = 2$): an intercept and a linear slope. The factor loading matrix was specified as $\Lambda_y = [\mathbf{1}, (0, 1, 2, 3)^\top]$, and a path diagram of the population model is presented in Figure 4.2. The growth factor means for Class 1 were fixed at $\alpha_1 = (48, 3)^\top$, representing an intercept of 48 and a linear slope of 3. The growth factor means for Class 2 (α_2) varied across conditions as a function of class separation (described in the Class Separation subsection). The within-class growth factor covariance matrix and time-specific residual variance were set to $\Psi_c = \text{diag}(18, 2)$ and $\Theta_c = 0.5\mathbf{I}_4$, respectively, and held equal across classes. Similar population models have been used in previous GMM simulation studies (Depaoli, 2013; Depaoli, Heo, et al., 2026; S. Kim et al., 2022) and in studies addressing missing data in GMMs (D. Y. Lee & Harring, 2023; Lu et al., 2011). The population values were determined based on Depaoli (2013, 2014) and Kaplan (2002).

Figure 4.2: Path diagram of the two-class linear growth mixture model used as the population model.



4.4.2 Sample Size

Sample size has been a critical factor in the evaluation of missing data methods for GMMs (D. Y. Lee & Harring, 2023; Lu & Zhang, 2014; Lu et al., 2011). I

examined three levels of total sample size: $n = 150$ (small), $n = 300$ (medium), and $n = 900$ (large). These values cover sample sizes considered in previous GMM literature (Depaoli, 2013, 2014; Depaoli, Heo, et al., 2026; S. Kim et al., 2021, 2022) and, particularly, missing data literature evaluating MI (Depaoli et al., 2025; Jia & Wu, 2019, 2023; Sterba, 2014).

4.4.3 Class Proportion

Beyond sample size, the relative sizes of latent classes can influence model performance in mixture modeling (Enders & Tofghi, 2008; Tueller & Lubke, 2010). Two conditions of class proportions were examined: equal ($\pi_1 = \pi_2 = .50$) and unequal ($\pi_1 = .70$, $\pi_2 = .30$). I considered the unequal condition to reflect scenarios involving a relative minority class, which is a setting of particular relevance given previous GMM simulation research with two-class models (Depaoli, Heo, et al., 2026; M. Li & Harring, 2017; Lu et al., 2011).

4.4.4 Class Separation

The degree to which latent classes are distinguishable, or class separation, can affect the performance of missing data methods in GMMs (D. Y. Lee & Harring, 2023). To quantify class separation, I used the multivariate Mahalanobis distance (MD), defined as

$$MD = [(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)^\top \boldsymbol{\Psi}^{-1}(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)]^{1/2}. \quad (4.9)$$

Three levels of class separation were considered: MD = 2.0 (low), MD = 3.0 (moderate), and MD = 4.0 (high). These values cover the range of class separation conditions examined in previous research (Depaoli, Heo, et al., 2026; S. Kim et al., 2021, 2022; D. Y. Lee & Harring, 2023). Across these levels, the growth factor means for Class 1 were fixed at $\boldsymbol{\alpha}_1 = (48, 3)^\top$, while the Class 2 means ($\boldsymbol{\alpha}_2$) were adjusted to achieve each target MD: $\boldsymbol{\alpha}_2 = (47, 0.191)^\top$ for MD = 2.0, $\boldsymbol{\alpha}_2 = (39, 0)^\top$ for MD = 3.0, and $\boldsymbol{\alpha}_2 = (33.613, 0)^\top$ for MD = 4.0. These values were computed using the common covariance matrix $\boldsymbol{\Psi} = \text{diag}(18, 2)$ specified in the population model.

4.4.5 Missing Data Proportion

Missing data were generated under a missing at random (MAR) mechanism at two levels of missing data proportion: 10% and 40% missingness at the final time point (y_4). These two levels represent small and large degrees of missingness, respectively, consistent with the amount of missing data reported in applied GMM research (Bauer, 2007). These levels also fall within the range of missing data proportions examined in GMM simulation studies using both attrition and non-attrition patterns (D. Y. Lee & Harring, 2023; Lu & Zhang, 2014; McNeish & Harring, 2021). Details of the missing data generation procedure are provided in the following paragraph.

Generating Attrition-Pattern Missing Data

When considering attrition patterns in simulation research, several different forms are feasible (Heo, Jia, & Depaoli, 2024, 2025; Ortega-Azurdoy et al., 2008). In this chapter, I adopted a linear, monotone attrition pattern in which the cumulative proportion of missing data increases at a constant rate across time points. Under the 10% attrition condition, the cumulative missing proportions were 0%, 3.33%, 6.67%, and 10% at y_1 through y_4 , respectively; under the 40% attrition condition, the corresponding proportions were 0%, 13.33%, 26.67%, and 40%.

To satisfy the MAR assumption, the probability of dropout at each time point was modeled as a function of the first observed score using logistic regression (Agresti, 2012):

$$\text{logit}[p(y_{it} \text{ is missing})] = b_{0t} + b_1 z_{i1}, \quad (4.10)$$

where z_{i1} is the within-class standardized value of y_{i1} , b_{0t} is the time-specific intercept, and b_1 is a regression weight. The regression weight was fixed at $b_1 = 1.48$ to establish a strong association between the cause of missingness and the probability of being missing (Enders & Mansolf, 2018), and the intercept values b_{0t} were calibrated to achieve the target proportions at each time point. Once an individual dropped out at a given time point, all subsequent observations were also set to missing, producing a monotone attrition pattern. This procedure was applied separately within each latent class so that each class independently reached the target missing data proportion at the final time point.

4.4.6 Missing Data Methods

Five missing data methods were compared, spanning three broad categories reviewed in the Missing Data Issues in Growth Mixture Models section: non-MI approaches that handle missingness within the estimation process (FIML, Bayesian estimation via MCMC), standard MI approaches (MI-JM, MI-FCS), and the proposed Bayesian nonparametric MI approach (MI-CRP).

Full Information Maximum Likelihood

FIML handles missing data by maximizing the observed-data likelihood, using all available information without discarding incomplete cases or requiring a separate imputation step (Enders, 2022). The two-class linear GMM was estimated with the robust maximum likelihood estimator (MLR) in *Mplus* (L. K. Muthén & Muthén, 2017).

Bayesian Estimation via Markov Chain Monte Carlo

Bayesian estimation handles missing data by treating them as additional unknown quantities in the joint posterior distribution, which is sampled via MCMC (Enders, 2022). I estimated the two-class linear GMM in *Mplus* using a single MCMC

chain. A diffuse symmetric Dirichlet prior, $\boldsymbol{\pi} \sim \text{Dirichlet}(10, 10)$, was placed on the class proportion parameter, and default diffuse priors were retained for the covariance matrices (D. Y. Lee & Harring, 2023). To enhance estimation accuracy, I specified informative normal priors for the growth factor means. Priors for the intercept means were centered at the population values with variances set to 5% of the mean hyperparameter (Depaoli, 2013), and priors for the slope means were likewise centered at the population values but with a fixed variance of 0.05 (Heo, Depaoli, et al., 2024). Because these population values differed across the class separation conditions, I adjusted the prior hyperparameters accordingly, as summarized in Table 4.2. The MCMC sampler was run for 40,000 iterations, with the first 20,000 discarded as burn-in and no thinning applied, and point estimates were obtained as posterior means.

Table 4.2: Prior Specifications for Growth Factor Means.

Class	Parameter	Mahalanobis Distance		
		$MD = 2.0$	$MD = 3.0$	$MD = 4.0$
Class 1	$\alpha_0^{(1)}$ (Intercept)	$\mathcal{N}(48, 2.400)$	$\mathcal{N}(48, 2.400)$	$\mathcal{N}(48, 2.400)$
	$\alpha_1^{(1)}$ (Slope)	$\mathcal{N}(3, 0.050)$	$\mathcal{N}(3, 0.050)$	$\mathcal{N}(3, 0.050)$
Class 2	$\alpha_0^{(2)}$ (Intercept)	$\mathcal{N}(47, 2.350)$	$\mathcal{N}(39, 1.950)$	$\mathcal{N}(33.613, 1.681)$
	$\alpha_1^{(2)}$ (Slope)	$\mathcal{N}(0.191, 0.050)$	$\mathcal{N}(0, 0.050)$	$\mathcal{N}(0, 0.050)$

Note. Each entry denotes a normal prior $\mathcal{N}(\mu_0, \sigma_0^2)$ with mean hyperparameter μ_0 and variance hyperparameter σ_0^2 .

Multiple Imputation via Joint Modeling

MI-JM was implemented using the Bayesian imputation procedure available in *Mplus* (Asparouhov & Muthén, 2010). The imputation model was an unrestricted variance-covariance model (i.e., the *H1* model), which jointly models all analysis variables without imposing structural constraints and is estimated via Bayesian MCMC. Upon convergence of the MCMC chain, determined by the default *Mplus* convergence criterion based on the potential scale reduction factor ($\text{PSR} < 1.05$), imputed data sets were generated by saving the sampled missing values at intervals of 100 iterations (Asparouhov & Muthén, 2010). A total of $M = 40$ imputed data sets were produced, following the recommendation of Graham (2009) and previous studies (D. Y. Lee & Harring, 2023). This number of imputations was also used for MI-FCS and MI-CRP.

Multiple Imputation via Fully Conditional Specification

MI-FCS was implemented using the *mice* package in R (van Buuren & Groothuis-Oudshoorn, 2011). The imputation method for each variable was predictive mean

matching (PMM; Little, 1988; van Buuren, 2018). PMM is a semi-parametric imputation method that replaces each missing value with an observed value drawn from a set of donor cases whose predicted values are closest to the predicted value of the incomplete case. Because imputed values are always taken from the observed data, PMM preserves the distributional properties of the original variables and avoids generating implausible values, making it a widely adopted default choice within FCS implementations (van Buuren, 2018). The number of iterations was set to $\text{maxit} = 10$, which was sufficient for convergence of the imputation chains across all conditions. A total of $M = 40$ imputed data sets were generated.

Multiple Imputation via the Chinese Restaurant Process

MI-CRP was implemented using custom code written in R with the `nimble` package (de Valpine et al., 2017). The imputation model was a DPMM with the CRP construction as described in the Bayesian Nonparametric Multiple Imputation section. For each CRP cluster k , a linear growth structure was assumed:

$$\mu_{it} = \mu_{0k} + \beta_k \cdot t, \quad t = 1, \dots, T, \quad (4.11)$$

where μ_{0k} is the cluster-specific intercept, β_k is the cluster-specific slope, and t indexes the time points. The following prior distributions were specified:

$$\alpha \sim \text{Gamma}(2, 2), \quad (4.12)$$

$$\mu_{0k} \sim \mathcal{N}(0, 1000), \quad (4.13)$$

$$\beta_k \sim \mathcal{N}(0, 1000), \quad (4.14)$$

$$\sigma_k^2 \sim \text{Inverse-Gamma}(1, 1), \quad (4.15)$$

for $k = 1, \dots, K_{\max}$, where $K_{\max}$ is the upper bound on the number of clusters. Diffuse normal priors were placed on the cluster-specific mean parameters (Qiu et al., 2025), and weakly informative inverse-gamma priors were specified for the variance parameters (Gelman, 2006). For the concentration parameter α , Si and Reiter (2013) used a $\text{Gamma}(0.25, 0.25)$ prior as a noninformative choice. However, because the sampled α values under this prior can be very small and thereby constrain the number of clusters, I instead adopted a $\text{Gamma}(2, 2)$ prior following the recommendation of Ishwaran (2000). This prior has a mean of 1, which yields an expected number of clusters on the order of $\log(1 + n)$ when $\alpha \approx 1$, providing sufficient flexibility to accommodate both small and large numbers of clusters. Although the number of latent classes in applied research in the social and behavioral sciences typically ranges from two to four (Tofighi & Enders, 2007), $K_{\max}$ was set to 20, 30, and 50 for $n = 150, 300$, and 900, respectively, to ensure sufficient flexibility while remaining computationally feasible. Because the residual variance σ_k^2 was cluster-specific and the CRP places no a priori upper limit on the number of clusters, the residual structure of the imputation model was more flexible than that of the linear GMM analysis model, in which four free time-specific residual variances were held equal across classes. The imputation model was thus no simpler than the analysis model.

Convergence of the MCMC sampler was assessed by monitoring the stability of cluster assignments $\mathbf{z} = (z_1, \dots, z_n)$ across iterations using normalized mutual information (NMI; McDaid et al., 2011; Vinh et al., 2009). Preliminary analyses indicated that NMI values stabilized within approximately 2,000 iterations. Given this, I doubled the burn-in to 4,000 iterations as a conservative choice. After burn-in, $M = 40$ completed data sets were saved at every 100 iterations (see also Asparouhov & Muthén, 2010). Preliminary analyses indicated that autocorrelation approached zero after a lag of approximately 23, confirming that an interval of 100 iterations was more than sufficient. This resulted in a total of $4,000 + 40 \times 100 = 8,000$ MCMC iterations per replication.

4.4.7 Analytic Settings

For the two non-MI approaches (FIML and Bayesian estimation via MCMC), missing data were handled directly within the estimation process, and the two-class linear GMM was fitted in a single step as described above. For the three MI-based approaches (MI-JM, MI-FCS, and MI-CRP), each of the $M = 40$ completed data sets was analyzed by fitting the two-class linear GMM. The number of latent classes was fixed at $C = 2$ across all methods and all conditions at the model fitting stage. With the exception of Bayesian estimation via MCMC, for which results were derived from the posterior distribution, all analysis models were estimated using maximum likelihood with robust standard errors (MLR; L. K. Muthén & Muthén, 2017). For the MI methods, point estimates and standard errors were pooled across the M completed data sets using Rubin’s rules (Rubin, 1987).

To guard against convergence to local maxima, 100 sets of random starting values were used for each MLR estimation, followed by 10 final stage optimizations. The random starts were generated by *Mplus* as perturbations of user-specified starting values that were informed by the population parameters. These user-specified starting values, together with the identifiability constraint $\alpha_1^{(2)} < \alpha_1^{(1)}$ on the class-specific intercept means, helped prevent potential label switching across replications (detailed below). The use of 100 initial random starts follows general guidelines recommended by Hipp and Bauer (2006) and is consistent with prior GMM research (Depaoli, 2013; McNeish, Harring, & Bauer, 2023; McNeish, Peña, et al., 2023). A sensitivity analysis using 500 random starts with 50 final stage optimizations yielded no differences in results. These starting value settings and identifiability constraints were applied consistently across all methods that used MLR estimation.

Label switching is a common issue in mixture modeling that occurs when the ordering of latent classes is arbitrarily rearranged (Stephens, 2000; Tueller et al., 2011). For all methods estimated via MLR (FIML, MI-JM, MI-FCS, and MI-CRP), the identifiability constraint $\alpha_1^{(2)} < \alpha_1^{(1)}$ on the class-specific intercept means was imposed to prevent label switching across replications (Cassiday et al., 2021), following the approach of D. Y. Lee and Harring (2023). For Bayesian estimation via MCMC, a single Markov chain was used to prevent between-chain label switching, and

the same parameter constraint was applied to address within-chain label switching, mirroring the approach of Depaoli, Heo, et al. (2026). The convergence criterion of $\text{PSR} < 1.05$ further ensured that within-chain label switching did not occur (Depaoli, 2013). For MI-CRP, because the CRP sampler assigns cluster indices without a fixed ordering, additional treatment was necessary to ensure consistent class labeling across the M imputed data sets. Following GMM estimation on each imputed data set, the estimated class assignments were aligned using the class assignment matrix approach of Tueller et al. (2011).¹

Data were generated in *Mplus* version 8.7 (L. K. Muthén & Muthén, 2017), and missing data were imposed in R (R Core Team, 2025). MI-JM imputation was performed within *Mplus*, while MI-FCS and MI-CRP imputation were carried out in R using the *mice* package (van Buuren & Groothuis-Oudshoorn, 2011) and the *nimble* package (de Valpine et al., 2017), respectively. All GMM estimation was conducted in *Mplus*.

4.4.8 Outcome Measures

I evaluated each method in terms of convergence rate, parameter recovery, and classification accuracy. Parameter recovery was assessed using relative bias and root mean square error, and classification accuracy was assessed using the correct class assignment rate. All outcome measures were computed for each condition across 500 replications, with parameter recovery and classification accuracy based on converged replications only.

Convergence Rate

I obtained the convergence rate as the proportion of replications in which the estimation procedure terminated with converged parameter solutions. For Bayesian estimation via MCMC, a replication was considered converged when the PSR fell below 1.05 for all parameters. For FIML, convergence required all parameter estimates to be admissible (Diallo et al., 2016) and the best log-likelihood value to be replicated (Depaoli et al., 2025). For the MI-based methods (MI-JM, MI-FCS, and MI-CRP), these same criteria were applied to each imputed data set, and a replication was retained only if at least 80% of imputations (32 out of 40) satisfied both criteria.

Relative Bias

Relative bias is the ratio of bias, defined as the difference between a point estimate and the true parameter value, to the absolute true value. The point estimate was the posterior mean for Bayesian estimation via MCMC and the maximum likelihood estimate from MLR for all other methods. Relative bias was computed for

¹The true class labels, known by design in the simulation, served as the reference for this alignment but were not used in either the imputation or the estimation process.

each parameter θ as

$$\text{Relative Bias} = \begin{cases} \frac{1}{R} \sum_{r=1}^R \frac{\hat{\theta}_r - \theta}{|\theta|} & \text{if } \theta \neq 0, \\ \frac{1}{R} \sum_{r=1}^R (\hat{\theta}_r - \theta) & \text{otherwise,} \end{cases} \quad (4.16)$$

where $\hat{\theta}_r$ is the parameter estimate from the r th replication, R is the number of converged replications, and θ is the true population value. When $\theta = 0$, relative bias is undefined, and raw bias is reported instead. Relative bias values within ± 0.1 were considered an acceptable threshold.

Root Mean Square Error

The root mean square error (RMSE) quantifies the average discrepancy between parameter estimates and true values, reflecting how well estimates align with the true value and the consistency of these estimates across replications. The RMSE was computed as

$$\text{RMSE} = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\theta}_r - \theta)^2}, \quad (4.17)$$

where $\hat{\theta}_r$ is the estimate from replication r .

Correct Class Assignment Rate

Classification accuracy was assessed using the correct class assignment rate (CCAR), defined as the proportion of individuals whose true class membership matched their estimated modal class assignment:

$$\text{CCAR} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(c_i = \hat{c}_i), \quad (4.18)$$

where c_i is the true class label and $\hat{c}_i$ is the modal class assignment for individual i . For the MI-based methods, the CCAR was computed for each valid imputed data set and then averaged across the valid imputations within each replication.

4.5 Simulation Results

I present the results of the simulation study in this section. Results are organized in the order of convergence rates, parameter recovery, and classification accuracy.

4.5.1 Convergence and Degenerate Solutions

The overall proportion of converged replications across all conditions and methods was 92.64% on average, with a median of 97.80%. Among the replications that satisfied the convergence criteria, a subset yielded degenerate solutions in which one latent class collapsed to fewer than 5% of the sample, resulting in a class proportion for the other class approaching one. When a class contains very few observations, its parameters become poorly estimated and the solution is unlikely to represent substantively meaningful subgroups (Bauer & Curran, 2003; Hipp & Bauer, 2006). I identified degenerate solutions using the 5% threshold from Nasserinejad et al. (2017), flagging any replication in which one class fell below this threshold.² After these exclusions, 90.20% of replications were valid on average (median = 96.00%).

Table 4.3 presents convergence, degenerate solution, and valid replication rates from selected conditions for brevity, averaged across class proportion (equal vs. unequal) and missing data proportion (10% vs. 40%).³ All three rates are percentages of 500 replications per condition, and the convergence rate minus the degenerate rate gives the valid replication rate. Full results across all conditions are available in the supplementary materials at <https://osf.io/8pbd5/>.

Because convergence difficulties were largely confined to the lowest class separation condition ($MD = 2.0$), I first focus on this condition. At this class separation level, FIML maintained the highest convergence rates among the non-MI methods. Bayesian-MCMC showed comparatively lower convergence that, unlike the other methods, did not improve with larger sample sizes under this class separation level. I inspected the *Mplus* output files for the nonconverged Bayesian-MCMC replications and found that the label-switching constraints could not be satisfied, which I revisit in the Discussion. For the MI-based methods, MI-CRP achieved the highest convergence at smaller sample sizes, followed by MI-JM and MI-FCS. Under 40% missing data at $n = 150$, MI-JM and MI-FCS valid replication rates dropped below 50% (see supplementary materials for condition-specific values). At moderate and high class separation ($MD = 3.0$ and 4.0), convergence rates generally improved and exceeded 90% for most methods at larger sample sizes.

