Using Two-Stage Path Analysis to Account for Measurement Error and Noninvariance

Hok Chio (Mark) Lai
University of Southern California

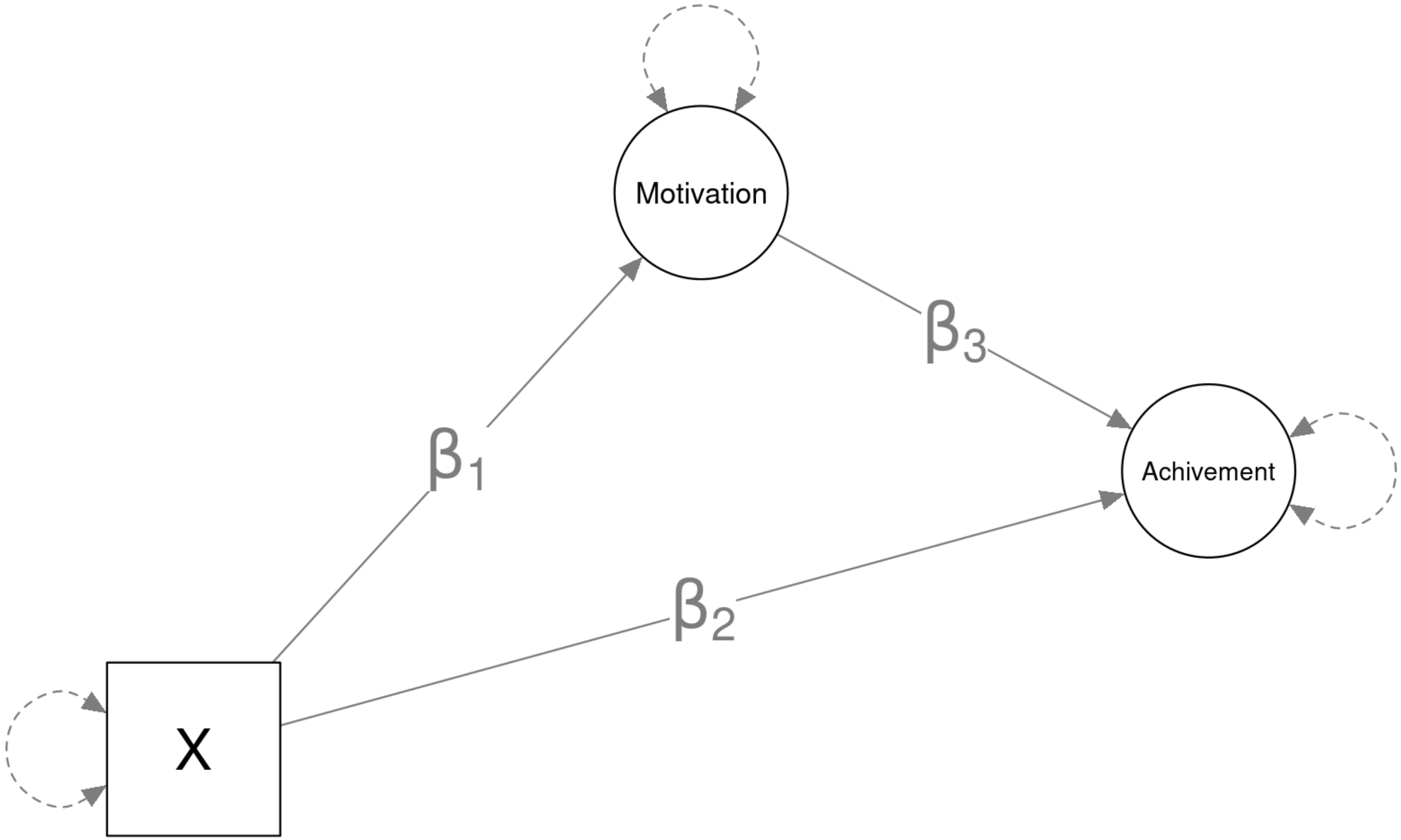
March 27, 2024

Roadmap

- 2S-PA as an alternative to joint SEM modeling
- Example 1: Categorical indicators violating measurement invariance
- Example 2: Growth modeling of latent constructs
- Extensions & Limitations

