THE UNIVERSITY OF RHODE ISLAND

Evaluating the Effect of Change on Change in Cross-Domain Latent Growth Curve Analysis with Missing Data

Parisa Rafiee, M.S., Manshu Yang, Ph.D.

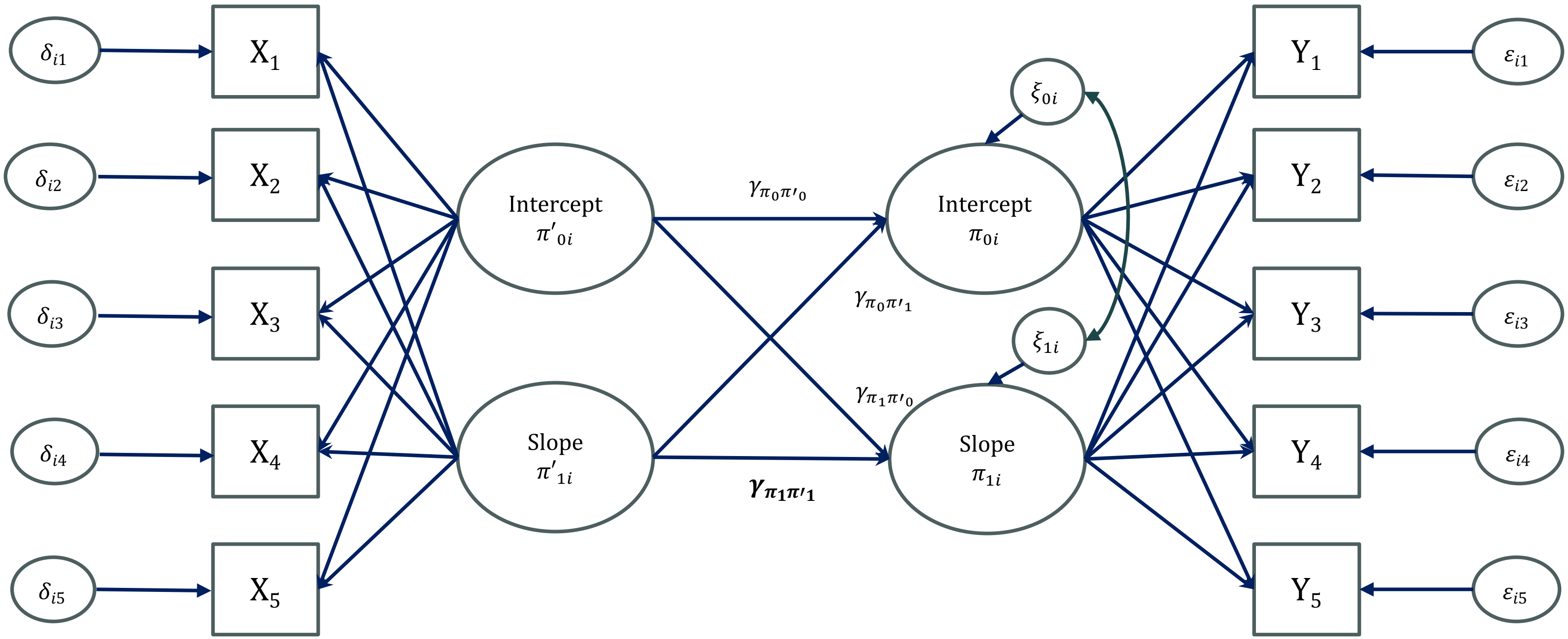
Modern Modeling Methods Conference
June 2024

THINK BIG WE DO™

Introduction

- Cross-domain latent growth curve (CD-LGC):
  - Examines the dynamic interplay between changes over time in outcomes and time-varying predictors.
  - Concurrently model changes in outcome and predictor variables.
  - Explore the relationship between their growth parameters.
  - Usually conducted under an SEM framework.