Degenerate solutions were similarly concentrated at $MD = 2.0$. All three MI methods produced higher degenerate rates at this level, and a small but nonzero rate persisted for MI-CRP even at higher class separation levels. Bayesian-MCMC yielded no degenerate solutions across any condition. In Bayesian estimation, when a class collapses, its parameters become weakly identified and are sampled from priors rather than informed by the data (Xiao et al., 2025). This leads to poor chain mixing, which

²It is fully acknowledged that excluding very small classes may be problematic if those classes represent genuinely existing subgroups. I discuss this issue and offer practical guidance in the Discussion section.

³Under specific conditions, individual values deviated from the averages presented in Table 4.3. For example, Bayesian-MCMC convergence at $n = 900$ with $MD = 2.0$ dropped to 32.2% under unequal class proportions and 40% missing data. More generally, convergence and valid replication rates were lower under 40% missing data than under 10% missing data, particularly for MI-JM and MI-FCS at $MD = 2.0$.

Table 4.3: Convergence, Degenerate Solution, and Valid Replication Rates from Selected Conditions.

n	MD	FIML			Bayesian-MCMC		
		Conv	Degen	Valid	Conv	Degen	Valid
150	2.0	85.5	7.6	77.9	69.7	0.0	69.7
	3.0	90.6	0.2	90.5	98.3	0.0	98.3
	4.0	91.8	0.0	91.8	98.2	0.0	98.2
300	2.0	96.4	4.1	92.3	67.6	0.0	67.6
	3.0	98.4	0.0	98.4	98.8	0.0	98.8
	4.0	98.5	0.0	98.5	98.6	0.0	98.6
900	2.0	99.9	0.1	99.8	66.2	0.0	66.2
	3.0	100.0	0.0	100.0	98.5	0.0	98.5
	4.0	100.0	0.0	100.0	99.3	0.0	99.3

n	MD	MI-JM			MI-FCS			MI-CRP		
		Conv	Degen	Valid	Conv	Degen	Valid	Conv	Degen	Valid
150	2.0	68.2	14.2	54.0	61.3	13.1	48.2	90.7	11.3	79.3
	3.0	87.9	0.5	87.5	79.8	0.7	79.2	97.3	2.6	94.7
	4.0	90.7	0.0	90.7	88.6	0.0	88.6	98.3	1.1	97.2
300	2.0	88.3	10.7	77.7	84.0	12.6	71.4	95.8	9.2	86.7
	3.0	97.3	0.0	97.3	94.6	0.0	94.6	99.6	2.6	97.0
	4.0	98.2	0.0	98.2	98.8	0.0	98.8	99.8	1.9	98.0
900	2.0	99.2	2.0	97.2	98.2	6.3	91.8	97.2	4.4	92.8
	3.0	100.0	0.0	100.0	99.8	0.0	99.8	99.7	3.1	96.6
	4.0	100.0	0.0	100.0	100.0	0.0	100.0	99.4	1.8	97.6

Note. All values are percentages out of 500 replications per condition. Conv = convergence rate. Degen = degenerate solution rate. Valid = valid replication rate. Values are averaged across class proportion (equal vs. unequal) and missing data proportion (10% vs. 40%).

is more likely to be filtered out by the $\text{PSR} < 1.05$ criterion rather than retained as a post-convergence degenerate solution.

The Valid column in Table 4.3 shows the percentage of replications, out of 500, used in subsequent analyses after excluding both nonconverged and degenerate solutions. Valid replication rates were highest at $\text{MD} \geq 3.0$ with larger sample sizes, where most methods retained close to 100% of replications. At $\text{MD} = 2.0$, valid rates were lower for MI-JM and MI-FCS at smaller sample sizes, though both methods recovered as n increased to 900. Bayesian-MCMC valid rates remained low regardless of sample size, due to convergence failures rather than degenerate solutions. MI-CRP retained the highest valid rates among the MI methods at smaller sample sizes at $\text{MD} = 2.0$, though degenerate exclusions reduced its valid rates even at higher class separation levels. The results reported in the subsequent sections on parameter recovery and classification accuracy are based on these valid replications (see also Chen et al., 2001).

4.5.2 Parameter Recovery

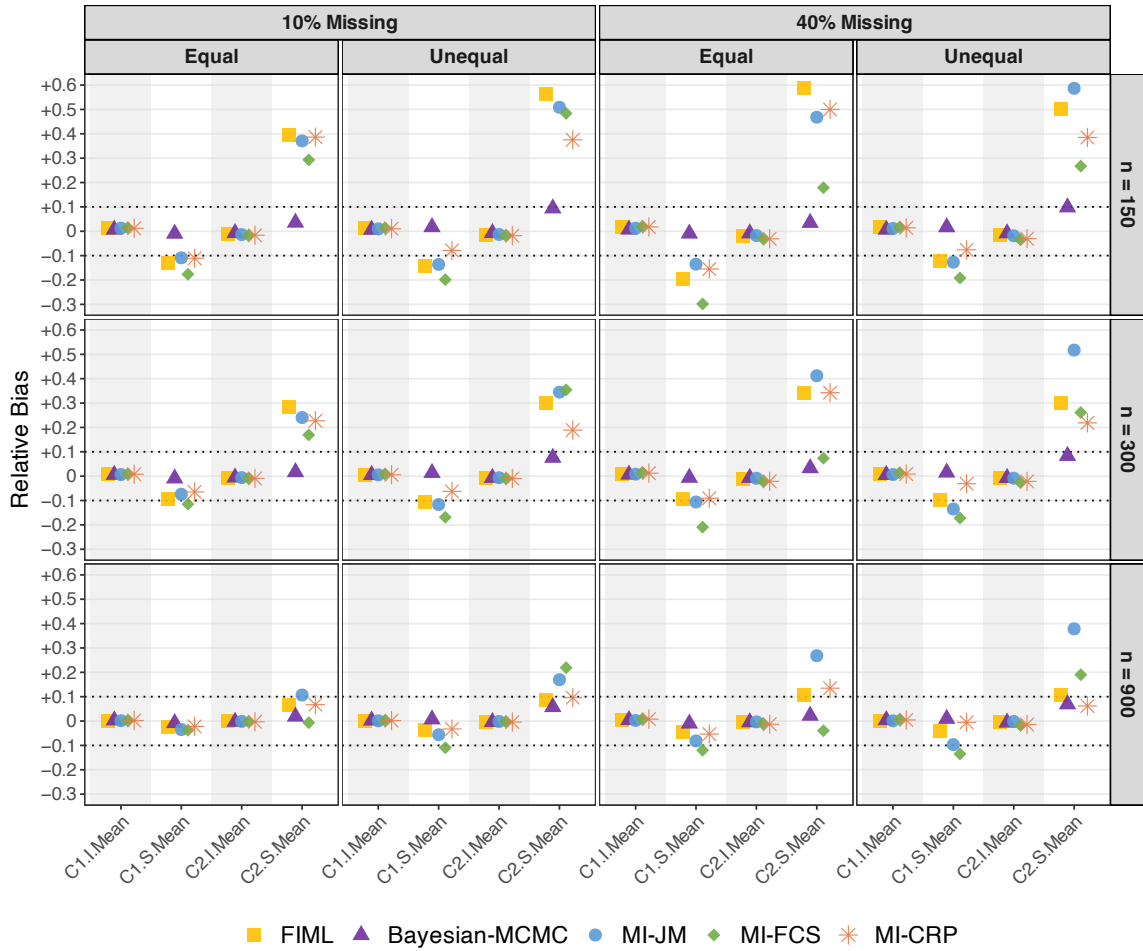
I evaluated parameter recovery for the growth factor means across both latent classes, the elements of the covariance matrix (i.e., variances of and covariance between the growth factors) that were held constant between classes, and class proportions. I first report relative bias, followed by RMSE.

Relative Bias

Relative bias values for the growth factor means are presented separately for each class separation condition in Figures 4.3, 4.4, and 4.5. Each figure follows the same layout: Rows correspond to the three sample size levels ($n = 150, 300$, and 900), and columns represent the four conditions derived from crossing missing data proportion (10% and 40%) with class proportion (equal and unequal). Within each panel, the x -axis indicates the parameter, and the y -axis indicates relative bias. The five methods are distinguished by color and shape, and alternating grey shading distinguishes adjacent parameters within each panel. Horizontal dotted lines at ± 0.1 mark the conventional threshold for acceptable relative bias, such that values falling between -0.1 and $+0.1$ are considered negligible. As noted in the Simulation Design section, raw bias was used in place of relative bias for parameters with population values of zero, for which relative bias is undefined. For the Class 2 slope mean ($\alpha_2^{(2)}$), I report raw bias across all separation conditions because the population value is small at $\text{MD} = 2.0$ ($\alpha_2^{(2)} = 0.191$) and zero at higher separations. As such, using relative bias at $\text{MD} = 2.0$ would produce inflated values that are not comparable across conditions.

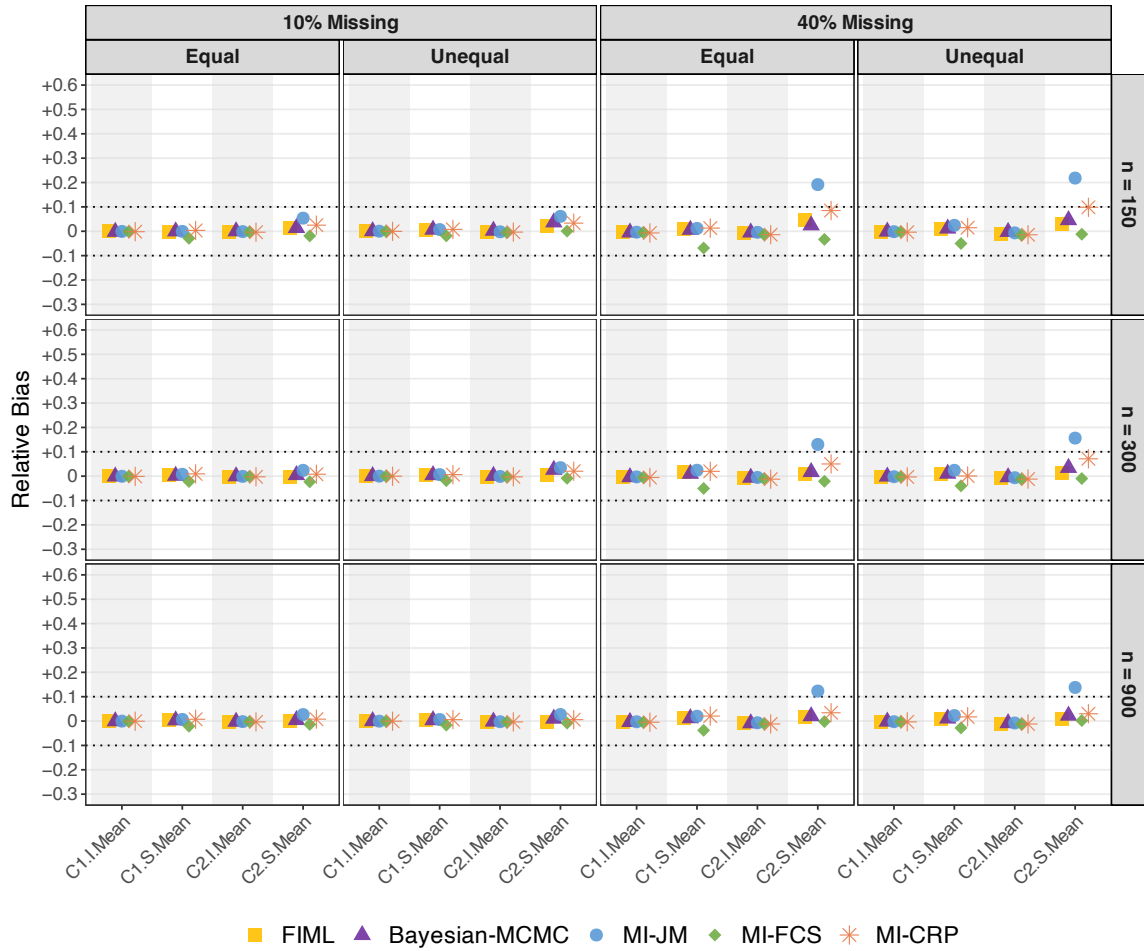
Growth factor means. I begin with the low class separation condition ($\text{MD} = 2.0$; Figure 4.3). At $n = 150$ with 10% missing data and equal class proportions, the intercept means for both classes (C1.I.Mean and C2.I.Mean) were accurately

Figure 4.3: Relative Bias for Growth Factor Means at MD = 2.0.



estimated by all five methods, with relative bias close to zero. The slope means showed more differentiation. The slope mean for Class 1 (C1.S.Mean) exceeded or approached the -0.1 threshold for FIML, MI-JM, MI-FCS, and MI-CRP, with MI-FCS showing the largest negative bias. Bayesian-MCMC, by contrast, produced near-zero bias for this parameter. The slope mean for Class 2 (C2.S.Mean) was the most biased, with all methods except Bayesian-MCMC showing positive bias of around or more than $+0.3$, whereas Bayesian-MCMC remained close to zero. Under unequal class proportions with 10% missing data, the same general pattern held, but with somewhat larger bias for the slope mean of Class 2. When missing data increased to 40%, the intercept means remained stable across both class proportion conditions. On the other hand, bias worsened for the slope parameters. MI-FCS showed the largest negative bias for the slope mean of Class 1, and bias for the slope mean of Class 2 increased for all methods except MI-FCS, which nonetheless retained a nonnegligible level of bias. At larger sample sizes, bias generally decreased. Yet, MI-JM still showed overestimation for the slope mean of Class 2, even at $n = 900$.

Figure 4.4: Relative Bias for Growth Factor Means at MD = 3.0.

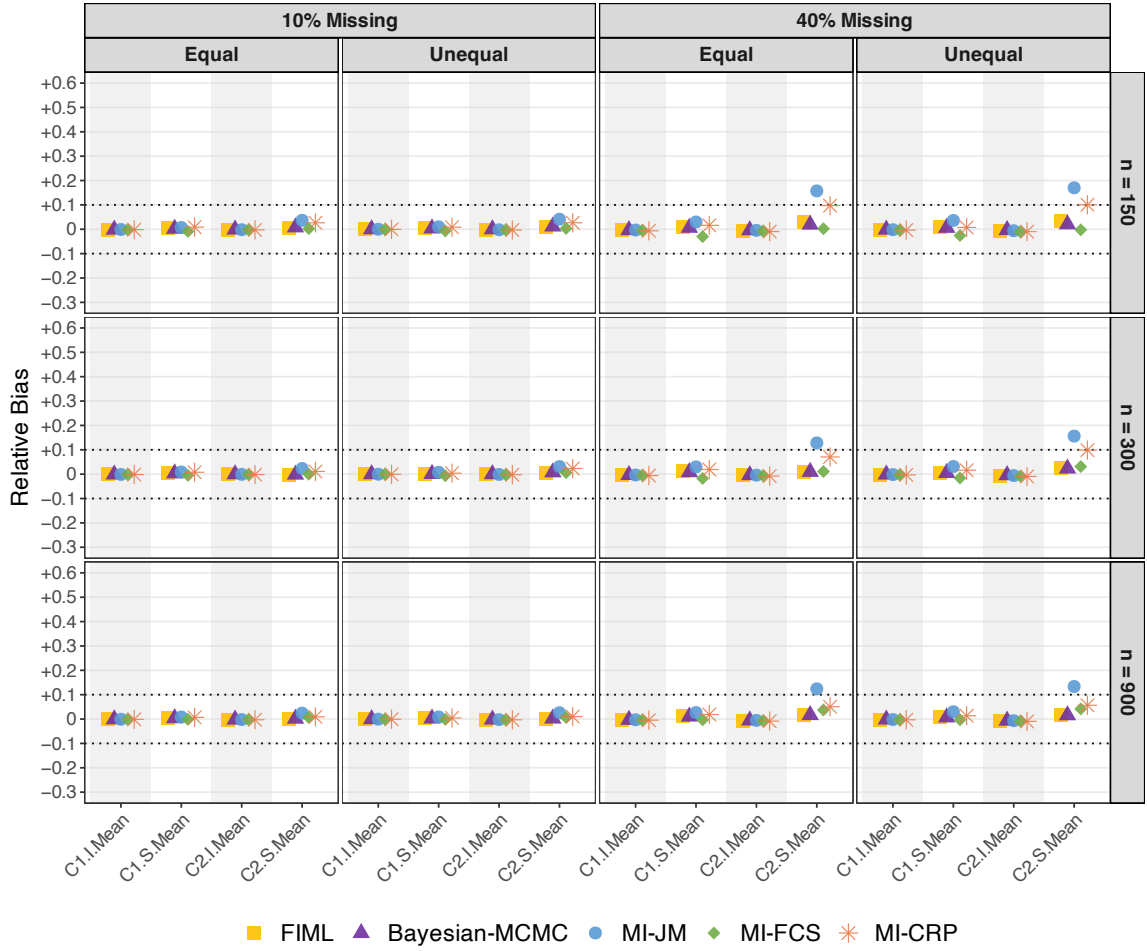


At moderate and high class separation (MD = 3.0 and 4.0; Figures 4.4 and 4.5), relative bias for the intercept means was negligible, and that for the slope means was substantially reduced across all conditions. Still, under 40% missing data, the slope mean for Class 2 showed positive bias for MI-JM that exceeded the threshold at both separation levels. This was still the case at larger sample sizes, although the bias was attenuated.

Growth factor variances and covariance. Relative bias for the variance and covariance parameters of the growth factors is shown in Figures 4.6, 4.7, and 4.8. The figure layout is the same as for the growth factor means, with the x -axis now indicating the intercept variance (I.Variance), slope variance (S.Variance), and covariance between the intercept and slope (Covariance).

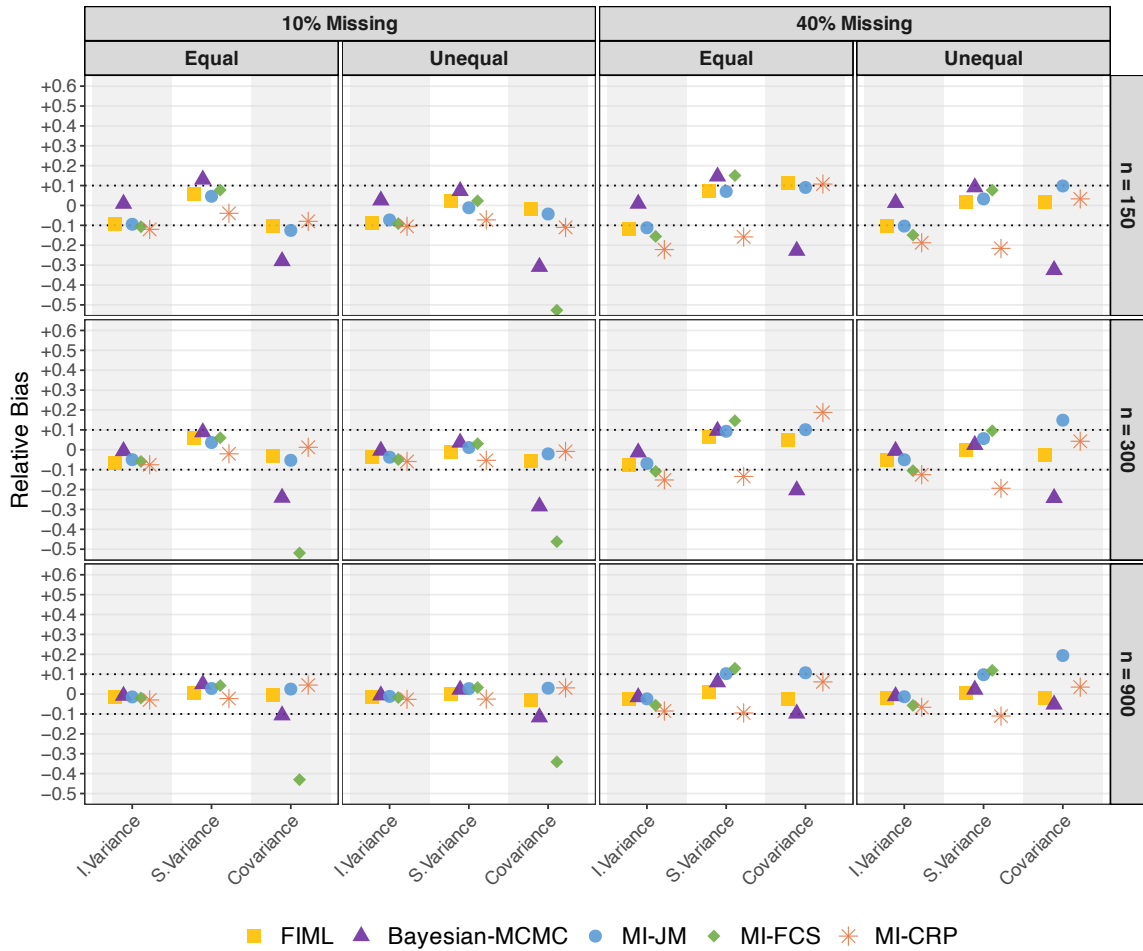
At MD = 2.0 (Figure 4.6), I describe the results beginning with $n = 150$. Under 10% missing data (left two panels), the intercept variance was slightly underestimated by FIML, MI-JM, MI-FCS, and MI-CRP, with bias near or just beyond

Figure 4.5: Relative Bias for Growth Factor Means at MD = 4.0.



the -0.1 threshold, whereas Bayesian-MCMC was close to zero. For the slope variance, Bayesian-MCMC showed a slight positive bias that exceeded the threshold under equal proportions, whereas the remaining methods fell within acceptable limits. The covariance was the most challenging parameter. MI-FCS produced pronounced negative bias, falling outside the plotted range (which spans -0.5 to $+0.6$ for visual clarity). Bayesian-MCMC also showed negative covariance bias around -0.30 . FIML, MI-JM, and MI-CRP were closer to zero, generally within or near the ± 0.1 threshold. Under 40% missing data (right two panels), the intercept and slope variance patterns were similar to the 10% condition but slightly amplified. The covariance, however, shifted noticeably. MI-FCS bias worsened further, falling far beyond the plotted range. Bayesian-MCMC maintained negative bias around -0.25 to -0.30 . FIML, MI-JM, and MI-CRP remained within or near the acceptable range. With $n = 300$ and 900 (middle and bottom rows), bias for the intercept and slope variance decreased toward zero for all methods. While this decreasing bias pattern was likewise observed for the covariance parameter, the bias for MI-FCS persisted even at

Figure 4.6: Relative Bias for Growth Factor Variances and Covariance at MD = 2.0.