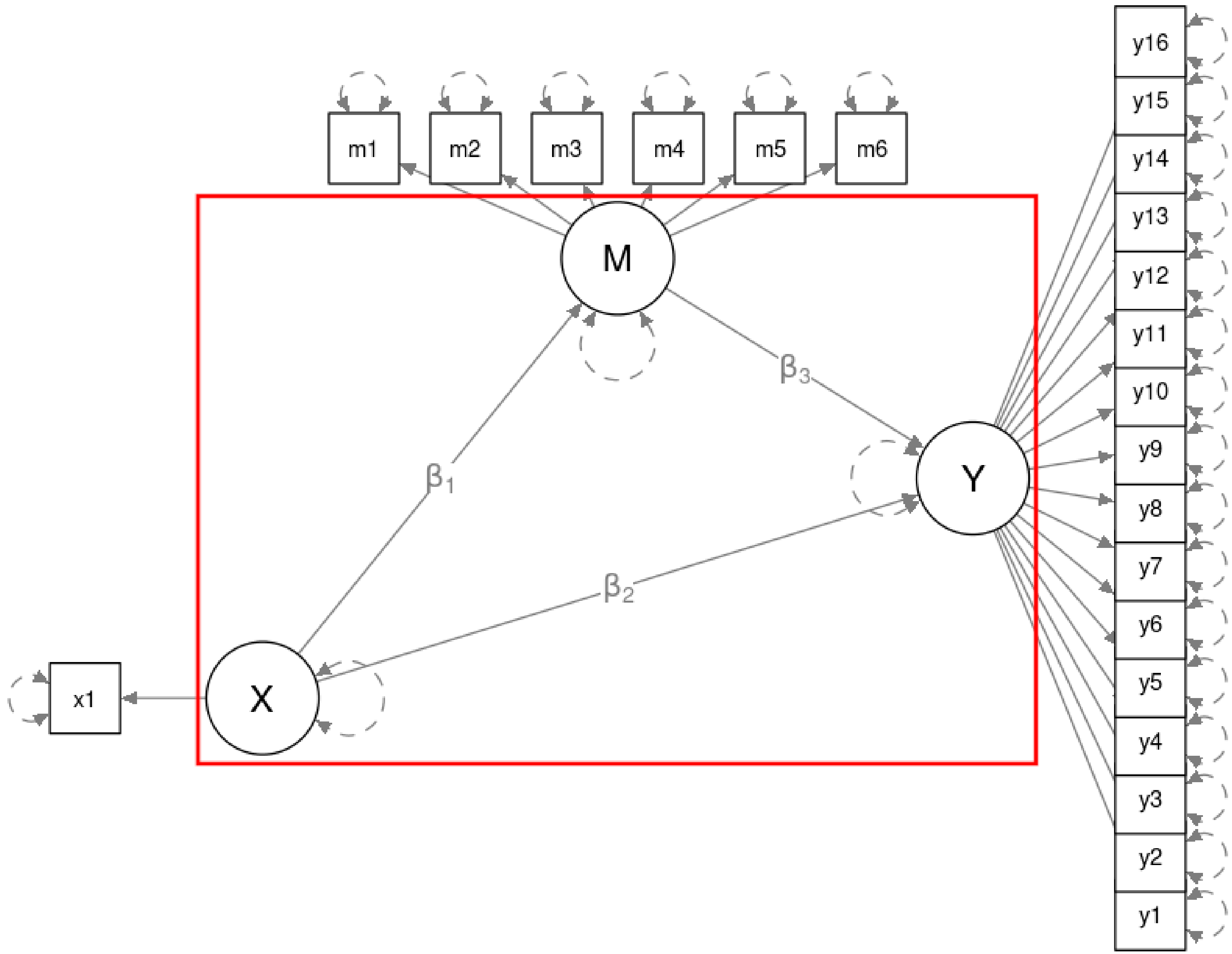
Source code of this presentation and the two examples can be found on GitHub:

Path Analysis



But constructs are typically not directly observed

Joint Measurement and Structural Modeling



Joint Modeling (JM) Not Always Practical

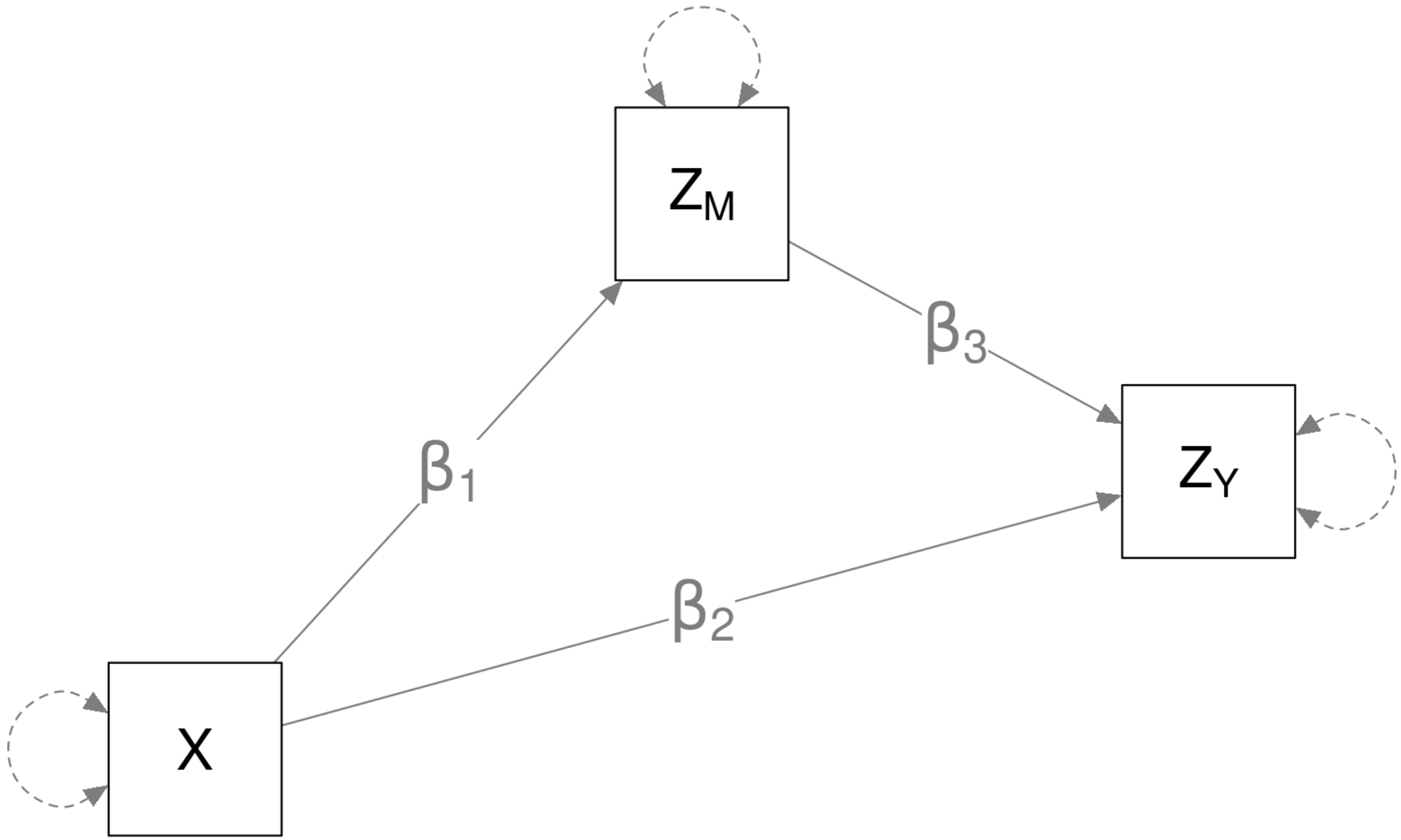
- Need a large model
- Need a large sample
  - Especially with binary/ordinal indicators