Model Specification



The Y-measurement model:

$$Y = \tau_y + \Lambda_y \eta + \varepsilon$$

$$Y = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4} \\ Y_{i5} \end{bmatrix}, \quad \eta = \begin{bmatrix} \pi_{0i} \\ \pi_{1i} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \\ \varepsilon_{i5} \end{bmatrix}$$

$$\tau_y = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \Lambda_y = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \\ 1 & t_4 \\ 1 & t_5 \end{bmatrix}$$

The X-measurement model:

$$X = \tau_x + \Lambda_x \xi + \delta$$

$$X = \begin{bmatrix} X_{i1} \\ X_{i2} \\ X_{i3} \\ X_{i4} \\ X_{i5} \end{bmatrix}, \quad \xi = \begin{bmatrix} \pi'_{0i} \\ \pi'_{1i} \end{bmatrix}, \quad \delta = \begin{bmatrix} \delta_{i1} \\ \delta_{i2} \\ \delta_{i3} \\ \delta_{i4} \\ \delta_{i5} \end{bmatrix}$$

$$\tau_x = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \Lambda_x = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \\ 1 & t_4 \\ 1 & t_5 \end{bmatrix}$$

Cross-domain Analysis of Change

$$\eta = \alpha + \Gamma\xi + \mathrm{B}\eta + \zeta$$

where the score vectors are:

$$\eta = \begin{bmatrix} \pi_{0i} \\ \pi_{1i} \end{bmatrix}, \quad \xi = \begin{bmatrix} \pi'_{0i} \\ \pi'_{1i} \end{bmatrix}, \quad \zeta = \begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix}$$

and parameter matrices are:

$$\alpha = \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix}, \qquad \Gamma = \begin{bmatrix} \gamma_{\pi_0\pi'_0} & \gamma_{\pi_0\pi'_1} \\ \gamma_{\pi_1\pi'_0} & \boldsymbol{\gamma_{\pi_1\pi'_1}} \end{bmatrix}, \qquad \mathrm{B} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

The Challenge of Missing Data

- Missing Data is common in longitudinal studies (CD-LGC).
- Can occur in the time-varying outcome, the time-varying predictor, or both.
- If inappropriately handled, missing data can result in:
  - → Biased and less precise estimates
  - → Reduced statistical power
  - → Diminished generalizability of results

Types of Missing Data

- Missing Completely at Random (MCAR)
- Missing at Random (MAR)
- Missing Not at Random (MNAR)

Estimation Methods

- Listwise Deletion (LD) using Maximum Likelihood
- Full Information Maximum Likelihood (FIML)
- Sequential Fully Bayesian (SFB)

- Complete data (pre-deletion) analysis acted as a benchmark—representing the best-case scenario without missing data.

Method 1: LD

- Implemented in R lavaan package as default.
- Only uses complete cases with observed values for all variables.
- Creates a likelihood function for complete cases.
- Maximizes the likelihood across complete cases to estimate the model parameters.

Method 2: FIML

- Implemented in many SEM software and packages.
- Widely used methods in handling MAR data.
- Uses all available data points to estimate the parameters of interest, without needing to impute missing values.
- Creates a likelihood function for each observation based on the observed data.
- Maximizes the overall likelihood across all observations to estimate the model parameters.

Method 3: SFB

- Utilizes all available data (observed and partially observed) to update estimates and uncertainties in a sequential manner.
- Missing data are considered parameters.
- Incorporates new data as it becomes available, to refine estimates over iterations.

Simulation Study

- Four factors were manipulated across 1000 replications:
  1. *Missing data scenario:* missingness in the time-varying predictor vs. missingness in the time-varying outcome
  2. *Number of participants:* n = 100, 200, or 400
  3. *Number of measurement occasions*: t = 5 or 9
  4. *Reliability of slope:* reliability = 0.5 or 0.8

Complete Data Generation

- Individual growth parameters (π_{0i}, π_{1i}, π'_{0i}, and π'_{1i}) generated using a multivariate normal distribution, and ε and δ using a normal distribution.
- X and Y were generated based on:

$$X = \Lambda_x \xi + \delta$$

$$Y = \Lambda_y \eta + \varepsilon$$

Missing Data Generation

- The probability of MAR data expressed as:

$$Pr(R_{Yi} = 1) = \Phi\,(\eta_{\,0} + \eta_1 \mathrm{X}_i + \eta_2 time_i) \qquad \text{if outcome Y is missing}$$

$$Pr(R_{\mathrm{X}i} = 1) = \Phi\,(\eta_{\,0} + \eta_1 Y_i + \eta_2 time_i) \qquad \text{if predictor X is missing}$$

  - $\eta_1 \rightarrow$ 1.815
  - $\eta_0, \eta_2 \rightarrow$ -1.31 and 0.085 (30% at mid timepoint, 40% at last timepoint)

- Indicator R is generated from a binomial distribution.
- R = 1 indicates missing data.