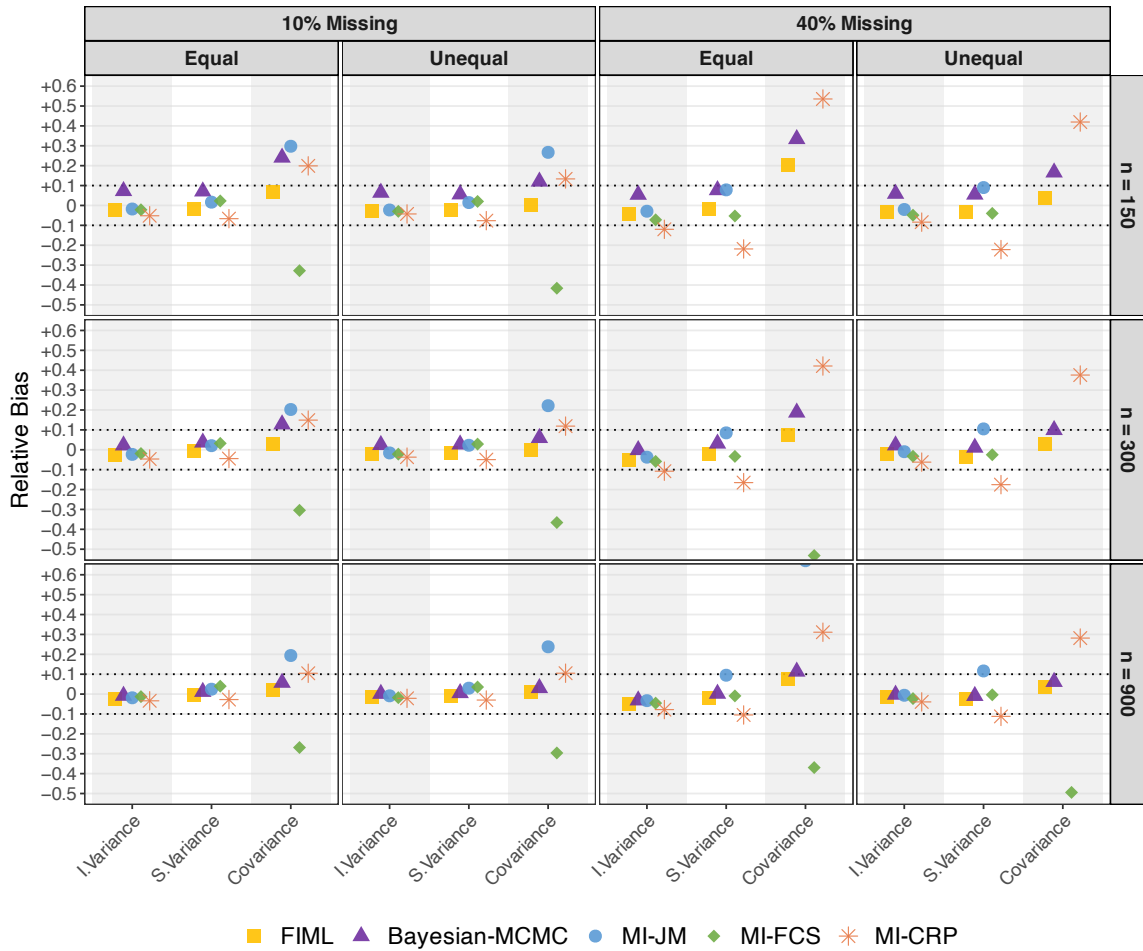


$n = 900$, remaining far outside the plotted range under 40% missing data.

At MD = 3.0 (Figure 4.7), for $n = 150$ (the first row), the intercept and slope variance parameters mostly fell within the ± 0.1 threshold. The exception was MI-CRP, which showed negative bias of around -0.2 for the slope variance under the 40% missing data condition. The covariance remained problematic. MI-FCS exhibited a large negative bias under 10% missing, which worsened substantially under 40% missing data, with values beyond the plotted range. MI-JM showed the opposite pattern, with substantial positive bias that also grew with missing data proportion, likewise reaching values outside the displayed range under 40% missing. Bayesian-MCMC and MI-CRP showed greater positive bias as the amount of missing data increased to 40%. FIML remained closest to zero across all conditions. These patterns of bias were attenuated as sample size increased. Yet, the extreme negative bias of MI-FCS and the extreme positive bias of MI-JM persisted.

As class separation increased to MD = 4.0 (Figure 4.8), the overall patterns observed for MD = 3.0 were similar such that variance parameters remained within acceptable limits, and the covariance showed somewhat smaller magnitudes. However,

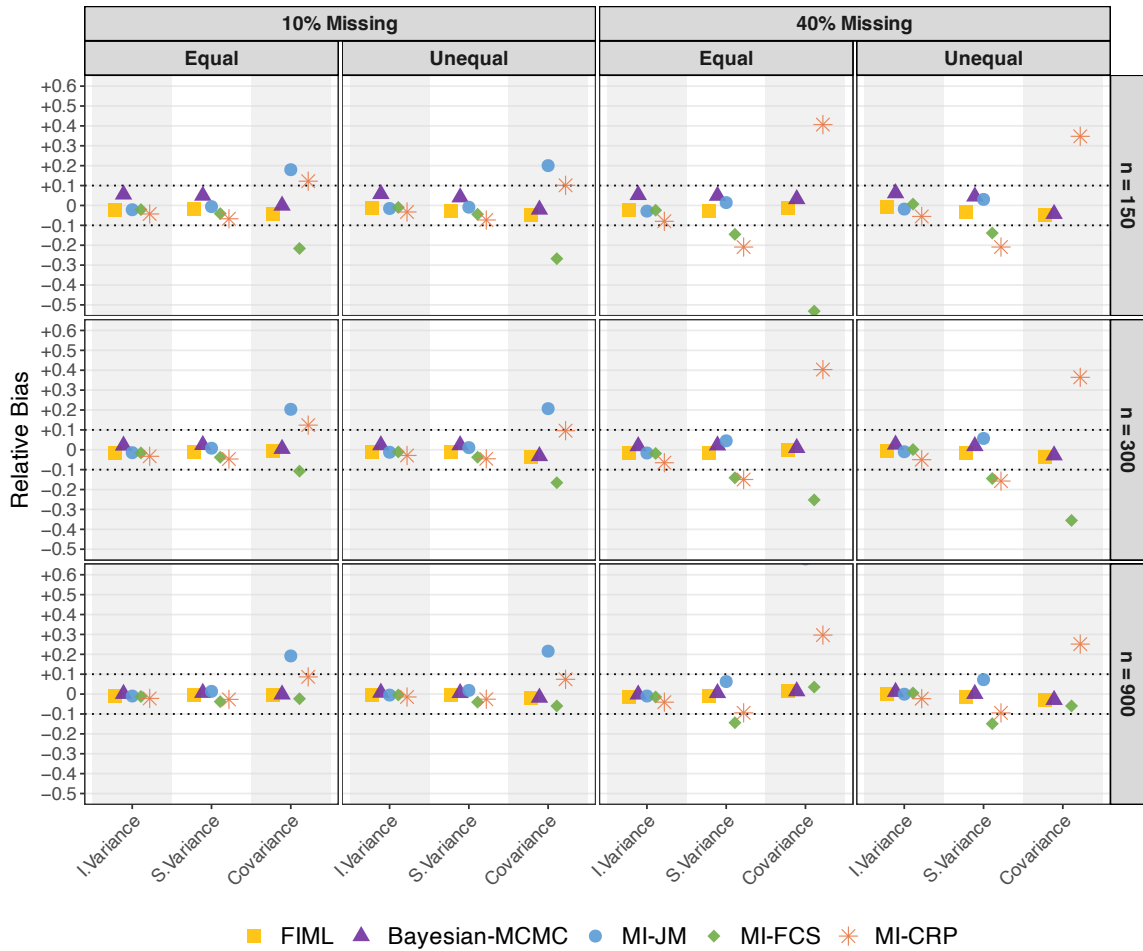
Figure 4.7: Relative Bias for Growth Factor Variances and Covariance at MD = 3.0.



MI-JM continued to produce positive covariance bias beyond the plotted range under 40% missing data across all sample sizes, whereas MI-FCS showed negative covariance bias beyond the plotted range at $n = 150$. MI-CRP showed moderate positive covariance bias that increased with missing data proportion. FIML and Bayesian-MCMC produced bias closer to zero. Larger sample sizes partially attenuated these differences but did not fully resolve the covariance bias for MI-FCS, MI-JM, and MI-CRP.

Class proportions. Because the two class proportions sum to one in a two-class model, the bias for one class determines the bias for the other. I report relative bias for the Class 1 proportion under equal class proportions (Figure 4.9) and for the Class 2 (minority class) proportion under unequal class proportions (Figure 4.10). Unlike the previous figures, all three class separation conditions appear along the x -axis within each panel, as class proportion is a single parameter that is more readily displayed across all separation levels within one figure. Rows correspond to the three

Figure 4.8: Relative Bias for Growth Factor Variances and Covariance at MD = 4.0.

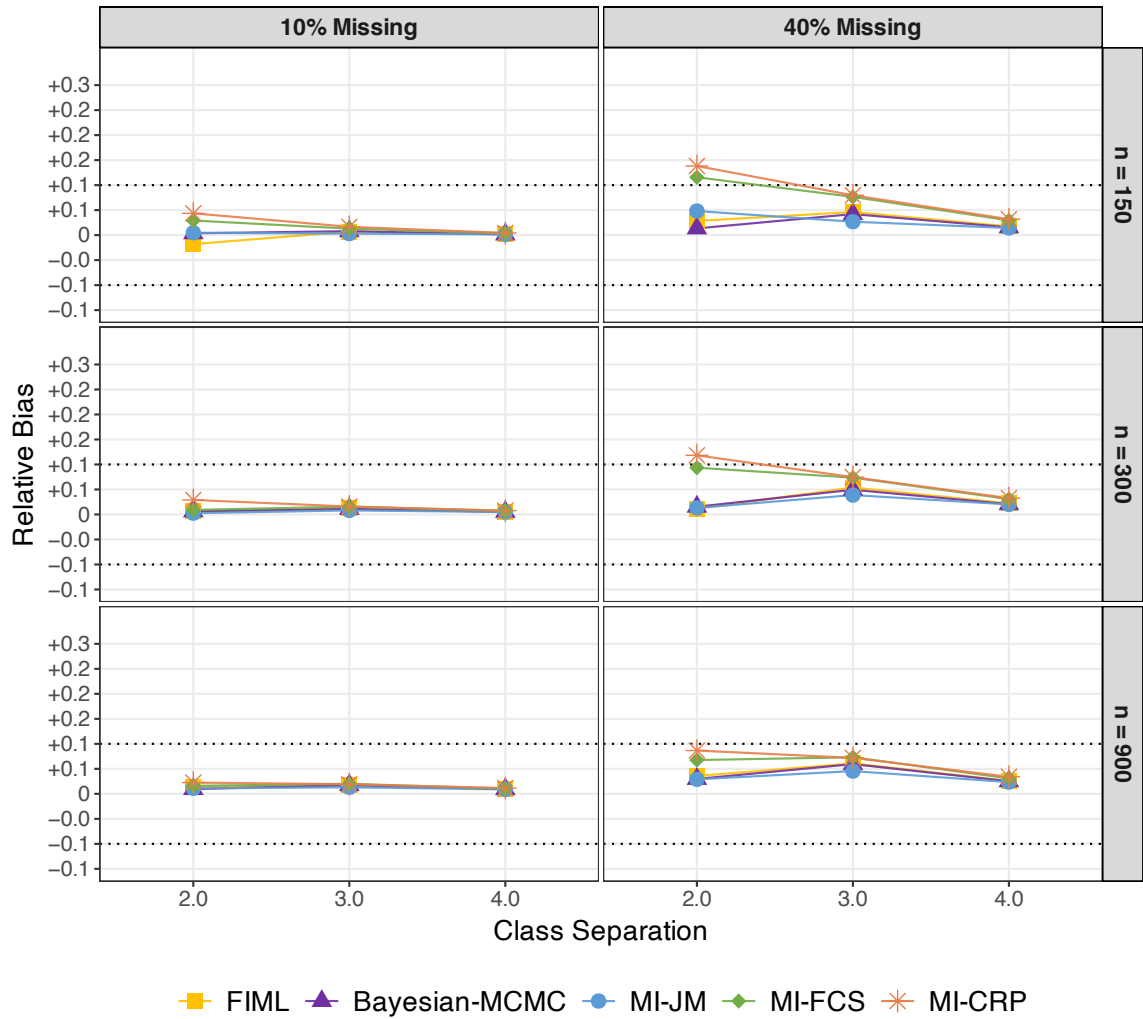


sample sizes ($n = 150, 300$, and 900), and columns represent the two missing data proportions (10% and 40%). Horizontal dotted lines at ± 0.1 mark the acceptable threshold for relative bias.

Under equal class proportions (Figure 4.9), relative bias for the Class 1 proportion was small and within the ± 0.1 threshold for all methods across all separation levels at $n = 150$ with 10% missing data. Under 40% missing data, MI-FCS and MI-CRP showed positive bias at MD = 2.0 that slightly exceeded the threshold, whereas the other methods remained closer to zero. As class separation increased, bias decreased for all methods. With larger sample sizes, bias was further reduced, and at $n = 900$, all methods produced bias within ± 0.1 across all conditions.

For the minority class proportion under unequal class proportions (Figure 4.10), bias was more pronounced. For $n = 150$ with 10% missing data, all methods tended to overestimate the minority class proportion at MD = 2.0, shifting the estimated class proportions closer to equality. FIML showed the largest such bias, followed by MI-JM and MI-FCS. MI-CRP and Bayesian-MCMC were the least biased. As class separation increased to MD = 3.0 and 4.0, bias decreased for all methods. Under

Figure 4.9: Relative Bias for Class 1 Proportion (Equal Proportions).

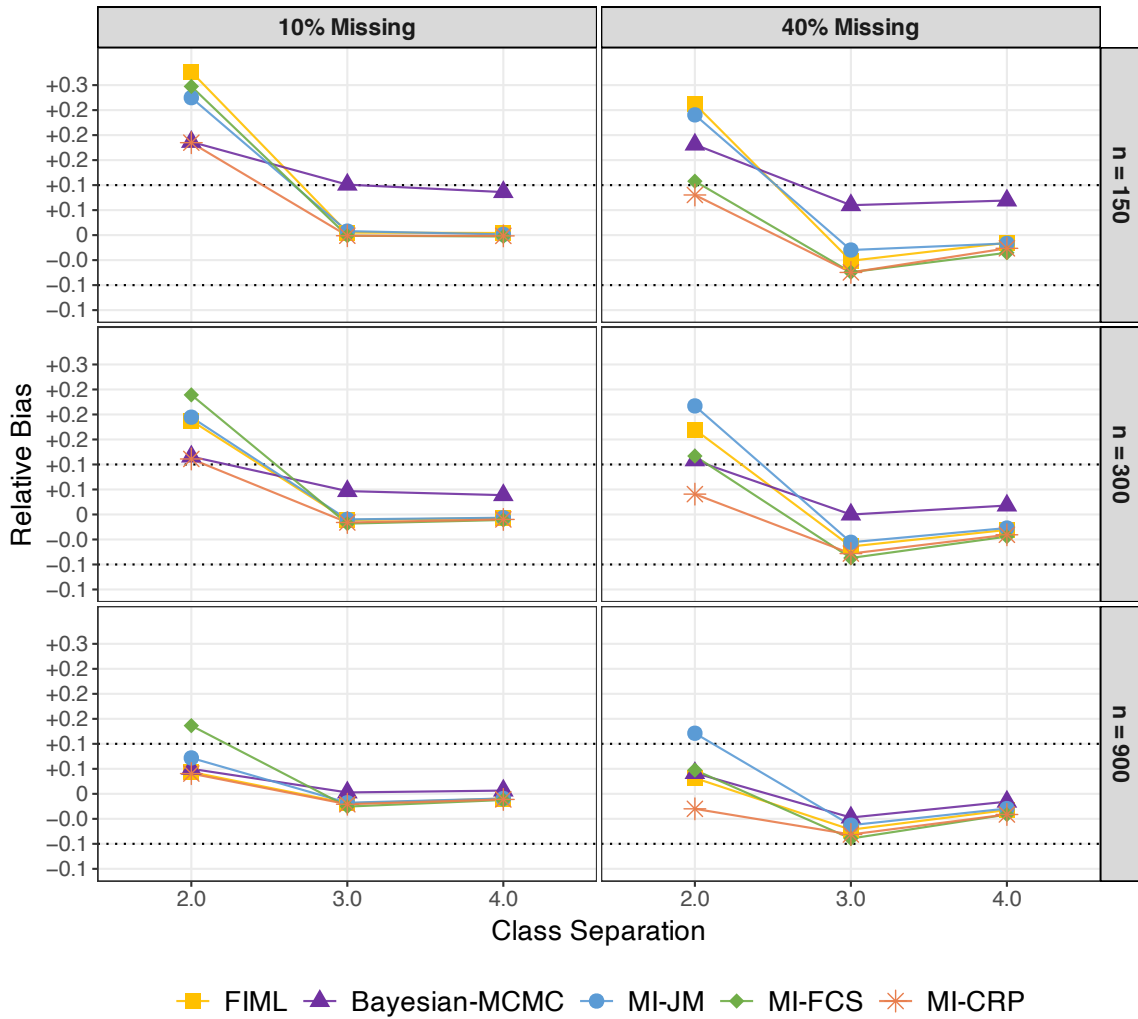


40% missing data, the tendency toward overestimation of the minority class persisted at MD = 2.0. FIML again showed the largest bias, and MI-CRP the smallest. With larger sample sizes, bias was reduced. When $n = 900$, most methods produced bias within ± 0.1 across all separation levels, though MI-FCS and MI-JM slightly exceeded the threshold at MD = 2.0.

Root Mean Square Error

I now turn to RMSE. Figures 4.11, 4.12, and 4.13 show RMSE for the growth factor means; Figures 4.14, 4.15, and 4.16 provide RMSE for the variances and covariance; and Figure 4.17 displays RMSE for the class proportions. The growth factor mean and variance/covariance figures follow the same panel layout described previously for relative bias.

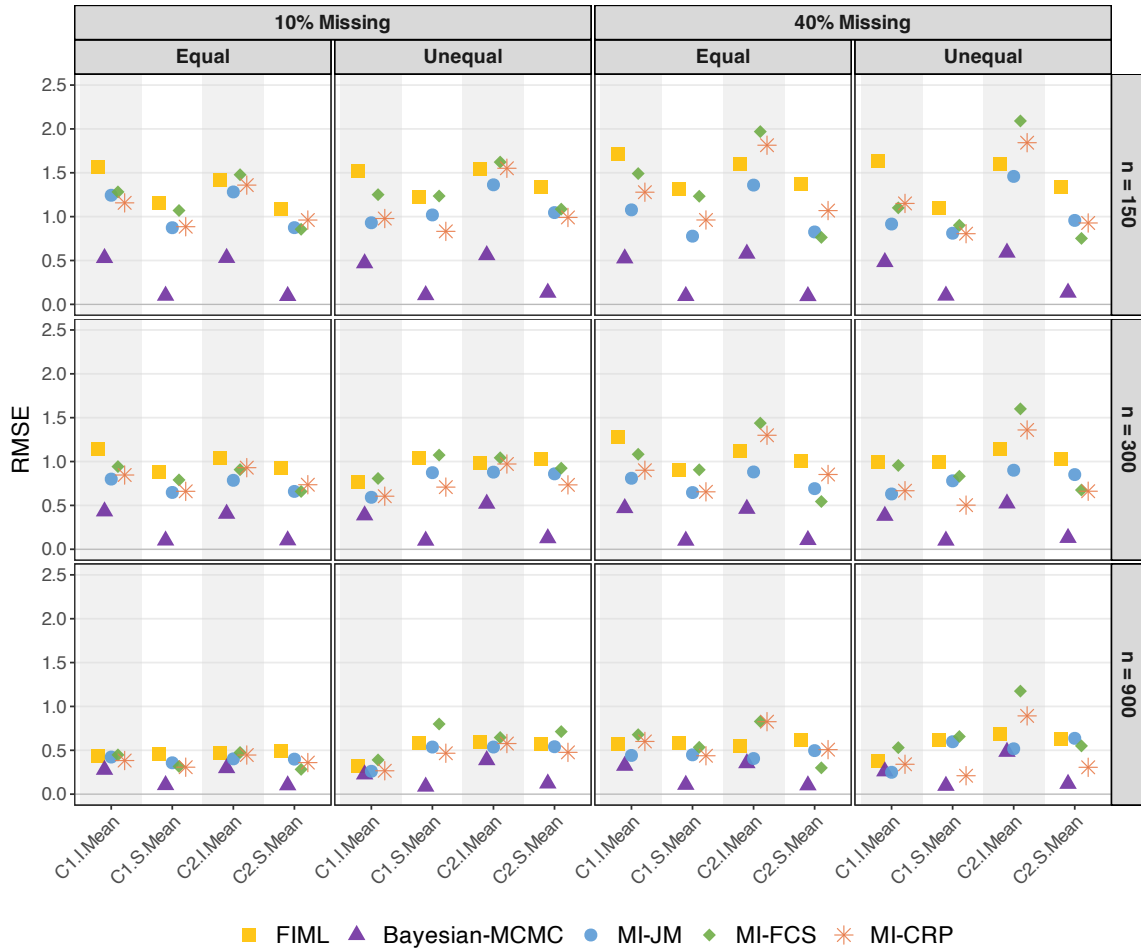
Figure 4.10: Relative Bias for Class 2 Proportion (Unequal Proportions).



Growth factor means. At MD = 2.0 (Figure 4.11), Bayesian-MCMC yielded the smallest RMSE values across all four mean parameters with $n = 150$ (first row). Under 10% missing data (left two panels), FIML tended to produce the largest RMSE, with MI-JM, MI-FCS, and MI-CRP falling in between, though the ordering among these three varied by parameter. Under 40% missing data (right two panels), RMSE increased for all methods but the patterns were broadly similar. At $n = 300$ and 900 (middle and bottom rows), RMSE decreased for all methods. The gap between Bayesian-MCMC and the rest narrowed at $n = 300$ but remained visible. By $n = 900$, RMSE values were substantially reduced. Bayesian-MCMC retained the lowest values.

At MD = 3.0 and 4.0 (Figures 4.12 and 4.13), RMSE values were lower overall and the five methods were more similar. For both separation levels, with $n = 150$, RMSE values were higher for the intercept mean of Class 2. By $n = 300$ and 900, all methods produced comparable RMSE values. The RMSE values were generally

Figure 4.11: Root Mean Square Error for Growth Factor Means at MD = 2.0.



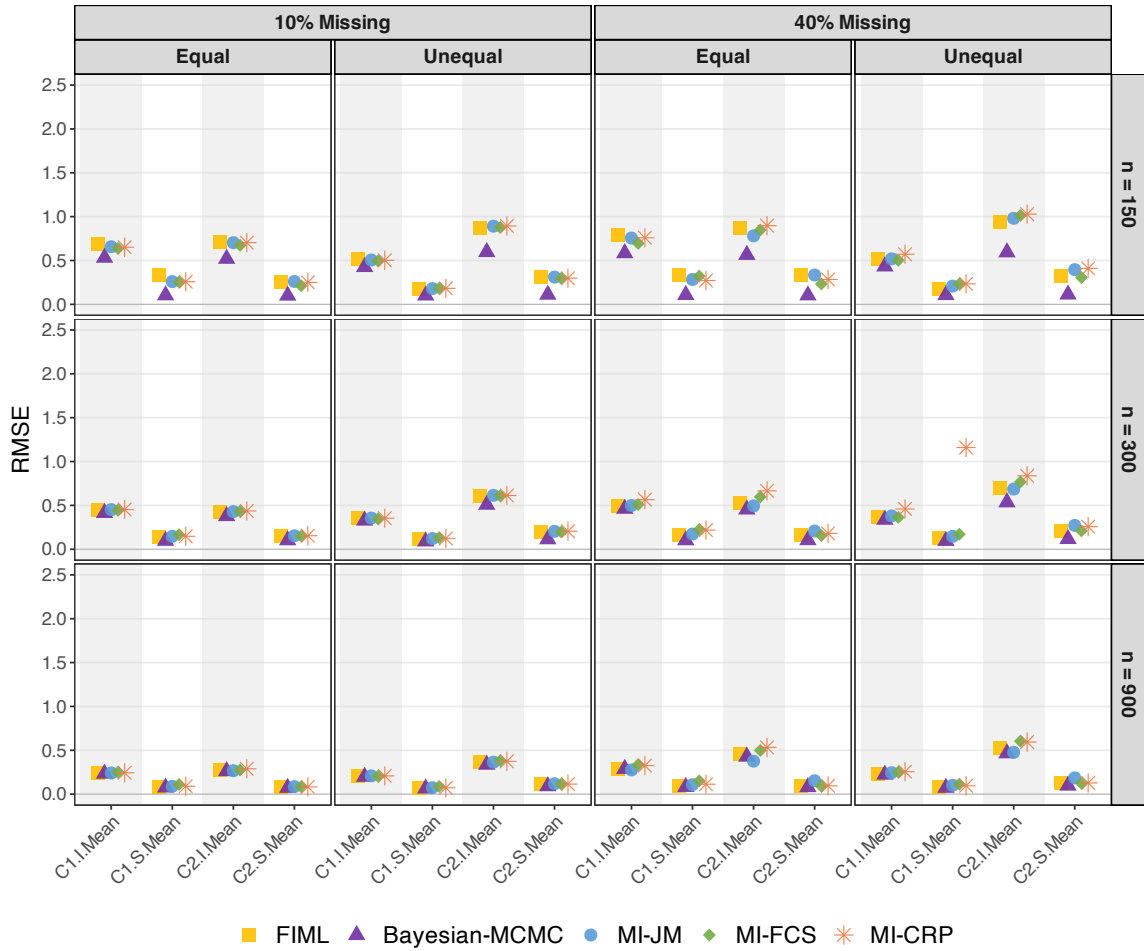
lower for the slope means than for the intercept means across both classes.

Growth factor variances and covariance. RMSE for the variance and covariance parameters across different class separation levels is reported in Figures 4.14, 4.15, and 4.16.

At MD = 2.0 (Figure 4.14), when $n = 150$ under 10% missing data (left two panels in first row), Bayesian-MCMC stood apart with lower intercept variance RMSE, and the remaining four methods clustered at higher levels. The slope variance and covariance had smaller RMSE overall, with Bayesian-MCMC again the lowest. Under 40% missing data (right two panels), RMSE rose for all methods, and MI-CRP had the highest RMSE for the intercept variance under equal proportions. Larger sample sizes reduced RMSE across the board, and by $n = 900$ (bottom row) the differences among methods were smaller.

At higher class separation levels (MD = 3.0 and 4.0; Figures 4.15 and 4.16), RMSE values tended to drop further, and the five methods were more tightly grouped.

Figure 4.12: Root Mean Square Error for Growth Factor Means at MD = 3.0.

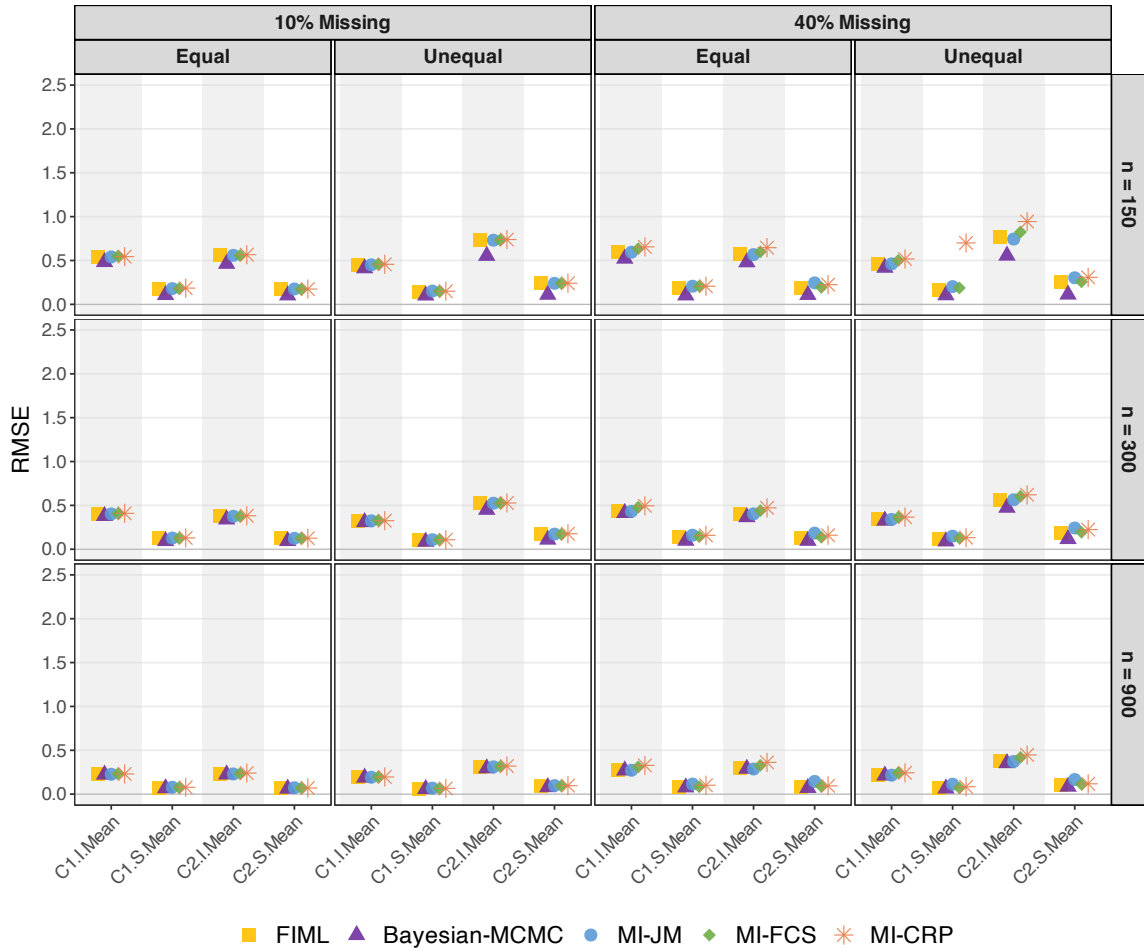


Among the parameters, the intercept variance showed higher RMSE than the slope variance and the covariance. By $n = 900$, all methods reached comparable RMSE levels for the variance and covariance parameters.

Class proportions. I present RMSE for the class proportions in Figure 4.17. Because the two class proportions sum to one, the estimation error is identical for both classes, and I report RMSE for the Class 1 proportion only. Rows correspond to sample sizes, and columns represent the crossing of missing data proportion and class proportion. The x -axis indicates class separation, the y -axis indicates RMSE, and methods are distinguished by color and shape.

RMSE for class proportions was highest at MD = 2.0 regardless of sample size and across all conditions. In general, Bayesian-MCMC consistently produced the smallest RMSE. FIML and MI-FCS tended to produce the largest values depending on conditions, with the remaining MI methods falling in between. With increasing class separation and sample size, RMSE decreased for all methods and the differences

Figure 4.13: Root Mean Square Error for Growth Factor Means at MD = 4.0.



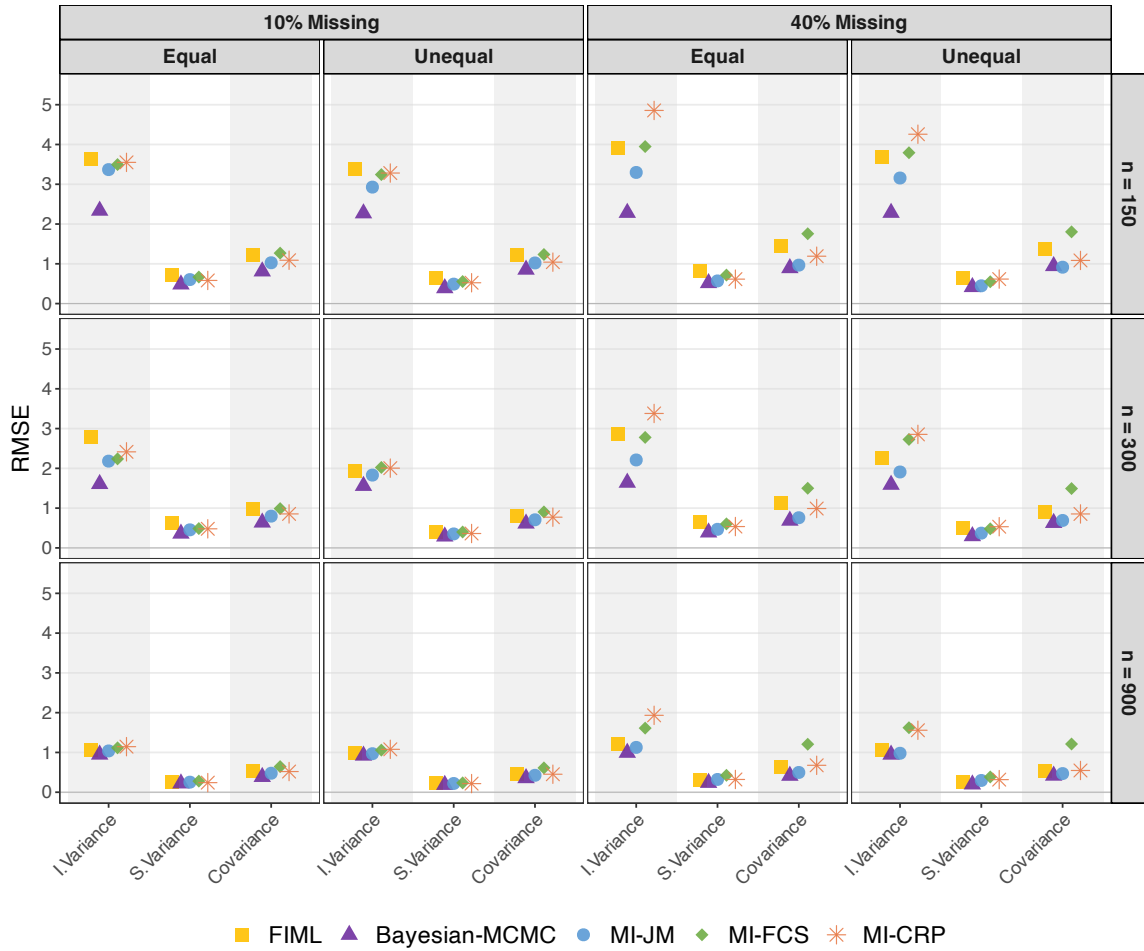
among them became negligible.

Summary of Parameter Recovery

The intercept means were well recovered by all methods, with relative bias close to zero across conditions. The slope means were more challenging. At low class separation, all methods except Bayesian-MCMC showed bias exceeding or approaching the ± 0.1 threshold, and the Class 2 slope mean showed a greater degree of bias. The covariance between the intercept and slope proved the most difficult parameter to recover. MI-FCS and MI-JM showed persistent and the most severe bias overall, and MI-CRP showed positive covariance bias to a lesser degree. FIML and Bayesian-MCMC generally produced the least bias. For the class proportions, all methods tended to pull estimated proportions toward equality under unequal conditions, with FIML showing the largest such tendency at low class separation.

RMSE results complemented the bias findings. Bayesian-MCMC yielded the lowest RMSE across nearly all parameters, consistent with the stability conferred by

Figure 4.14: Root Mean Square Error for Growth Factor Variances and Covariance at MD = 2.0.



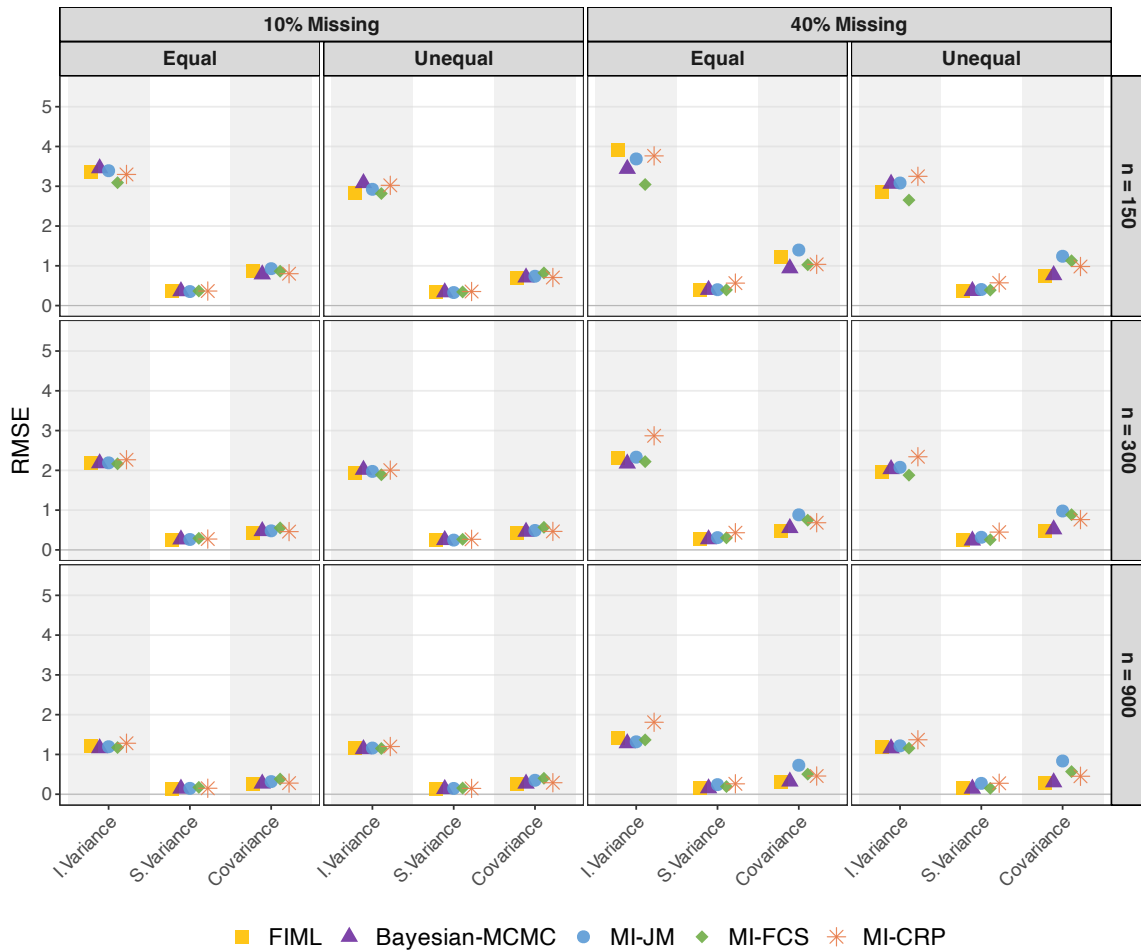
informative priors. FIML produced the highest RMSE for the growth factor means at low class separation. The MI-based methods produced intermediate RMSE values, with no single MI approach consistently outperforming the others.

Taken together, Bayesian-MCMC achieved the best parameter recovery overall, benefiting from correct model specification and informative prior specification. Among the MI approaches, MI-CRP did not produce uniformly lower bias or RMSE than MI-JM and MI-FCS, but it avoided the extreme covariance bias of MI-FCS and maintained bias levels comparable to the other methods.

4.5.3 Classification Accuracy

For classification accuracy, CCAR values for each condition are presented in Figure 4.18, after excluding replications with nonconverged and degenerate solutions. Columns represent the four conditions derived from crossing the missing data proportion with class proportion, and rows correspond to the three sample size levels.

Figure 4.15: Root Mean Square Error for Growth Factor Variances and Covariance at MD = 3.0.