CD-LGC Model

- After data generation, a CD-LGC model was fitted to each dataset.

$$\eta = \alpha + \Gamma\xi + \mathrm{B}\eta + \zeta \quad , \qquad \Gamma = \begin{bmatrix} \gamma_{\pi_0\pi'_0} & \gamma_{\pi_0\pi'_1} \\ \gamma_{\pi_1\pi'_0} & \gamma_{\pi_1\pi'_1} \end{bmatrix}$$

- Primary parameter of interest was $\gamma_{\pi_1\pi'_1}$.
- $\gamma_{\pi_1\pi'_1}$ was estimated via LD, FIML and SFB.

→ **LD** and **FIML**: R lavaan package (Rosseel, 2012)

→ **SFB**: R rjagss package (Plummer, 2023)

Evaluation Statistics

- To evaluate the performance of the estimation methods, three key statistics were used:
  - Relative Bias
  - Mean Squared Error (MSE)
  - Type I Error Rate

Results: missing data in predictor

Table 1

Relative bias and MSE for $\gamma_{\pi_1 \pi'_1}$, when X is missing (reliability = 0.5)

Sample Size	Method	t = 5				t = 9			
		Complete Data		pmiss = 30%		Complete Data		pmiss = 30%	
		% Bias	MSE	% Bias	MSE	% Bias	MSE	% Bias	MSE
100	LD	**11.20**	0.12	**-38.38**	0.23	2.33	0.11	**-36.72**	0.08
	FIML	**11.20**	0.12	-4.21	0.15	2.33	0.11	1.60	0.24
	SFB	**63.88**	0.28	**55.61**	0.32	**50.12**	0.25	**51.31**	0.31
200	LD	8.09	0.06	**-31.33**	0.18	4.99	0.05	**-65.35**	0.31
	FIML	8.09	0.06	8.32	0.09	4.99	0.05	6.52	0.07
	SFB	**56.71**	0.17	**69.69**	0.24	**47.47**	0.13	**69.99**	0.23
400	LD	3.32	0.02	**-17.53**	0.12	2.58	0.02	**-34.62**	0.22
	FIML	3.32	0.02	7.84	0.05	2.58	0.02	8.52	0.04
	SFB	**28.26**	0.06	**60.61**	0.15	**21.90**	0.05	**54.82**	0.13

Results: missing data in predictor

Table 2

Relative bias and MSE for $\gamma_{\pi_1 \pi'_1}$, when X is missing (reliability = 0.8)

Sample Size	Method	t = 5				t = 9			
		Complete Data		pmiss = 30%		Complete Data		pmiss = 30%	
		% Bias	MSE	% Bias	MSE	% Bias	MSE	% Bias	MSE
100	LD	1.78	0.02	**-13.06**	0.11	3.21	0.02	**-20.69**	0.15
	FIML	1.78	0.02	3.76	0.03	3.21	0.02	3.82	0.03
	SFB	7.65	0.03	**19.89**	0.06	8.08	0.02	**16.57**	0.05
200	LD	0.49	0.01	**-15.42**	0.04	1.33	0.01	**-16.50**	0.08
	FIML	0.49	0.01	2.01	0.01	1.33	0.01	-0.72	0.01
	SFB	2.68	0.01	7.95	0.02	3.37	0.01	3.70	0.02
400	LD	0.52	0.00	**-15.90**	0.02	0.51	0.00	**-18.55**	0.04
	FIML	0.52	0.00	1.42	0.01	0.51	0.00	0.87	0.01
	SFB	1.55	0.01	3.77	0.01	1.44	0.00	2.88	0.01

Results: missing data in outcome

Table 3

Relative bias and MSE for $\gamma_{\pi_1\pi'_1}$, when Y is missing (reliability = 0.5)

Sample Size	Method	t = 5				t = 9			
		Complete Data		pmiss = 30%		Complete Data		pmiss = 30%	
		% Bias	MSE	% Bias	MSE	% Bias	MSE	% Bias	MSE
100	LD	**11.20**	0.12	**-23.49**	0.51	2.33	0.11	**15.93**	0.00
	FIML	**11.20**	0.12	0.42	0.14	2.33	0.11	4.57	0.13
	SFB	**63.88**	0.28	**46.99**	0.29	**50.12**	0.25	**48.12**	0.28
200	LD	8.09	0.06	4.90	0.30	4.99	0.05	-2.65	0.51
	FIML	8.09	0.06	1.48	0.08	4.99	0.05	7.18	0.06
	SFB	**56.71**	0.17	**38.25**	0.18	**47.47**	0.13	**43.77**	0.15
400	LD	3.32	0.02	**12.85**	0.19	2.58	0.02	**11.92**	0.32
	FIML	3.32	0.02	1.43	0.03	2.58	0.02	1.65	0.03
	SFB	**28.26**	0.06	**24.72**	0.07	**21.90**	0.05	**20.97**	0.06