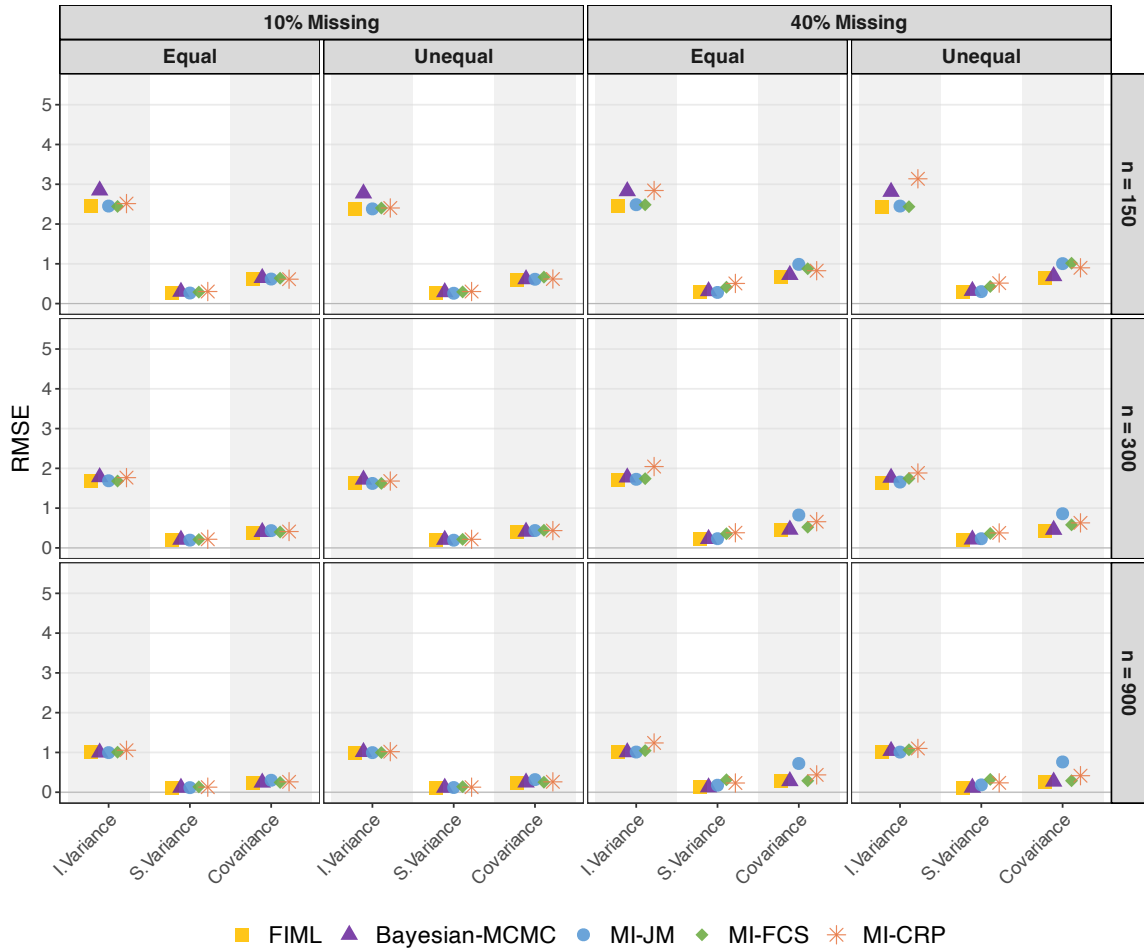


Within each panel, the x -axis represents class separation from MD = 2.0 through 4.0, and the y -axis indicates classification accuracy quantified by CCAR. At each level of class separation, the boxplots from left to right correspond to FIML, Bayesian estimation via MCMC, MI-JM, MI-FCS, and MI-CRP. I organize the results by sample size, beginning with the smallest ($n = 150$) and proceeding to larger samples.

Small Sample Size ($n = 150$)

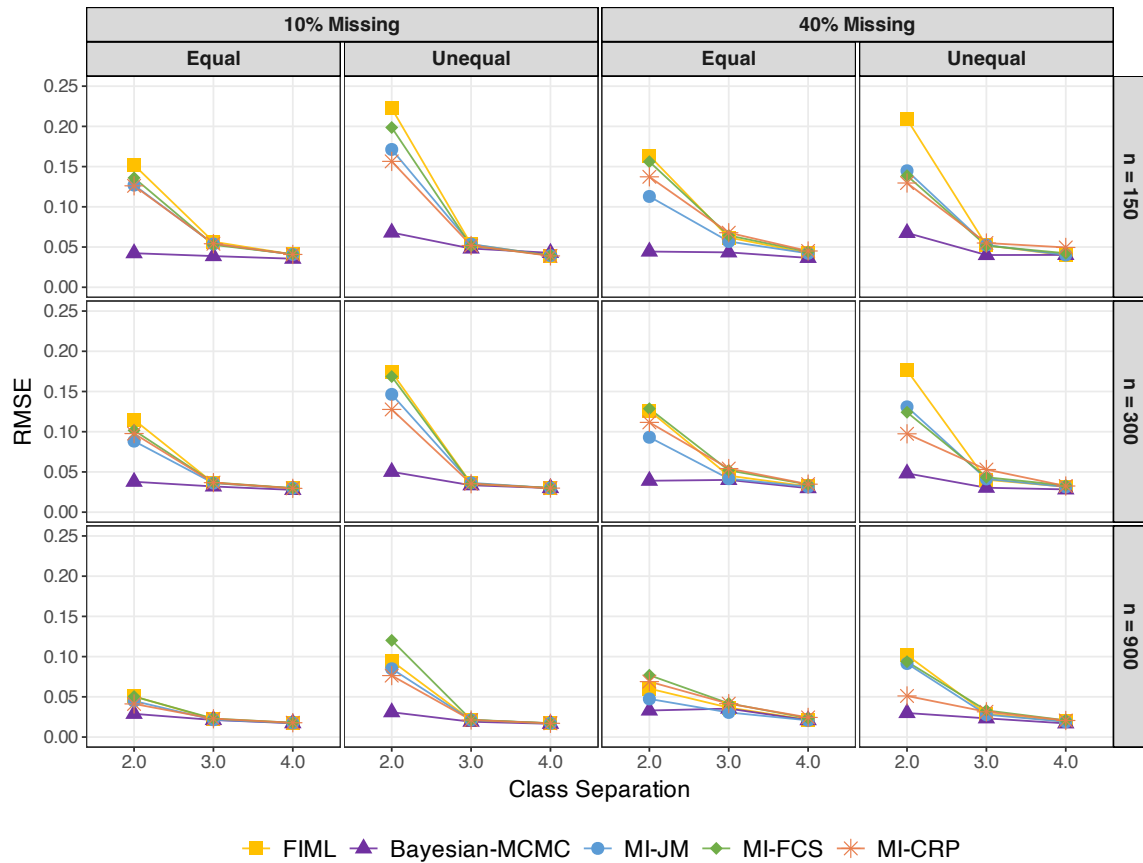
I first consider conditions with 10% missing data and equal class proportions. At low class separation (MD = 2.0), the five methods differed substantially in both their median accuracy and variability. Bayesian-MCMC achieved the highest median CCAR, with a relatively compact interquartile range. FIML produced a comparable yet slightly lower median, but its distribution was much wider, with numerous outliers falling below 50%. This greater variability for FIML indicates that, although FIML can recover class membership reasonably well on average, individual replica-

Figure 4.16: Root Mean Square Error for Growth Factor Variances and Covariance at MD = 4.0.



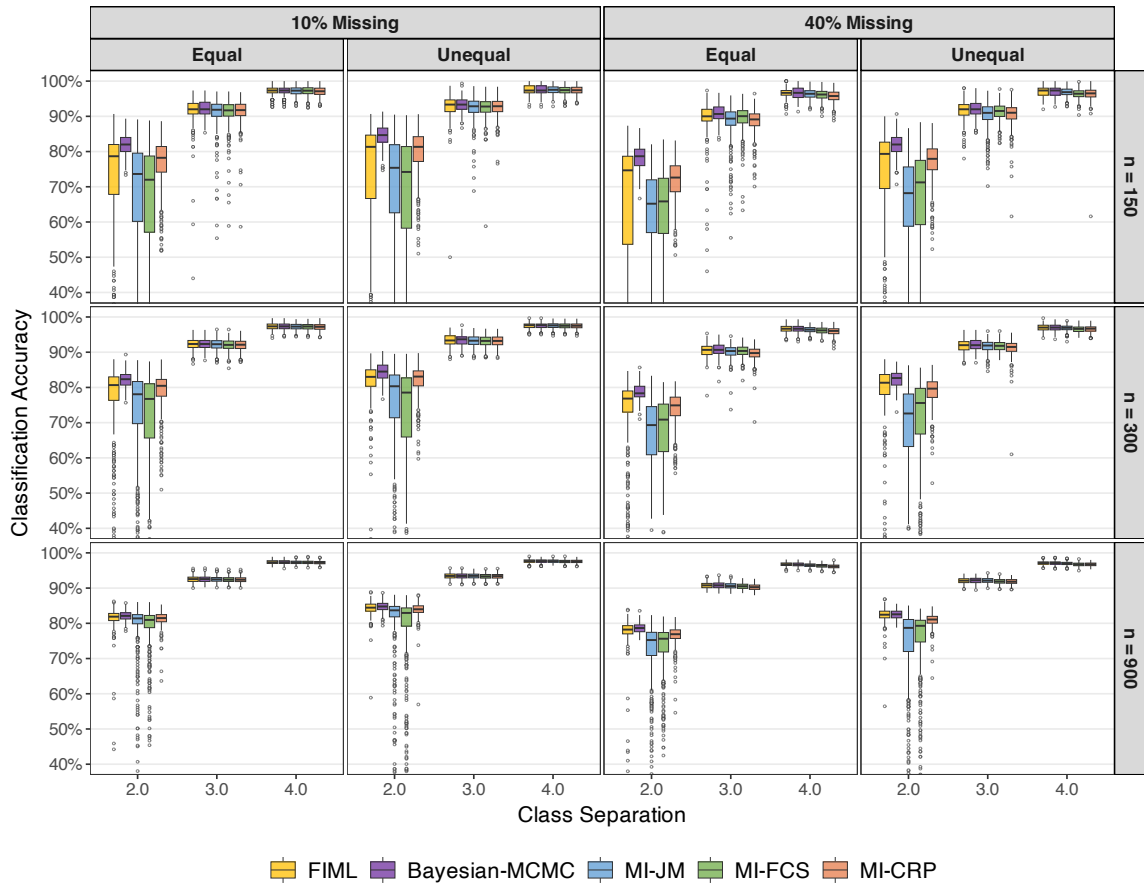
tions sometimes yielded poor classification. MI-JM and MI-FCS both exhibited the lowest median accuracy and even wider interquartile ranges than FIML. The proposed MI-CRP achieved a median CCAR close to that of FIML with a tighter distribution than FIML, MI-JM, and MI-FCS. At moderate class separation (MD = 3.0), all five methods yielded comparable accuracy with their medians above 90%. Although some outliers remained, the boxplots were considerably more compact than at MD = 2.0. At high class separation (MD = 4.0), the methods were nearly indistinguishable, with medians reaching closer to 100% and very narrow interquartile ranges. Under unequal class proportions, the overall pattern was similar to the equal-proportion condition. At MD = 2.0, Bayesian-MCMC again led with the highest median among the methods, followed by MI-CRP and FIML, then MI-JM and MI-FCS. Regarding the variability, Bayesian-MCMC was the tightest, followed by MI-CRP, and FIML, MI-JM, and MI-FCS had the widest variability. As class separation increased to MD = 3.0 and 4.0, the methods performed similarly, and the differences observed at MD = 2.0 largely disappeared.

Figure 4.17: Root Mean Square Error for Class Proportions.



When the missing data proportion increased to 40% under equal class proportions, accuracy declined across all methods. When $MD = 2.0$, Bayesian-MCMC remained the most accurate and retained a compact distribution. FIML showed a lower median with a further widened interquartile range, and its first quartile reached near 50%. MI-JM and MI-FCS again had the lowest medians, near 65%, but were less variable than FIML. MI-CRP showed a median slightly below that of FIML, but its distribution remained more concentrated than those of FIML, MI-JM, and MI-FCS. At $MD = 3.0$ and 4.0 , all methods showed increasing median accuracies and the outliers decreased. When the class proportions were unequal, the general patterns were similar to those of equal proportions. However, overall classification performance was better, particularly at $MD = 2.0$. At this separation level, the notable difference compared to the equal class proportion condition was that FIML had a less variable interquartile range. At higher class separation levels, differences narrowed at $MD = 3.0$ and were minimal at $MD = 4.0$.

Figure 4.18: Correct Class Assignment Rates Across Simulation Conditions.



Medium Sample Size ($n = 300$)

Increasing the sample size to $n = 300$ generally improved classification accuracy and reduced variability for all methods. Such changes were more visible at low class separation ($MD = 2.0$). Under 10% missing data with equal proportions and $MD = 2.0$, Bayesian-MCMC maintained its predominance, followed by FIML and MI-CRP. Here, both FIML and MI-CRP had comparable variability but MI-CRP had fewer outliers than FIML. MI-JM and MI-FCS improved relative to the $n = 150$ condition but still trailed the other methods. At $MD \geq 3.0$, all methods achieved median accuracy above 90%, with minimal differences. The pattern under unequal proportions was similar, with MI-CRP performing on par with FIML and close to Bayesian-MCMC at $MD = 2.0$. With 40% missing data, the same ordering held but the gap among the methods was somewhat narrower than at $n = 150$, and FIML showed much less variability but still had many outliers. For $MD \geq 3.0$, all methods converged toward comparable accuracy.

Large Sample Size ($n = 900$)

At the largest sample size, accuracy improved further and the differences among methods diminished. The patterns under different class proportion conditions were similar within each missing data proportion condition. The main difference between the two missing data conditions was that under 40% missing data, MI-JM and MI-FCS showed greater variability compared to 10% missing data. Furthermore, MI-JM and MI-FCS produced more outliers than the other three methods across both missing data conditions, indicating that a fair number of replications still yielded poor classification. However, the median accuracies were more similar across the methods than at smaller sample sizes. It is worth noting that Bayesian-MCMC yielded the fewest converged replications under 40% missing data with unequal proportions, meaning that its accuracy estimates were based on a smaller and potentially selective subset of replications. This may have contributed to its consistently high median accuracy under this condition, as replications that failed the PSR convergence criterion were excluded. At higher class separation levels, all methods performed very similarly, and the differences were minimal.

Summary of Classification Accuracy

Class separation was the strongest determinant of CCAR, and differences among methods emerged primarily at $MD = 2.0$ and diminished as separation increased. Smaller sample sizes and higher missing data proportions amplified these differences. A consistent ordering emerged under the most challenging conditions. Bayesian-MCMC achieved the highest and most stable accuracy, due to informative and accurate priors for the growth factor means. FIML also produced high median accuracy but exhibited considerably wider variability across replications. MI-JM and MI-FCS produced the lowest accuracy and the widest variability. Under these conditions, MI-CRP consistently outperformed MI-JM and MI-FCS and approached the accuracy of Bayesian-MCMC and FIML with more stable performance than FIML.

4.6 Discussion

With GMMs, researchers can obtain a more nuanced account of interindividual differences in intraindividual change by identifying latent classes of growth trajectories. Importantly, the quality of inferences drawn from GMMs depends on how missing data are handled. Despite the recognized importance of this issue, the methodological landscape for addressing missing data in GMMs has remained incomplete. This chapter proposed a Bayesian nonparametric multiple imputation approach using the Chinese restaurant process algorithm, which I termed MI-CRP, for handling missing data in GMMs. Three methodological considerations motivated this work: (a) retaining all observed data rather than discarding incomplete cases, (b) mitigating the potential circularity that arises when the latent class structure must be specified as part of the missing data treatment, and (c) generating multiply imputed data sets

from an imputation model that accounts for population heterogeneity. To examine the performance of the proposed MI-CRP approach, I conducted a Monte Carlo simulation comparing MI-CRP with four existing methods: FIML, Bayesian estimation via MCMC, MI-JM, and MI-FCS.

The results revealed four broad findings. First, convergence difficulties and degenerate solutions were concentrated at low class separation ($MD = 2.0$), where valid replication rates dropped below 50% for MI-JM and MI-FCS at the smallest sample size with 40% missing data. MI-CRP achieved the highest convergence and valid replication rates among the MI methods at smaller sample sizes, and Bayesian-MCMC showed a distinctive pattern in which convergence did not improve with larger samples at this separation level. Second, class separation and sample size were the dominant design factors influencing parameter recovery and classification accuracy across all methods. Third, when compared with non-MI methods (Bayesian estimation via MCMC and FIML), MI-CRP achieved comparable parameter recovery for the growth factor means and class proportions. Classification accuracy also approached that of these non-MI methods, but was slightly lower than that of Bayesian-MCMC, which benefited from informative prior specifications, and less variable than FIML at low class separation and smaller sample sizes. Fourth, among the MI-based methods, MI-CRP consistently outperformed MI-JM and MI-FCS in classification accuracy, with higher median accuracy and less variability. MI-CRP also avoided the extreme covariance bias that characterized MI-FCS and MI-JM, though it did not produce uniformly lower bias or RMSE across all parameters. I discuss these findings and situate this work in the extant literature below.

4.6.1 Convergence Difficulties and Degenerate Solutions

Convergence difficulties were largely concentrated in the low class separation condition ($MD = 2.0$), where the latent classes overlapped most and the estimation problem was most challenging. At this level, MI-JM and MI-FCS produced the lowest valid replication rates, dropping below 50% at $n = 150$ with 40% missing data (see supplementary materials). These two methods impute missing values without accounting for population heterogeneity, and when the resulting completed data sets are submitted to a two-class GMM, the imputed values may not adequately reflect the between-class differences needed for the mixture model to converge. MI-CRP, by contrast, maintained higher convergence rates than MI-JM and MI-FCS under the same conditions, suggesting that preserving population heterogeneity during imputation can facilitate convergence of the downstream analytic model.

A distinctive pattern emerged for Bayesian-MCMC at $MD = 2.0$. Unlike the other methods, whose convergence generally improved with larger sample sizes, Bayesian-MCMC convergence did not improve and in some cases worsened as n increased. At $n = 900$ with unequal class proportions and 40% missing data, Bayesian-MCMC convergence fell to approximately 30%. When I checked the *Mplus* output files, those output files had messages saying that parameters were not able to be

generated satisfying the constraints (set for label switching). This practically means that the posteriors for the growth factor means overlap a lot. This suggests that even with larger sample sizes, the plausible values generated from the Bayesian estimation might not be adequate for the model to be estimated normally if class separation is low. Given that convergence rates increased with larger sample sizes even in low class separation conditions when the correct model was fit (Depaoli, Heo, et al., 2026), this finding suggests that attrition-pattern missingness introduces additional estimation difficulties that persist even when informative priors for the growth factors are specified. FIML, by contrast, maintained relatively high convergence rates throughout. This contrasts with the finding of D. Y. Lee and Harring (2023), where FIML produced relatively low convergence rates, falling as low as 39% at the smallest sample size ($n = 100$). Given that D. Y. Lee and Harring (2023) employed a non-attrition missing data pattern, this discrepancy suggests that the missing data pattern itself may influence FIML convergence behavior, with attrition-pattern missingness appearing to be more conducive to convergence than non-monotone patterns.

Among the converged solutions, degenerate cases in which one class collapsed to fewer than 5% of the sample were excluded from subsequent analyses following the 5% threshold of Nasserinejad et al. (2017). These cases were mostly confined to $MD = 2.0$. In this simulation, the smallest true class proportion was 30% under the unequal condition, making it straightforward to distinguish degenerate solutions from the intended class structure. In applied settings, however, the boundary between a degenerate solution and a genuinely small latent class is far less clear. E. Kim et al. (2026) recently investigated this issue in a comprehensive simulation examining GMM performance in detecting small latent classes. Their results showed that when GMM assumptions were satisfied, specifically normality within class and correctly specified linear growth trajectories, a well-separated small class ($MD = 6.0$) could be reliably detected even at proportions as low as 1% to 3%. On the other hand, recovery of small classes was considerably more difficult under more moderate separation ($MD = 3.0$). E. Kim et al. (2026) thus cautioned that applying a fixed cutoff to exclude small classes carries different risks depending on the degree of class separation. When separation is large, excluding a small class may discard a substantively meaningful subgroup; when separation is moderate or low, small classes may be difficult to distinguish from estimation artifacts, and researchers may be justified in their exclusion decisions. In this chapter, the class separation conditions of $MD = 2.0$ to 4.0 fall within the range where such justification is reasonable. However, I emphasize that the decision to retain or exclude a small class should not rest on a numerical cutoff alone. Substantive and theoretical considerations must inform this judgment (see a similar sentiment in Heo, Depaoli, et al., 2024).

4.6.2 Recovery of Growth Factor Parameters

The intercept means were well recovered by all five methods across nearly all conditions, a finding consistent with the broader GMM literature in which intercept

parameters tend to be the most stable (Depaoli, 2013; D. Y. Lee & Harring, 2023). The slope means proved more challenging. Bias was most pronounced at the smallest class separation ($MD = 2.0$) with small sample sizes ($n = 150$), high missing data proportions (40%), and unequal class proportions. Under these conditions, the effective number of observed data points for the smaller class was severely limited by attrition, leaving little information from which to recover the slope trajectory. That parameters other than the intercepts are harder to recover aligns with the findings of D. Y. Lee and Harring (2023). All methods except Bayesian-MCMC exceeded the ± 0.1 threshold for the slope means at $MD = 2.0$ with small samples. The favorable performance of Bayesian-MCMC on these parameters was facilitated by the informative priors specified for the growth factor means. D. Y. Lee and Harring (2023) noted that priors for the means did not substantially alter the recovery of the mean parameters themselves. Yet, prior specifications for different sets of parameters can interact in ways that jointly influence estimation, a point I return to below.

The growth factor covariance proved to be the single most difficult parameter to recover across all methods. MI-FCS exhibited persistent and large negative bias on this parameter, reaching values that exceeded the plotted range of the figures, particularly at smaller sample sizes. Because MI-FCS imputes each variable conditionally without a joint distributional model, it does not directly preserve the covariance structure among the growth factors, and this limitation becomes more pronounced when the data contain mixture heterogeneity that the conditional models do not accommodate. MI-JM showed a complementary pattern, with positive covariance bias that grew with increasing missing data proportion. MI-CRP also showed positive covariance bias, especially at $MD = 3.0$ and 4.0 , though the magnitudes were generally smaller than those of MI-JM. That covariance parameters are inherently difficult to recover in GMMs is well documented. Depaoli (2013) reported that variance and covariance parameters were not well recovered even without missing data, and D. Y. Lee and Harring (2023) similarly found that the intercept-slope covariance showed bias across methods.

Bayesian-MCMC achieved the best parameter recovery overall, with near-zero relative bias for the growth factor means and the lowest RMSE across nearly all parameters. This strong performance, however, was obtained under informative priors centered at the true population values. I specified these priors to maximize estimation efficiency, following the approach of Depaoli (2013) and Heo, Depaoli, et al. (2024). In practice, researchers rarely have access to the true population values, and the degree to which informative priors improve estimation depends on how accurately the prior hyperparameters reflect the underlying parameters (Depaoli, 2013, 2014). Indeed, D. Y. Lee and Harring (2023) noted the importance of prior specification in Bayesian estimation of GMMs with missing data, and the results should be understood in light of the prior settings.

Among the MI-based approaches, no single method uniformly dominated the others in parameter recovery. MI-CRP did not produce consistently lower bias or RMSE than MI-JM and MI-FCS across all parameters. One consideration here is

that valid replication rates differed across methods (Table 4.3), so parameter recovery was not evaluated on the same set of replications. MI-JM and MI-FCS retained fewer valid replications at low class separation with smaller sample sizes, and the replications that survived may represent conditions more favorable to estimation. That said, what MI-CRP did achieve was a more balanced profile. MI-CRP avoided the extreme negative covariance bias of MI-FCS and the growing positive covariance bias of MI-JM, while maintaining bias levels for the means and class proportions that were generally comparable to those of the other methods. When this parameter recovery profile is considered alongside the classification accuracy results, MI-CRP can be deemed a method that trades modest imprecision in certain variance-covariance parameters, which are inherently difficult to recover, for gains in classification accuracy.

4.6.3 Classifying Individuals While Handling Missingness

A central finding of this chapter is that MI-CRP consistently outperformed MI-JM and MI-FCS in classification accuracy across the range of conditions examined, with the advantage most pronounced under challenging conditions of low class separation, small sample sizes, and high missing data proportions. The ordering among methods was most evident at $MD = 2.0$. Under these conditions, Bayesian-MCMC achieved the highest and most stable accuracy, followed by FIML, which produced slightly lower medians but considerably wider variability across replications. MI-CRP produced slightly lower median accuracy than FIML but with notably less variability, offering more consistent classification performance across replications, while outperforming MI-JM and MI-FCS at $n = 150$. As class separation increased to $MD \geq 3.0$, all methods converged toward similarly high accuracy levels, and with larger sample sizes, classification accuracy increased overall.

The classification advantage of MI-CRP over MI-JM and MI-FCS can be understood through the lens of imputation model compatibility. Standard MI approaches do not account for the mixture structure underlying the data, effectively smoothing over the between-class differences that GMMs are designed to detect. When the completed data sets are then submitted to a mixture model, the imputed values may blur the boundaries between classes, leading to reduced classification accuracy. MI-CRP addresses this by allowing the imputation model to discover latent clusters through the CRP, generating imputed values that respect subgroup boundaries. This reasoning is consistent with the broader literature on nonparametric Bayesian methods for mixture data. Si and Reiter (2013) demonstrated that a DPMM for multiple imputation of categorical data produced more accurate estimates and better coverage than standard chained equations imputation, precisely because the nonparametric model captured complex dependencies that the parametric model missed. Qiu et al. (2025) showed that a CRP-based approach to latent class analysis correctly identified the true number of classes across a range of conditions with accurate parameter recovery, further supporting the capacity of Bayesian nonparametric models to accommodate

latent heterogeneity. These results extend these findings to the specific problem of missing data in GMMs and demonstrate that the benefits of nonparametric imputation carry over to classification accuracy in the growth mixture context. These MI-CRP results suggest that the limitation of standard MI documented by D. Y. Lee and Harring (2023) can be substantially mitigated by incorporating a heterogeneity-preserving imputation model.

I additionally note that the classification advantage of the non-MI approaches (Bayesian-MCMC and FIML) should be interpreted with an important caveat. Both methods fit the correct two-class GMM directly to the incomplete data, meaning that the analytic model matches the data-generating process. In applied settings where the true number of classes is unknown, researchers must determine C through model comparison, and the resulting class structure may be influenced by the missing data pattern. MI-CRP offers a different approach in this regard. Because the imputation model does not require the number of classes to be specified, researchers can first generate completed data sets that respect population heterogeneity and then explore the class structure on these completed data without committing to a particular class solution during the imputation phase.

4.6.4 Implications for Applied Practice

The findings in this chapter raise several considerations for applied researchers. One important question is how to handle missing values when the latent class structure is not yet known. FIML and Bayesian estimation via MCMC, which address incomplete data as part of the model fitting process, performed well in the simulation and may be preferred when the analytic model is well specified. This aligns with broader findings in structural equation modeling that, under correct model specification, FIML yields estimates close to those from complete-data analysis (T. Lee & Shi, 2021). However, both approaches require the researcher to specify the number of latent classes prior to estimation, and Bayesian-MCMC additionally requires careful prior specification. In applied settings, the number of classes is rarely known in advance, and the process of determining it often involves fitting and comparing multiple models with different class solutions. When missing data are present, this model comparison process may itself be influenced by how the incomplete data are handled, introducing a potential circularity between class enumeration and missing data treatment. Moreover, prior specification for Bayesian-MCMC may not be intuitive and requires careful prior elicitation. When the research design calls for completed data sets, for instance, to support follow-up analyses beyond the primary GMM, MI-based approaches have an additional practical benefit.

Inference from multiply imputed data sets provides an alternative by separating the treatment of missing data from the analytic model. In that regard, the results demonstrate that the choice of imputation model matters considerably. MI-JM and MI-FCS, which do not account for population heterogeneity during imputation, produced lower classification accuracy and wider variability than the other methods.

The resulting completed data sets may obscure meaningful subgroup differences. MI-CRP addresses this limitation by accommodating population heterogeneity through the Chinese restaurant process algorithm in the imputation phase. Without requiring the researcher to pre-specify the number of mixture components, MI-CRP flexibly captures subgroup structure in the data during imputation.

MI-CRP may provide further advantages that go beyond what the raw accuracy numbers suggest. Unlike FIML, MI-CRP exhibited less variability in classification performance across replications. Unlike Bayesian-MCMC, MI-CRP does not require the specification of informative priors for model parameters of interest. And unlike both non-MI approaches, MI-CRP does not require the number of latent classes to be fixed before missing data are addressed, reducing the risk of misspecification at the outset. Moreover, compared to standard MI-based methods (MI-JM and MI-FCS), MI-CRP produced comparable bias for the growth factor means and achieved consistently higher classification accuracy with less variability. From a practical standpoint, these results suggest that when researchers plan to use MI as a preprocessing step before fitting GMMs, adopting an imputation model that can accommodate population heterogeneity is preferable. Thus, unless researchers are confident that their analytic model is correctly specified or are in a position to specify realistic informative priors for Bayesian estimation, MI-CRP serves as a more practical alternative that preserves the heterogeneity characteristics of the data during imputation. Therefore, MI-CRP provides a viable approach that balances flexibility with practical applicability.

4.6.5 Open Questions and Future Directions

The MI-CRP imputation model used in this chapter assumed a linear growth structure within each cluster. In longitudinal research, linear growth is often a reasonable starting point for modeling change over time (Grimm et al., 2016), making this choice well grounded and realistic. A natural extension would be to evaluate MI-CRP in more complex settings, such as nonlinear growth mixtures, models with covariates or distal outcomes, and cross-sectional mixture modeling frameworks (e.g., Depaoli, 2013; Heo, Depaoli, et al., 2024; Petras & Masyn, 2010; Qiu et al., 2025; Tueller & Lubke, 2010). Examining how MI-CRP performs in the presence of outliers (S. Kim et al., 2022; Lu & Zhang, 2014) or under models with different latent covariance structures (McNeish & Harring, 2021; McNeish, Harring, & Bauer, 2023) would further clarify the boundaries of the approach.

Another important direction concerns the type of outcome variable. Although MI-CRP was developed and evaluated here for continuous data within the GMM framework, the underlying Bayesian nonparametric approach applies more broadly. MI-CRP can be adapted to categorical variables, including binary, ordinal, and nominal items, beyond the growth mixture setting. Bayesian nonparametric models have been shown to perform well in categorical multiple imputation (Si & Reiter, 2013), lending support to this extension. For instance, the distribution of binary responses may shape imputation performance. More balanced patterns (e.g., 50/50) are ex-

pected to pose less challenge than highly imbalanced ones (e.g., 80/20 or 90/10), in which sparse cells may strain the imputation model. Investigating MI-CRP across these response patterns would be a meaningful next study.

The simulation generated missing data under a MAR mechanism with an attrition pattern. MAR is a working assumption in many applied longitudinal studies (e.g., Peeters et al., 2015). In practice, real data often involve nonignorable missing data or combinations of different missing data mechanisms (Enders, 2011; Graham, 2009), and ongoing research continues to address these complexities (Gomer & Yuan, 2023). Even within the same mechanism, different subtypes may exist (Gomer & Yuan, 2021). Whether the proposed Bayesian nonparametric MI approach proves additionally useful or encounters new limitations under alternative mechanisms and subtypes, for both parameter recovery and classification purposes, remains a promising area with considerable room for investigation.

Beyond parameter recovery and classification accuracy, which represent traditional topics in GMM research, another compelling direction involves class enumeration (Depaoli, Heo, et al., 2026; Nylund et al., 2007; Tofighi & Enders, 2007). How a nonparametric imputation model influences the determination of the number of latent classes in the presence of missing data is not yet known. One related study by Depaoli, Qiu, et al. (2026) compared the DPMM and the mixture of finite mixtures model to traditional information criteria for class selection in GMMs, but without missing data. One theoretical consideration is whether the DPMM’s adaptive cluster discovery during imputation may yield additional clusters beyond those corresponding to the substantive classes of the analysis model, with implications for class enumeration in subsequent analyses that warrant future investigation. Given that missing data are pervasive in longitudinal research, extending the investigation of class enumeration to the missing data context is a complement to this line of research.

Finally, this chapter used the CRP construction as one example of the Bayesian nonparametric algorithm. There is a range of Bayesian nonparametric algorithms beyond the CRP, as introduced in Gershman and Blei (2012), that could potentially be adapted for missing data handling. An alternative construction could include, for instance, the stick-breaking process (Sethuraman, 1994), which offers a different computational approach and has been applied in handling missing data by Si and Reiter (2013). Whether these alternative constructions offer computational or inferential advantages for multiple imputation in the GMM and other mixture modeling contexts is an open question worth investigating.

4.6.6 Concluding Remarks

Handling missing data in the context of GMMs has remained a methodological dilemma. I addressed this dilemma by proposing MI-CRP, a Bayesian nonparametric multiple imputation approach that retains all available data, mitigates the potential circularity of specifying the latent class structure during missing data treatment, and accommodates population heterogeneity during the imputation phase so that uncer-

tainty about missing values is properly propagated into downstream analyses. More broadly, this chapter contributes to a growing body of work that seeks to advance the utility of GMMs through principled methodological innovation. The proposed framework holds potential not merely for handling missing data in GMMs but for applications in complex modeling settings beyond GMMs. I anticipate that this work will encourage continued advancement and development in this domain.

Chapter 5

General Discussion

Two practical gaps have long remained in the implementation of growth mixture modeling: a lack of principled guidance for selecting predictors of heterogeneous growth trajectories, and a shortage of tools well-suited for handling missing data in the presence of such heterogeneity. This dissertation developed and evaluated advanced Bayesian methods to meet both challenges. Study 1 proposed Bayesian variable selection via shrinkage priors for predictors of class membership and within-class growth factors. Study 2 developed multiple imputation via Bayesian nonparametrics grounded in the Chinese restaurant process. Having presented the two studies in their own chapters, I now synthesize their findings from a broader perspective. I articulate the contributions of this dissertation and reflect on limitations and prospects. I close with an outlook on the dissertation as a launching point for sustained innovation in Bayesian latent variable mixture modeling.

5.1 Contributions

This dissertation contributes to the growth mixture modeling literature along methodological, empirical, practical, and broader disciplinary lines.

Methodologically, this dissertation introduced two advanced Bayesian methods for growth mixture modeling, each targeting one of the two implementation gaps described earlier. In the absence of an established practice for predictor selection in growth mixture modeling, Study 1 proposed a Bayesian variable selection framework via shrinkage priors that jointly operates on predictors of latent class membership and predictors of within-class growth factors. This framework enables a unified treatment of predictor integration in growth mixture models (GMMs). Study 2 developed multiple imputation via the Chinese restaurant process (MI-CRP), a Bayesian nonparametric approach to handling missing data that does not require the number of latent classes to be pre-specified and respects population heterogeneity during imputation. The approach separates the treatment of missing data from the determination of the latent class structure, which resolves the circularity that arises when incomplete data influence class enumeration during model fitting. The two methods share a con-

ceptual thread, in that both deploy specific priors, sparsity-inducing priors in Study 1 and an infinite-dimensional Dirichlet process prior in Study 2, to extend Bayesian thinking from parameter estimation or model selection to implementation problems that have long constrained applied work.

Empirically, the simulation studies provided comparative evidence previously absent in the GMM literature. Study 1 evaluated seven shrinkage priors (ridge, lasso, hyperlasso, elastic net, horseshoe, regularized horseshoe, and spike-and-slab) under a single set of conditions, generating the first comprehensive head-to-head benchmark of Bayesian variable selection in the GMM literature. Study 2 placed the proposed MI-CRP approach alongside four existing missing data methods, namely full information maximum likelihood (FIML), Bayesian estimation via Markov chain Monte Carlo (MCMC), multiple imputation via joint modeling, and multiple imputation via fully conditional specification. On the predictor selection side, the horseshoe, regularized horseshoe, and spike-and-slab consistently struck the strongest balance between detecting nonzero effects and shrinking null effects. The regularized horseshoe, spike-and-slab, and horseshoe showed superior predictive accuracy, though the regularized horseshoe and spike-and-slab came at the cost of lower convergence rates. Classification accuracy, in contrast, was similar across all seven priors. On the missing data side, MI-CRP avoided the extreme covariance bias that existing multiple imputation methods exhibited. MI-CRP also consistently outperformed the other multiple imputation methods in classification accuracy. Its performance approached that of Bayesian estimation via MCMC (with informative priors) and FIML, with less variability than FIML. Beyond these study-specific highlights, several patterns cut across the two studies. Sample size was a consistent driver of methodological differences. Class separation and class proportion mattered in different ways depending on the outcome examined. Each study surfaced its own distinguishing conditions, where predictor effect pattern and predictor intercorrelation shaped differences among priors in Study 1, and missing data proportion shaped differences among missing data methods in Study 2. Knowing where the methods yielded specific results and how those results differ across factors affords a detailed view that guides both future comparative simulation research and applied researchers in weighing these intricacies against their own data and method choices.

Practically, both studies were designed with applied researchers in mind. Study 1 translated its findings into concrete recommendations for prior choice. Study 2 gave similar guidance on when MI-CRP is likely to improve upon existing approaches. Both methods were implemented in open-source software, **JAGS** for Study 1 and **NIMBLE** for Study 2. This choice ensures that applied researchers can access and adapt these methods for their own research without proprietary barriers. The methods presented in this dissertation can be further disseminated through tutorials and primers. Together, the methodological innovations of this dissertation are not intended to remain confined to simulation. They are paired with practical scaffolding so that researchers can adopt, modify, and extend them.

At a broader disciplinary level, this dissertation illustrates how Bayesian think-

ing can resolve implementation problems that have persisted in GMM research. Shrinkage priors reframe predictor selection as a regularization of coefficients within a single model. Dirichlet process (DP) priors reframe missing data handling as a flexible generative process that allows the data to speak about heterogeneity during multiple imputation. In both approaches, Bayesian reasoning moves beyond parameter estimation or model selection to draw on broader methodological frameworks, including Bayesian variable selection and Bayesian nonparametrics. Such applications are not confined to the two methods developed here, and they can continue to extend the latent variable mixture modeling framework.

5.2 Limitations and Prospects

I acknowledge three limitations of this dissertation and point out elements that reflect the conceptual posture of the work. First, the methods address implementation gaps within the GMM paradigm but do not inform when mixture modeling itself is the right approach for a given research question. In other words, the methods operate under the working assumption that discrete latent classes characterize population heterogeneity, but they provide no diagnostic for whether that assumption holds in a specific application. Researchers must still determine, through substantive reasoning or external validation, whether a discrete-class view of heterogeneity is appropriate for their research question. Constructive responses have shown that, with a strong theoretical basis and expert-elicited priors on class trajectories, GMMs can recover theoretically meaningful classes that would otherwise be missed (van de Schoot et al., 2017, 2018).

Second, the priors examined in this dissertation shift, rather than eliminate, the burden of prior specification. For example, shrinkage priors require the specification of hyperparameters, the DP prior requires a concentration parameter, and other model parameters require their own priors. Although guidance can be drawn from prior work, from the settings used in this dissertation, or from other sources, responsibility for principled specification ultimately rests with the researcher. For shrinkage priors such as the regularized horseshoe, previous work has offered some guidance on setting its hyperparameters (Piironen & Vehtari, 2017), but a comprehensive examination of these priors and their possible interactions across mixture models remains little understood.

Third, Bayesian mixture modeling is inherently complex. As GMM formulations grow more intricate and interact with particular characteristics of the data (e.g., highly unbalanced class proportions with small minority classes, low class separation), computational demands compound. This may further limit how readily the methods introduced here can be implemented. This dissertation has focused on commonly encountered implementation problems, with more elaborate formulations and varied data characteristics remaining for future investigation. More fundamentally, ensuring stable estimation across the idiosyncratic features of any given dataset will require continued attention.

These limitations point to directions that future methodological innovations should pursue. As model complexity grows, the rising computational burden calls for algorithmic strategies that can keep Bayesian growth mixture modeling within practical reach. Advances in algorithmic efficiency such as variational Bayes represent promising approaches for making such modeling tractable at scale (Blei & Jordan, 2006; Blei et al., 2017). In addition, whether such algorithmic advances can extend Bayesian latent variable mixture modeling to so-called big data (Adjerid & Kelley, 2018) stands as a major challenge for the field.

The types of data on which latent variable mixture models can be applied are also changing rapidly. Traditional psychological datasets are increasingly complemented by new sources. These include, but are not limited to, network data (Heo, Simons, & Liu, 2025; H. Liu et al., 2018; Xu & Zhang, 2025), text data (H. Liu, Tsang, et al., 2025; Wilcox et al., 2023), intensive longitudinal data (Ernst et al., 2024; Hamaker, 2025), among others. Whether Bayesian latent variable mixture frameworks can accommodate these emerging data types, and if so, what they imply for applied and methodological research, marks an important frontier.

In parallel, machine learning, deep learning, and large language models are continuously reshaping the methodological landscape (Demszky et al., 2023; Jacobucci et al., 2023; Pargent et al., 2023; Tong & Zhang, 2026; Urban & Gates, 2021; Wyman & Zhang, 2025). Integrating these tools with Bayesian mixture modeling can shape the entire workflow, from prior specification to posterior inference (for examples in other modeling contexts, see Sekulovski et al., 2026). Realizing their full potential within Bayesian mixture modeling, with the methodological rigor it demands, remains in the hands of researchers as a compelling prospect.