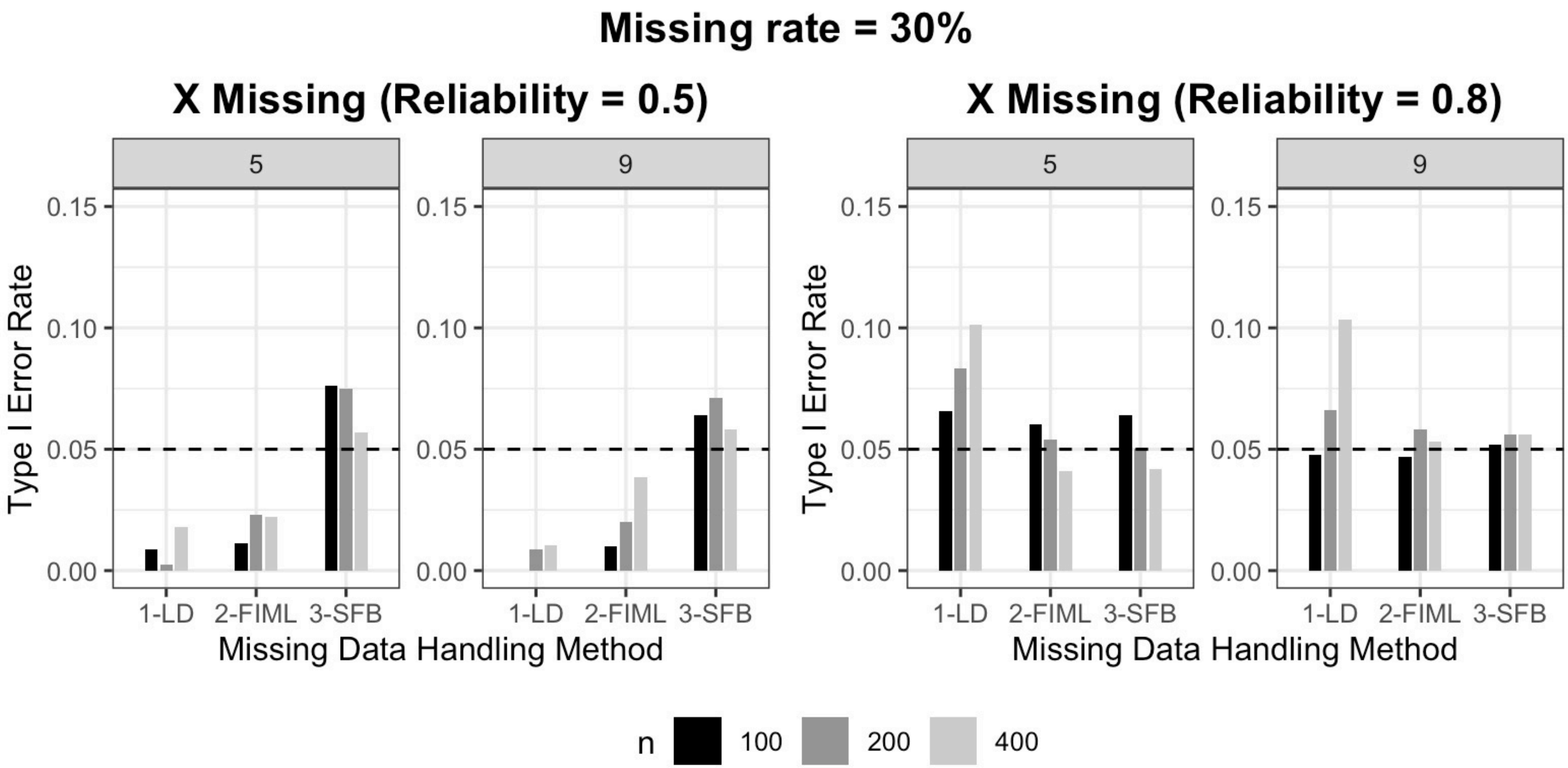
Results: missing data in outcome

Table 4

Relative bias and MSE for $\gamma_{\pi_1 \pi'_1}$, when Y is missing (reliability = 0.8)

Sample Size	Method	t = 5				t = 9			
		Complete Data		pmiss = 30%		Complete Data		pmiss = 30%	
		% Bias	MSE	% Bias	MSE	% Bias	MSE	% Bias	MSE
100	LD	1.78	0.02	**12.44**	0.17	3.21	0.02	-1.61	0.28
	FIML	1.78	0.02	4.40	0.03	3.21	0.02	1.68	0.03
	SFB	7.65	0.03	**11.17**	0.04	8.08	0.02	7.61	0.04
200	LD	0.49	0.01	6.66	0.07	1.33	0.01	**16.91**	0.15
	FIML	0.49	0.01	-1.05	0.01	1.33	0.01	0.87	0.01
	SFB	2.68	0.01	1.60	0.02	3.37	0.01	3.23	0.01
400	LD	0.52	0.00	6.12	0.03	0.51	0.00	**14.56**	0.05
	FIML	0.52	0.00	0.34	0.01	0.51	0.00	0.13	0.01
	SFB	1.55	0.01	1.60	0.01	1.44	0.00	1.26	0.01

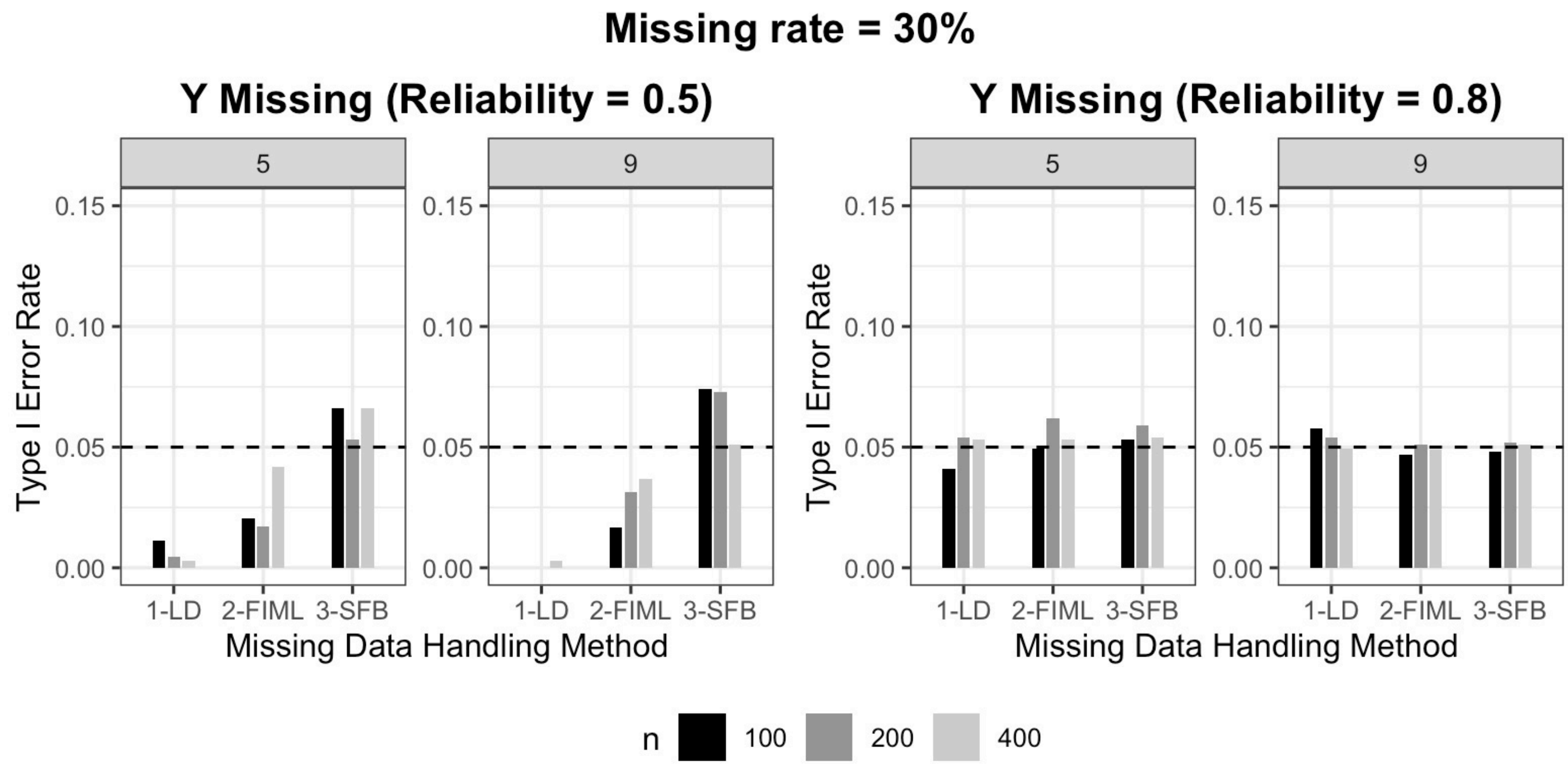
Results: Type I error rate



THINK BIG WE DO℠

THE UNIVERSITY OF RHODE ISLAND

Results: Type I error rate



Discussion

- Boosting reliability from 0.5 to 0.8 enhances the performance of all methods.
- **LD:**
  - Severe biases in the presence of missing data.
- **SFB:**
  - Performs well (in terms of bias and Type I error) with
    - high reliability (0.8)
    - larger sample size (N = 200 +)
- **FIML:**
  - Consistently provides the lowest bias and MSE (even for N = 100)
  - Deflated Type I error rate given low reliability (0.5)
  - Most effective method for handling relatively reliable data with intermittent missingness.

Future Directions

- Assess how other patterns of missingness, such as monotone dropout would affect the performance of estimation methods.
- Assess how MNAR missingness would influence the estimation of key parameters.
- Apply these methods to real-world datasets to validate the effectiveness of these techniques in practical scenarios.

References

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, *112*(4), 545.

Enders, C. K. (2022). Applied missing data analysis. Guilford Publications.

Enders, C. K., Du, H., & Keller, B. T. (2020). A model-based imputation procedure for multilevel regression models with random coefficients, interaction effects, and nonlinear terms. Psychological Methods, 25(1), 88.