A final prospect concerns the intersection between methodological advances and cumulative theory development. It has been emphasized that theoretical considerations should anchor the implementation of Bayesian GMMs (Heo, Depaoli, et al., 2024; van de Schoot et al., 2017, 2018). As open science practices are increasingly adopted simultaneously with ongoing methodological advances (Pfadt et al., 2025), substantive theories can be bolstered or face closer scrutiny. Realizing this potential calls for a structured framework that can guide cumulative theory development within Bayesian latent variable mixture modeling (for a related discussion, see Heo & Jeon, 2026).

5.3 Outlook

This dissertation has advanced methods for the implementation of Bayesian growth mixture modeling. I envision the methods developed and evaluated here as a launching point for sustained innovation in applied and methodological research. From this foundation, Bayesian growth mixture modeling and, more broadly, latent variable mixture modeling are poised to keep pace with the evolving demands of psychological science and other academic disciplines.

References

- Adjerid, I., & Kelley, K. (2018). Big data in psychology: A framework for research advancement. *American Psychologist*, 73(7), 899–917. <https://doi.org/10.1037/amp0000190>
- Agresti, A. (2012). *Categorical data analysis* (3rd). John Wiley & Sons.
- Aldous, D. J. (1985). Exchangeability and related topics. In P. L. Hennequin (Ed.), *École d'été de probabilités de Saint-Flour XIII—1983* (pp. 1–198, Vol. 1117). Springer. <https://doi.org/10.1007/BFb0099421>
- Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 2(6), 1152–1174. <https://doi.org/10.1214/aos/1176342871>
- Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 243–277). Psychology Press.
- Asparouhov, T., & Muthén, B. (2010). Multiple imputation with Mplus [Technical report]. <https://www.statmodel.com/download/Imputations7.pdf>
- Asparouhov, T., & Muthén, B. (2021). Advances in Bayesian model fit evaluation for structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(1), 1–14. <https://doi.org/10.1080/10705511.2020.1764360>
- Bain, C., & Shi, D. (2024). Investigating variable selection techniques under missing data: A simulation study. In H. Hwang, H. Wu, & T. Sweet (Eds.), *Quantitative psychology* (pp. 109–119). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-55548-0_11
- Bainter, S. A., McCauley, T. G., Fahmy, M. M., Goodman, Z. T., Kupis, L. B., & Rao, J. S. (2023). Comparing Bayesian variable selection to lasso approaches for applications in psychology. *Psychometrika*, 88(3), 1032–1055. <https://doi.org/10.1007/s11336-023-09914-9>
- Baltes, P. B., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 1–39). Academic Press.
- Barboza, G. E. (2020). Child maltreatment, delinquent behavior, and school factors as predictors of depressive symptoms from adolescence to adulthood: A growth mixture model. *Youth & Society*, 52(1), 27–54. <https://doi.org/10.1177/0044118X17721803>

- Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, 42(4), 757–786. <https://doi.org/10.1080/00273170701710338>
- 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. <https://doi.org/10.1037/1082-989X.8.3.338>
- Bhadra, A., Datta, J., Polson, N. G., & Willard, B. (2017). The horseshoe+ estimator of ultra-sparse signals. *Bayesian Analysis*, 12(4), 1105–1131. <https://doi.org/10.1214/16-BA1028>
- Bhadra, A., Datta, J., Polson, N. G., & Willard, B. (2019). Lasso meets horseshoe: A survey. *Statistical Science*, 34(3), 405–427. <https://doi.org/10.1214/19-STS700>
- Bibriescas, N., & Whittaker, T. A. (2023). Detecting latent classes through mediation in regression mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, 30(3), 449–457. <https://doi.org/10.1080/10705511.2022.2137027>
- Blei, D. M., & Jordan, M. I. (2006). Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, 1(1), 121–143. <https://doi.org/10.1214/06-BA104>
- Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518), 859–877. <https://doi.org/10.1080/01621459.2017.1285773>
- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. John Wiley & Sons. <https://doi.org/10.1002/0471746096>
- Caesens, G., Morin, A. J. S., & Stinglhamber, F. (2020). Longitudinal trajectories of perceived organizational support: A growth mixture analysis. *Journal of Managerial Psychology*, 35(6), 481–495. <https://doi.org/10.1108/JMP-01-2020-0027>
- 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. <https://doi.org/10.1002/sim.3915>
- Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2), 465–480. <https://doi.org/10.1093/biomet/asq017>
- Cassiday, K. R., Cho, Y., & Harring, J. R. (2021). A comparison of label switching algorithms in the context of growth mixture models. *Educational and Psychological Measurement*, 81(4), 668–697. <https://doi.org/10.1177/0013164420970614>
- Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods & Research*, 29(4), 468–508. <https://doi.org/10.1177/0049124101029004003>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>

- Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Wiley. <https://doi.org/10.1002/9780470567333>
- Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development, 11*(2), 121–136. <https://doi.org/10.1080/15248371003699969>
- de Valpine, P., Turek, D., Paciorek, C., Anderson-Bergman, C., Temple Lang, D., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics, 26*(2), 403–413. <https://doi.org/10.1080/10618600.2016.1172487>
- 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. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>
- Demszky, D., Yang, D., Yeager, D. S., Bryan, C. J., Clapper, M., Chandhok, S., Eichstaedt, J. C., Hecht, C., Jamieson, J., Johnson, M., Jones, M., Krettek-Cobb, D., Lai, L., JonesMitchell, N., Ong, D. C., Dweck, C. S., Gross, J. J., & Pennebaker, J. W. (2023). Using large language models in psychology. *Nature Reviews Psychology, 2*(11), 688–701. <https://doi.org/10.1038/s44159-023-00241-5>
- Denwood, M. J. (2016). *runjags*: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *Journal of Statistical Software, 71*(9), 1–25. <https://doi.org/10.18637/jss.v071.i09>
- Depaoli, S. (2013). Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus Bayesian estimation. *Psychological Methods, 18*(2), 186–219. <https://doi.org/10.1037/a0031609>
- Depaoli, S. (2014). The impact of inaccurate “informative” priors for growth parameters in Bayesian growth mixture modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(2), 239–252. <https://doi.org/10.1080/10705511.2014.882686>
- Depaoli, S. (2021). *Bayesian structural equation modeling*. Guilford Press.
- Depaoli, S., Heo, I., Jauregui, M., Liu, H., & Jia, F. (2026). A comprehensive evaluation of model selection indices for class enumeration in Bayesian latent growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal, 33*(2), 157–176. <https://doi.org/10.1080/10705511.2025.2566135>
- Depaoli, S., Jia, F., & Heo, I. (2023). Detecting model misspecifications in Bayesian piecewise growth models. *Structural Equation Modeling: A Multidisciplinary Journal, 30*(4), 574–591. <https://doi.org/10.1080/10705511.2022.2144865>
- Depaoli, S., Jia, F., & Visser, M. (2025). Addressing missing data in latent class analysis when using a three-step estimation approach. *Structural Equation*

- Modeling: A Multidisciplinary Journal*, 32(2), 287–303. <https://doi.org/10.1080/10705511.2024.2410240>
- Depaoli, S., Qiu, M., Liu, H., & Jauregui, M. (2026). Class selection in growth mixture models: Comparing information criteria to nonparametric and parametric Bayesian approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, 33(1), 64–80. <https://doi.org/10.1080/10705511.2025.2563187>
- Depaoli, S., Yang, Y., & Felt, J. (2017). Using Bayesian statistics to model uncertainty in mixture models: A sensitivity analysis of priors. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(2), 198–215. <https://doi.org/10.1080/10705511.2016.1250640>
- Derksen, S., & Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, 45(2), 265–282. <https://doi.org/10.1111/j.2044-8317.1992.tb00992.x>
- Diallo, T. M. O., Morin, A. J. S., & Lu, H. (2016). Impact of misspecifications of the latent variance–covariance and residual matrices on the class enumeration accuracy of growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(4), 507–531. <https://doi.org/10.1080/10705511.2016.1169188>
- Diallo, T. M. O., Morin, A. J. S., & Lu, H. (2017a). The impact of total and partial inclusion or exclusion of active and inactive time invariant covariates in growth mixture models. *Psychological Methods*, 22(1), 166–190. <https://doi.org/10.1037/met0000084>
- Diallo, T. M. O., Morin, A. J. S., & Lu, H. (2017b). Performance of growth mixture models in the presence of time-varying covariates. *Behavior Research Methods*, 49(5), 1951–1965. <https://doi.org/10.3758/s13428-016-0823-0>
- DiPerna, J. C., Lei, P.-W., & Reid, E. E. (2007). Kindergarten predictors of mathematical growth in the primary grades: An investigation using the Early Childhood Longitudinal Study–Kindergarten cohort. *Journal of Educational Psychology*, 99(2), 369–379. <https://doi.org/10.1037/0022-0663.99.2.369>
- Enders, C. K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods*, 16(1), 1–16. <https://doi.org/10.1037/a0022640>
- Enders, C. K. (2022). *Applied missing data analysis* (2nd). 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. <https://doi.org/10.1080/10705511.2011.532695>
- Enders, C. K., & Mansolf, M. (2018). Assessing the fit of structural equation models with multiply imputed data. *Psychological Methods*, 23(1), 76–93. <https://doi.org/10.1037/met0000102>
- 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. <https://doi.org/10.1080/10705510701758281>

- Ernst, A. F., Timmerman, M. E., Ji, F., Jeronimus, B. F., & Albers, C. J. (2024). Mixture multilevel vector-autoregressive modeling. *Psychological Methods*, 29(1), 137–154. <https://doi.org/10.1037/met0000551>
- Feng, X.-N., Wu, H.-T., & Song, X.-Y. (2017). Bayesian adaptive lasso for ordinal regression with latent variables. *Sociological Methods & Research*, 46(4), 926–953. <https://doi.org/10.1177/0049124115610349>
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2), 209–230. <https://doi.org/10.1214/aos/1176342360>
- Frankfurt, S., Frazier, P., Syed, M., & Jung, K. R. (2016). Using group-based trajectory and growth mixture modeling to identify classes of change trajectories. *The Counseling Psychologist*, 44(5), 622–660. <https://doi.org/10.1177/001100016658097>
- Frühwirth-Schnatter, S., & Malsiner-Walli, G. (2019). From here to infinity: Sparse finite versus Dirichlet process mixtures in model-based clustering. *Advances in Data Analysis and Classification*, 13(1), 33–64. <https://doi.org/10.1007/s11634-018-0329-y>
- Gamerman, D., & Lopes, H. F. (2006). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference* (2nd ed.). Chapman & Hall/CRC. <https://doi.org/10.1201/9781482296426>
- Garnier-Villareal, M., & Jorgensen, T. D. (2020). Adapting fit indices for Bayesian structural equation modeling: Comparison to maximum likelihood. *Psychological Methods*, 25(1), 46–70. <https://doi.org/10.1037/met0000224>
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3), 515–534. <https://doi.org/10.1214/06-BA117A>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd). Chapman & Hall/CRC.
- Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6), 997–1016. <https://doi.org/10.1007/s11222-013-9416-2>
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472. <https://doi.org/10.1214/ss/1177011136>
- George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423), 881–889. <https://doi.org/10.1080/01621459.1993.10476353>
- Gershman, S. J., & Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, 56(1), 1–12. <https://doi.org/10.1016/j.jmp.2011.08.004>
- Ghosh, J., Li, Y., & Mitra, R. (2018). On the use of Cauchy prior distributions for Bayesian logistic regression. *Bayesian Analysis*, 13(2), 359–383. <https://doi.org/10.1214/17-BA1051>

- Gomer, B., & Yuan, K.-H. (2021). Subtypes of the missing not at random missing data mechanism. *Psychological Methods*, 26(5), 559–598. <https://doi.org/10.1037/met0000377>
- Gomer, B., & Yuan, K.-H. (2023). A realistic evaluation of methods for handling missing data when there is a mixture of MCAR, MAR, and MNAR mechanisms in the same dataset. *Multivariate Behavioral Research*, 58(5), 988–1013. <https://doi.org/10.1080/00273171.2022.2158776>
- Gomez, R., Skilbeck, C., Thomas, M., & Slatyer, M. (2017). Growth mixture modeling of depression symptoms following traumatic brain injury. *Frontiers in Psychology*, 8, 1320. <https://doi.org/10.3389/fpsyg.2017.01320>
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549–576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>
- Green, P. J., & Richardson, S. (2001). Modelling heterogeneity with and without the Dirichlet process. *Scandinavian Journal of Statistics*, 28(2), 355–375. <https://doi.org/10.1111/1467-9469.00242>
- Griffin, J. E., & Brown, P. J. (2011). Bayesian hyper-lassos with non-convex penalization. *Australian & New Zealand Journal of Statistics*, 53(4), 423–442. <https://doi.org/10.1111/j.1467-842X.2011.00641.x>
- Grimm, K. J., Ram, N., & Estabrook, R. (2016). *Growth modeling: Structural equation and multilevel modeling approaches*. Guilford Press.
- Hamaker, E. L. (2025). Analysis of intensive longitudinal data: Putting psychological processes in perspective. *Annual Review of Clinical Psychology*, 21(1), 379–405. <https://doi.org/10.1146/annurev-clinpsy-081423-022947>
- Harel, O. (2007). Inferences on missing information under multiple imputation and two-stage multiple imputation. *Statistical Methodology*, 4(1), 75–89. <https://doi.org/10.1016/j.stamet.2006.03.002>
- Harra, K., & Kaplan, D. (2024). On the performance of horseshoe priors for inducing sparsity in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 31(4), 667–684. <https://doi.org/10.1080/10705511.2023.2280895>
- Harring, J. R., & Hodis, F. A. (2016). Mixture modeling: Applications in educational psychology. *Educational Psychologist*, 51(3–4), 354–367. <https://doi.org/10.1080/00461520.2016.1207176>
- Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: The lasso and generalizations*. CRC Press. <https://doi.org/10.1201/b18401>
- Henson, J. M., Pearson, M. R., & Carey, K. B. (2015). Defining and characterizing differences in college alcohol intervention efficacy: A growth mixture modeling application. *Journal of Consulting and Clinical Psychology*, 83(2), 370–381. <https://doi.org/10.1037/a0038897>
- 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. <https://doi.org/10.1080/10705510709336744>
- Heo, I., Depaoli, S., Jia, F., & Liu, H. (2024). Bayesian approach to piecewise growth mixture modeling: Issues and applications in school psychology. *Journal of School Psychology*, 107, 101366. <https://doi.org/10.1016/j.jsp.2024.101366>
- Heo, I., & Jeon, M. (2026). Rethinking the evaluation of psychological theories: A commentary on Calderon et al.'s registered replication report. *Advances in Methods and Practices in Psychological Science*, 9(2), 1–7. <https://doi.org/10.1177/25152459261440485>
- Heo, I., Jia, F., & Depaoli, S. (2024). Performance of model fit and selection indices for Bayesian piecewise growth modeling with missing data. *Structural Equation Modeling: A Multidisciplinary Journal*, 31(3), 455–476. <https://doi.org/10.1080/10705511.2023.2264514>
- Heo, I., Jia, F., & Depaoli, S. (2025). Recovering knot placements in Bayesian piecewise growth models with missing data. *Behavior Research Methods*, 57(7), 1–27. <https://doi.org/10.3758/s13428-025-02716-0>
- Heo, I., Simons, J.-W., & Liu, H. (2025). A tutorial on Bayesian model averaging for exponential random graph models. *British Journal of Mathematical and Statistical Psychology*. Advance online publication. <https://doi.org/10.1111/bmsp.70007>
- Hinne, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E.-J. (2020). A conceptual introduction to Bayesian model averaging. *Advances in Methods and Practices in Psychological Science*, 3(2), 200–215. <https://doi.org/10.1177/2515245919898657>
- Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods*, 11(1), 36–53. <https://doi.org/10.1037/1082-989X.11.1.36>
- Hjort, N. L., Holmes, C., Müller, P., & Walker, S. G. (2010). *Bayesian nonparametrics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511802478>
- Hodis, F. A., Meyer, L. H., McClure, J., Weir, K. F., & Walkey, F. H. (2011). A longitudinal investigation of motivation and secondary school achievement using growth mixture modeling. *Journal of Educational Psychology*, 103(2), 312–323. <https://doi.org/10.1037/a0022547>
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67. <https://doi.org/10.1080/00401706.1970.10488634>
- Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14(4), 382–417. <https://doi.org/10.1214/ss/1009212519>
- Hong, S., & You, S. (2012). Understanding Latino children's heterogeneous academic growth trajectories: Latent growth mixture modeling approach. *The Journal of Educational Research*, 105(4), 235–244. <https://doi.org/10.1080/00220671.2011.584921>

- Hsiang, T. C. (1975). A Bayesian view on ridge regression. *Journal of the Royal Statistical Society Series D: The Statistician*, 24(4), 267–268. <https://doi.org/10.2307/2987923>
- Hu, J., Leite, W. L., & Gao, M. (2017). An evaluation of the use of covariates to assist in class enumeration in linear growth mixture modeling. *Behavior Research Methods*, 49(3), 1179–1190. <https://doi.org/10.3758/s13428-016-0778-1>
- Huang, D., Brecht, M.-L., Hara, M., & Hser, Y.-I. (2010). Influences of covariates on growth mixture modeling. *Journal of Drug Issues*, 40(1), 173–194. <https://doi.org/10.1177/002204261004000110>
- Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1), 193–218. <https://doi.org/10.1007/BF01908075>
- Ibrahim, J. G., & Molenberghs, G. (2009). Missing data methods in longitudinal studies: A review. *Test*, 18(1), 1–43. <https://doi.org/10.1007/s11749-009-0138-x>
- Ishwaran, H. (2000). Inference for the random effects in Bayesian generalized linear mixed models. *ASA Proceedings of the Bayesian Statistical Science Section*, 1–10.
- Jacobucci, R. (2017). Regsem: Regularized structural equation modeling. *arXiv preprint arXiv:1703.08489*. <https://doi.org/10.48550/arXiv.1703.08489>
- Jacobucci, R., Brandmaier, A. M., & Kievit, R. A. (2019). A practical guide to variable selection in structural equation modeling by using regularized multiple-indicators, multiple-causes models. *Advances in Methods and Practices in Psychological Science*, 2(1), 55–76. <https://doi.org/10.1177/2515245919826527>
- Jacobucci, R., & Grimm, K. J. (2018). Comparison of frequentist and Bayesian regularization in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(4), 639–649. <https://doi.org/10.1080/10705511.2017.1410822>
- Jacobucci, R., Grimm, K. J., & Zhang, Z. (2023). *Machine learning for social and behavioral research*. Guilford Publications.
- Jia, F., & Wu, W. (2019). Evaluating methods for handling missing ordinal data in structural equation modeling. *Behavior Research Methods*, 51(5), 2337–2355. <https://doi.org/10.3758/s13428-018-1187-4>
- Jia, F., & Wu, W. (2023). A comparison of multiple imputation strategies to deal with missing nonnormal data in structural equation modeling. *Behavior Research Methods*, 55(6), 3100–3119. <https://doi.org/10.3758/s13428-022-01936-y>
- Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, 2(1), 302–317. <https://doi.org/10.1111/j.1751-9004.2007.00054.x>
- Kaplan, D. (2002). Methodological advances in the analysis of individual growth with relevance to education policy. *Peabody Journal of Education*, 77(4), 189–215. https://doi.org/10.1207/S15327930PJE7704_9
- Kaplan, D. (2023). *Bayesian statistics for the social sciences*. Guilford Publications.