Erler, N. S., Rizopoulos, D., Jaddoe, V. W. V, Franco, O. H., & Lesaffre, E. M. E. H. (2017). Bayesian imputation of time-varying covariates in linear mixed models. Statistical Methods in Medical Research, 28(2), 555–568. https://doi.org/10.1177/0962280217730851

Erler, N. S., Rizopoulos, D., Rosmalen, J. van, Jaddoe, V. W. V, Franco, O. H., & Lesaffre, E. M. E. H. (2016). Dealing with missing covariates in epidemiologic studies: a comparison between multiple imputation and a full Bayesian approach. Statistics in Medicine, 35(17), 2955–2974.

Ibrahim, J. G., & Molenberghs, G. (2009). Missing data methods in longitudinal studies: a review. Test, 18(1), 1–43.

Plummer M (2023). rjags: Bayesian Graphical Models using MCMC. R package version 4-15,

References

Kaplan, D. (1988). The impact of specification error on the estimation, testing, and improvement of structural equation models. Multivariate Behavioral Research, 23(1), 69–86.

Little, R. J. A., & Rubin, D. B. (2019). Statistical analysis with missing data (Vol. 793). John Wiley & Sons.

Lüdtke, O., Robitzsch, A., & West, S. G. (2020). Regression models involving nonlinear effects with missing data: A sequential modeling approach using Bayesian estimation. Psychological Methods, 25(2), 157.

Schafer, J. L., & Graham, J. W. (2002b). Missing data: our view of the state of the art. Psychol Methods, 7, 147–177.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. In Journal of Statistical Software (Vol. 48, Issue 2, pp. 1–36).

Singer, J. D., & Willett, J. B. (2003). Applied longitudinal data analysis: Modeling change and event occurrence. Oxford university press.

Zhou, L., Wang, M., & Zhang, Z. (2021). Intensive longitudinal data analyses with dynamic structural equation modeling. Organizational Research Methods, 24(2), 219–250.

Zhou, X.-H., Zhou, C., Lui, D., & Ding, X. (2014). Applied missing data analysis in the health sciences. John Wiley & Sons.

Thank you!

For further information, please contact:

Parisa Rafiee
Email: prsrafiee@uri.edu

THINK BIG WE DO℠

THE UNIVERSITY OF RHODE ISLAND