- Khalili, A., & Chen, J. (2007). Variable selection in finite mixture of regression models. *Journal of the American Statistical Association*, 102(479), 1025–1038. <https://doi.org/10.1198/016214507000000590>
- Kim, E., Howard Kirby, C., Evudottir, E., Wang, Y., Tong, X., Goretzko, D., Ferron, J., & Jing, X. (2026). A small latent class in growth mixture modeling: A simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*. Advance online publication. <https://doi.org/10.1080/10705511.2026.2629849>
- Kim, S., Tong, X., & Ke, Z. (2021). Exploring class enumeration in Bayesian growth mixture modeling based on conditional medians. *Frontiers in Education*, 6, 624149. <https://doi.org/10.3389/feduc.2021.624149>
- Kim, S., Tong, X., Zhou, J., & Boichuk, J. P. (2022). Conditional median-based Bayesian growth mixture modeling for nonnormal data. *Behavior Research Methods*, 54(3), 1291–1305. <https://doi.org/10.3758/s13428-021-01655-w>
- Kohli, N., & Sullivan, A. L. (2019). Linear-linear piecewise growth mixture models with unknown random knots: A primer for school psychology. *Journal of School Psychology*, 73, 89–100. <https://doi.org/10.1016/j.jsp.2019.03.004>
- Kristman, V., Manno, M., & Côté, P. (2004). Loss to follow-up in cohort studies: How much is too much? *European Journal of Epidemiology*, 19(8), 751–760. <https://doi.org/10.1023/B:EJEP.0000036568.02655.f8>
- Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine*, 7(1–2), 305–315. <https://doi.org/10.1002/sim.4780070131>
- Lee, D. Y., & Harring, J. R. (2023). Handling missing data in growth mixture models. *Journal of Educational and Behavioral Statistics*, 48(3), 320–348. <https://doi.org/10.3102/10769986221149140>
- Lee, T., & Shi, D. (2021). A comparison of full information maximum likelihood and multiple imputation in structural equation modeling with missing data. *Psychological Methods*, 26(4), 466–485. <https://doi.org/10.1037/met0000381>
- Li, L., & Hser, Y.-I. (2011). On inclusion of covariates for class enumeration of growth mixture models. *Multivariate Behavioral Research*, 46(2), 266–302. <https://doi.org/10.1080/00273171.2011.556549>
- Li, M., & Harring, J. R. (2017). Investigating approaches to estimating covariate effects in growth mixture modeling: A simulation study. *Educational and Psychological Measurement*, 77(5), 766–791. <https://doi.org/10.1177/0013164416653789>
- Lindley, D. V. (1972). *Bayesian statistics, a review*. SIAM.
- Lindley, D. V. (1980). Jeffreys's contribution to modern statistical thought. In A. Zellner (Ed.), *Bayesian analysis in econometrics and statistics: Essays in honor of Harold Jeffreys* (pp. 35–39). North-Holland.
- Little, R. J. A. (1988). Missing-data adjustments in large surveys. *Journal of Business & Economic Statistics*, 6(3), 287–296. <https://doi.org/10.1080/07350015.1988.10509663>

- Little, R. J. A. (1992). Regression with missing X's: A review. *Journal of the American Statistical Association*, 87(420), 1227–1237. <https://doi.org/10.1080/01621459.1992.10476282>
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. John Wiley & Sons.
- Little, R. J. A., & Zhang, N. (2011). Subsample ignorable likelihood for regression analysis with missing data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60(4), 591–605. <https://doi.org/10.1111/j.1467-9876.2011.00763.x>
- Liu, H., Heo, I., Ivanov, A., & Depaoli, S. (2025). Model assumption violations in Bayesian latent mediation analysis: An exploration of Bayesian SEM fit indices and PPP. *Structural Equation Modeling: A Multidisciplinary Journal*, 32(5), 866–896. <https://doi.org/10.1080/10705511.2025.2503789>
- Liu, H., Jin, I. H., & Zhang, Z. (2018). Structural equation modeling of social networks: Specification, estimation, and application. *Multivariate Behavioral Research*, 53(5), 714–730. <https://doi.org/10.1080/00273171.2018.1479629>
- Liu, H., Tsang, S., Wood, A., & Tong, X. (2025). Longitudinal sentiment analysis with conversation textual data. *Fudan Journal of the Humanities and Social Sciences*, 18(1), 193–214. <https://doi.org/10.1007/s40647-024-00417-0>
- Liu, L. C., Hedeker, D., Segawa, E., & Flay, B. R. (2010). Evaluation of longitudinal intervention effects: An example of latent growth mixture models for ordinal drug-use outcomes. *Journal of Drug Issues*, 40(1), 27–43. <https://doi.org/10.1177/002204261004000103>
- Lu, Z. L., & Zhang, Z. (2014). Robust growth mixture models with non-ignorable missingness: Models, estimation, selection, and application. *Computational Statistics & Data Analysis*, 71, 220–240. <https://doi.org/10.1016/j.csda.2013.07.036>
- 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. <https://doi.org/10.1080/00273171.2011.589261>
- Lubke, G., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, 10(1), 21–39. <https://doi.org/10.1037/1082-989X.10.1.21>
- Magee, C. A., Caputi, P., & Iverson, D. C. (2013). The longitudinal relationship between sleep duration and body mass index in children: A growth mixture modeling approach. *Journal of Developmental and Behavioral Pediatrics*, 34(3), 165–173. <https://doi.org/10.1097/DBP.0b013e318289aa51>
- Marvin, L., Liu, H., & Depaoli, S. (2023). Using Bayesian piecewise growth curve models to handle complex nonlinear trajectories. *Journal of Behavioral Data Science*, 3(1), 1–33. <https://doi.org/10.35566/jbds/v3n1/marvin>
- Mattison, R. E., Woods, A. D., Morgan, P. L., Farkas, G., & Hillemeier, M. M. (2023). Longitudinal trajectories of reading and mathematics achievement for students with learning disabilities. *Journal of Learning Disabilities*, 56(2), 132–144. <https://doi.org/10.1177/00222194221085668>

- McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, 58(1), 110–133. <https://doi.org/10.2307/1130295>
- McCormick, E. M., Curran, P. J., & Hancock, G. R. (2024). Latent growth factors as predictors of distal outcomes. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000642>
- McDaid, A. F., Greene, D., & Hurley, N. (2011). Normalized mutual information to evaluate overlapping community finding algorithms [arXiv preprint arXiv:1110.2515]. <https://doi.org/10.48550/arXiv.1110.2515>
- McNeish, D., & Harring, J. R. (2021). Improving convergence in growth mixture models without covariance structure constraints. *Statistical Methods in Medical Research*, 30(4), 994–1012. <https://doi.org/10.1177/0962280220981747>
- McNeish, D., Harring, J. R., & Bauer, D. J. (2023). Nonconvergence, covariance constraints, and class enumeration in growth mixture models. *Psychological Methods*, 28(4), 962–992. <https://doi.org/10.1037/met0000456>
- McNeish, D., Peña, A., Vander Wyst, K. B., Ayers, S. L., Olson, M. L., & Shaibi, G. Q. (2023). Facilitating growth mixture model convergence in preventive interventions. *Prevention Science*, 24(3), 505–516. <https://doi.org/10.1007/s11121-021-01262-3>
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55(1), 107–122. <https://doi.org/10.1007/BF02294746>
- Milligan, G. W., & Cooper, M. C. (1986). A study of the comparability of external criteria for hierarchical cluster analysis. *Multivariate Behavioral Research*, 21(4), 441–458. https://doi.org/10.1207/s15327906mbr2104_5
- Mitchell, T. J., & Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404), 1023–1032. <https://doi.org/10.1080/01621459.1988.10478694>
- Muthén, B. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, 54(4), 557–585. <https://doi.org/10.1007/BF02296397>
- Muthén, B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, 29(1), 81–117. <https://doi.org/10.2333/bhmk.29.81>
- Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *The SAGE handbook of quantitative methodology for the social sciences* (pp. 345–368). SAGE Publications.
- Muthén, B., Asparouhov, T., Hunter, A. M., & Leuchter, A. F. (2011). Growth modeling with nonignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychological Methods*, 16(1), 17–33. <https://doi.org/10.1037/a0022634>
- Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S.-T., Yang, C.-C., Wang, C.-P., Kellam, S. G., Carlin, J. B., & Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics*, 3(4), 459–475. <https://doi.org/10.1093/biostatistics/3.4.459>

- Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55(2), 463–469. <https://doi.org/10.1111/j.0006-341X.1999.00463.x>
- Muthén, L. K., & Muthén, B. (2017). *Mplus user's guide* (8th). Los Angeles, CA, Muthén & Muthén.
- Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, 4(2), 139–157. <https://doi.org/10.1037/1082-989X.4.2.139>
- Nasserinejad, K., van Rosmalen, J., de Kort, W., & Lesaffre, E. (2017). Comparison of criteria for choosing the number of classes in Bayesian finite mixture models. *PloS One*, 12(1), e0168838. <https://doi.org/10.1371/journal.pone.0168838>
- Nicholson, J. S., Deboeck, P. R., & Howard, W. (2017). Attrition in developmental psychology: A review of modern missing data reporting and practices. *International Journal of Behavioral Development*, 41(1), 143–153. <https://doi.org/10.1177/0165025415618275>
- 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. <https://doi.org/10.1080/10705510701575396>
- Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, 4(4), 440–461. <https://doi.org/10.1037/tps0000176>
- O'Hara, R. B., & Sillanpää, M. J. (2009). A review of Bayesian variable selection methods: What, how and which. *Bayesian Analysis*, 4(1), 85–117. <https://doi.org/10.1214/09-BA403>
- Ortega-Azurduy, S. A., Tan, F. E. S., & Berger, M. P. F. (2008). The effect of dropout on the efficiency of D-optimal designs of linear mixed models. *Statistics in Medicine*, 27(14), 2601–2617. <https://doi.org/10.1002/sim.3108>
- Pargent, F., Schoedel, R., & Stachl, C. (2023). Best practices in supervised machine learning: A tutorial for psychologists. *Advances in Methods and Practices in Psychological Science*, 6(3), 25152459231162559. <https://doi.org/10.1177/25152459231162559>
- Park, T., & Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, 103(482), 681–686. <https://doi.org/10.1198/016214508000000337>
- Peeters, M., Zondervan-Zwijnenburg, M., Vink, G., & Van de Schoot, R. (2015). How to handle missing data: A comparison of different approaches. *European Journal of Developmental Psychology*, 12(4), 377–394. <https://doi.org/10.1080/17405629.2015.1049526>
- Petras, H., & Masyn, K. E. (2010). General growth mixture analysis with antecedents and consequences of change. In A. R. Piquero & D. Weisburd (Eds.), *Handbook of quantitative criminology* (1st ed., pp. 69–100). Springer. https://doi.org/10.1007/978-0-387-77650-7_5

- 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. <https://doi.org/10.1080/10705511.2012.659618>
- Pfadt, J. M., Bartoš, F., Godmann, H. R., Waaijers, M., Groot, L., Heo, I., Mensink, L., Nak, J., de Ruiter, J. P., Sarafoglou, A., Siepe, B., Arena, G., Akrong, E., Aust, F., van den Bergh, D., Brenner, W., Doekemeijer, R., Donzallaz, M. C., van Doorn, J., . . . Wagenmakers, E.-J. (2025). A methodological metamorphosis: The rapid rise of Bayesian inference and open science practices in psychology. *PsyArXiv*, 1–47. https://doi.org/10.31234/osf.io/ck3js_v1
- Piironen, J., & Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, 11(2), 5018–5051. <https://doi.org/10.1214/17-EJS1337SI>
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*. <https://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf>
- Polson, N. G., & Scott, J. G. (2011). Shrink globally, act locally: Sparse Bayesian regularization and prediction. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, & M. West (Eds.), *Bayesian statistics 9* (pp. 501–538). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199694587.003.0017>
- Polson, N. G., & Scott, J. G. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis*, 7(4), 887–902. <https://doi.org/10.1214/12-BA730>
- Polson, N. G., Scott, J. G., & Windle, J. (2014). The Bayesian bridge. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(4), 713–733. <https://doi.org/10.1111/rssb.12042>
- Qiu, M., Paganin, S., Ohn, I., & Lin, L. (2025). Bayesian nonparametric latent class analysis with different item types. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000728>
- R Core Team. (2025). *R: A language and environment for statistical computing* [R Foundation for Statistical Computing]. Vienna, Austria. <https://www.R-project.org/>
- Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437), 179–191. <https://doi.org/10.1080/01621459.1997.10473615>
- Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, 27(1), 85–95.
- Ram, N., & Grimm, K. J. (2009). Methods and measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved

- groups. *International Journal of Behavioral Development*, 33(6), 565–576. <https://doi.org/10.1177/0165025409343765>
- Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336), 846–850. <https://doi.org/10.1080/01621459.1971.10482356>
- Robert, C. P., & Casella, G. (2004). *Monte Carlo statistical methods* (2nd ed.). Springer. <https://doi.org/10.1007/978-1-4757-4145-2>
- Rodriguez, D., Moss, H. B., & Audrain-McGovern, J. (2005). Developmental heterogeneity in adolescent depressive symptoms: Associations with smoking behavior. *Psychosomatic Medicine*, 67(2), 200–210. <https://doi.org/10.1097/01.psy.0000156929.83810.01>
- Rosenthal, J. A. (1996). Qualitative descriptors of strength of association and effect size. *Journal of Social Service Research*, 21(4), 37–59. https://doi.org/10.1300/J079v21n04_02
- Rousseau, J., & Mengersen, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(5), 689–710. <https://doi.org/10.1111/j.1467-9868.2011.00781.x>
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592. <https://doi.org/10.1093/biomet/63.3.581>
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. CRC Press.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147–177. <https://doi.org/10.1037/1082-989X.7.2.147>
- 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. <https://doi.org/10.1198/106186002760180608>
- Scrucca, L., Fraley, C., Murphy, T. B., & Raftery, A. E. (2023). *Model-based clustering, classification, and density estimation using mclust in R*. Chapman; Hall/CRC. <https://doi.org/10.1201/9781003277965>
- Sekulovski, N., Bartoš, F., van den Bergh, D., Arena, G., Godmann, H. R., Goyal, V., Pfadt, J. M., Marsman, M., & Raftery, A. E. (2025). Comparing variable selection and model averaging methods for logistic regression. *arXiv preprint arXiv:2511.23216*. <https://doi.org/10.48550/arXiv.2511.23216>
- Sekulovski, N., Waaijers, M., & Arena, G. (2026). LLM-based prior elicitation for Bayesian graphical modeling. *British Journal of Mathematical and Statistical Psychology*. Advance online publication. <https://doi.org/10.1111/bmsp.70045>
- Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4(2), 639–650.
- Shen, Z. (2000). *Nested multiple imputation* [Doctoral dissertation, Harvard University, Department of Statistics].

- Shi, D., DiStefano, C., Zheng, X., Liu, R., & Jiang, Z. (2021). Fitting latent growth models with small sample sizes and non-normal missing data. *International Journal of Behavioral Development*, 45(2), 179–192. <https://doi.org/10.1177/0165025420979365>
- Shi, D., Shi, D., & Fairchild, A. J. (2023). Variable selection for mediators under a Bayesian mediation model. *Structural Equation Modeling: A Multidisciplinary Journal*, 30(6), 887–900. <https://doi.org/10.1080/10705511.2022.2164285>
- Si, Y., & Reiter, J. P. (2013). Nonparametric Bayesian multiple imputation for incomplete categorical variables in large-scale assessment surveys. *Journal of Educational and Behavioral Statistics*, 38(5), 499–521. <https://doi.org/10.3102/1076998613480394>
- Steinley, D. (2004). Properties of the Hubert-Arabie adjusted Rand index. *Psychological Methods*, 9(3), 386–396. <https://doi.org/10.1037/1082-989X.9.3.386>
- Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4), 795–809. <https://doi.org/10.1111/1467-9868.00265>
- Sterba, S. K. (2014). Handling missing covariates in conditional mixture models under missing at random assumptions. *Multivariate Behavioral Research*, 49(6), 614–632. <https://doi.org/10.1080/00273171.2014.950719>
- 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 & Adolescent Psychology*, 45(2), 167–175. <https://doi.org/10.1080/15374416.2014.958839>
- Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398), 528–540. <https://doi.org/10.1080/01621459.1987.10478458>
- Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement*, 55(4), 525–534. <https://doi.org/10.1177/0013164495055004001>
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Tofighi, D., & Enders, C. K. (2007). Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models*. Information Age Publishing. <https://doi.org/10.1108/978-1-60752-634-620251017>
- Tong, L., & Zhang, Z. (2026). Neural network analysis of psychological data: A step-by-step guide. *Multivariate Behavioral Research*, 61(2), 335–355. <https://doi.org/10.1080/00273171.2025.2587379>
- Tueller, S., Drotar, S., & Lubke, G. (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. <https://doi.org/10.1080/10705511.2011.534695>

- Tueller, S., & Lubke, G. (2010). Evaluation of structural equation mixture models: Parameter estimates and correct class assignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 17(2), 165–192. <https://doi.org/10.1080/10705511003659318>
- Twisk, J., & de Vente, W. (2002). Attrition in longitudinal studies: How to deal with missing data. *Journal of Clinical Epidemiology*, 55(4), 329–337. [https://doi.org/10.1016/S0895-4356\(01\)00476-0](https://doi.org/10.1016/S0895-4356(01)00476-0)
- Urban, C. J., & Gates, K. M. (2021). Deep learning: A primer for psychologists. *Psychological Methods*, 26(6), 743–773. <https://doi.org/10.1037/met0000374>
- Vallejo, G., Ato, M., Fernández, M. P., & Livacic-Rojas, P. E. (2019). Sample size estimation for heterogeneous growth curve models with attrition. *Behavior Research Methods*, 51(3), 1216–1243. <https://doi.org/10.3758/s13428-018-1059-y>
- van Erp, S., Oberski, D. L., & Mulder, J. (2019). Shrinkage priors for Bayesian penalized regression. *Journal of Mathematical Psychology*, 89, 31–50. <https://doi.org/10.1016/j.jmp.2018.12.004>
- van Buuren, S. (2018). *Flexible imputation of missing data*. CRC Press.
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67. <https://doi.org/10.18637/jss.v045.i03>
- van de Schoot, R., Sijbrandij, M., Depaoli, S., Winter, S. D., Olff, M., & Van Loey, N. E. (2018). Bayesian PTSD-trajectory analysis with informed priors based on a systematic literature search and expert elicitation. *Multivariate Behavioral Research*, 53(2), 267–291. <https://doi.org/10.1080/00273171.2017.1412293>
- van de Schoot, R., Sijbrandij, M., Winter, S. D., Depaoli, S., & Vermunt, J. K. (2017). The GRoLTS-checklist: Guidelines for reporting on latent trajectory studies. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(3), 451–467. <https://doi.org/10.1080/10705511.2016.1247646>
- Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5), 1413–1432. <https://doi.org/10.1007/s11222-016-9696-4>
- Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18(4), 450–469. <https://doi.org/10.1093/pan/mpq025>
- Vinh, N. X., Epps, J., & Bailey, J. (2009). Information theoretic measures for clusterings comparison: Is a correction for chance necessary? *Proceedings of the 26th Annual International Conference on Machine Learning*, 1073–1080. <https://doi.org/10.1145/1553374.1553511>
- Wake, S. K., Zewotir, T., & Muluneh, E. K. (2023). Analysis of heterogeneous growth changes in longitudinal height of children. *Journal of Health, Population and Nutrition*, 42(1), 78. <https://doi.org/10.1186/s41043-023-00425-y>

- Wang, M., & Bodner, T. E. (2007). Growth mixture modeling: Identifying and predicting unobserved subpopulations with longitudinal data. *Organizational Research Methods*, 10(4), 635–656. <https://doi.org/10.1177/1094428106289397>
- Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571–3594. <https://www.jmlr.org/papers/volume11/watanabe10a/watanabe10a.pdf>
- Wei, T., Liu, X., & Barnard-Brak, L. (2015). Gender differences in mathematics and reading trajectories among children from kindergarten to eighth grade. *Research in Education*, 93(1), 77–89. <https://doi.org/10.7227/RIE.0015>
- West, M. (1987). On scale mixtures of normal distributions. *Biometrika*, 74(3), 646–648. <https://doi.org/10.1093/biomet/74.3.646>
- Whittingham, M. J., Stephens, P. A., Bradbury, R. B., & Freckleton, R. P. (2006). Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology*, 75(5), 1182–1189. <https://doi.org/10.1111/j.1365-2656.2006.01141.x>
- Wilcox, K. T., Jacobucci, R., Zhang, Z., & Ammerman, B. A. (2023). Supervised latent Dirichlet allocation with covariates: A Bayesian structural and measurement model of text and covariates. *Psychological Methods*, 28(5), 1178–1206. <https://doi.org/10.1037/met0000541>
- Winter, S. D., & Depaoli, S. (2022). Sensitivity of Bayesian model fit indices to the prior specification of latent growth models. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(5), 667–686. <https://doi.org/10.1080/10705511.2022.2032078>
- Wolpert, D. H., & Strauss, C. E. M. (1996). What Bayes has to say about the evidence procedure. In *Maximum entropy and bayesian methods: Santa barbara, california, usa, 1993* (pp. 61–78). Springer.
- Wyman, A., & Zhang, Z. (2025). A tutorial on the use of artificial intelligence tools for facial emotion recognition in R. *Multivariate Behavioral Research*, 60(3), 641–655. <https://doi.org/10.1080/00273171.2025.2455497>
- Xiao, X., Rabe-Hesketh, S., & Skrondal, A. (2025). Bayesian identification and estimation of growth mixture models. *Psychometrika*, 90(2), 442–475. <https://doi.org/10.1017/psy.2025.11>
- Xu, Z., & Zhang, Z. (2025). Structural equation models with social networks. *Structural Equation Modeling: A Multidisciplinary Journal*, 33(1), 124–133. <https://doi.org/10.1080/10705511.2025.2488030>
- Zhang, L., & Liang, X. (2024). Bayesian regularization in multiple-indicators multiple-causes models. *Psychological Methods*, 29(4), 679–703. <https://doi.org/10.1037/met0000594>
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>