Convergence Issues

```
Warning messages:
1: In lav_samplestats_step2(UNI = FIT, wt = wt, ov.names = ov.names,  :
  lavaan WARNING: correlation between variables y1 and m3 is (nearly) 1.0
2: In lav_samplestats_step2(UNI = FIT, wt = wt, ov.names = ov.names,  :
  lavaan WARNING: correlation between variables y1 and m4 is (nearly) 1.0
3: In lav_samplestats_step2(UNI = FIT, wt = wt, ov.names = ov.names,  :
  lavaan WARNING: correlation between variables y1 and m6 is (nearly) 1.0
4: In muthen1984(Data = X[[g]], wt = WT[[g]], ov.names = ov.names[[g]],  :
  lavaan WARNING: trouble constructing W matrix; used generalized inverse for A11 submatrix
5: In lavaan::lavaan(model = sem_mod, data = test_dat, ordered = paste0("y",  :
  lavaan WARNING:
    the optimizer warns that a solution has NOT been found!
```

```
lavaan 0.6-9 did NOT end normally after 9183 iterations
** WARNING ** Estimates below are most likely unreliable
```

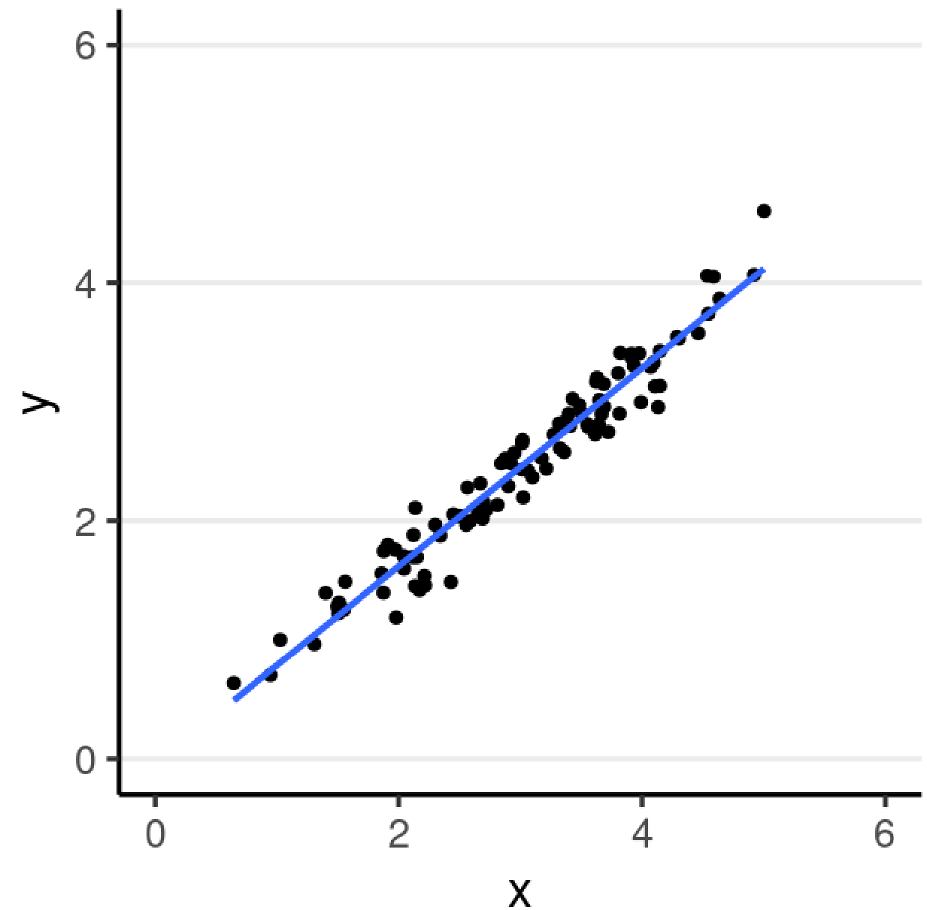
Alternative 1: Using Composite Scores



But, imperfect measurement leads to **biased** and **spurious** results

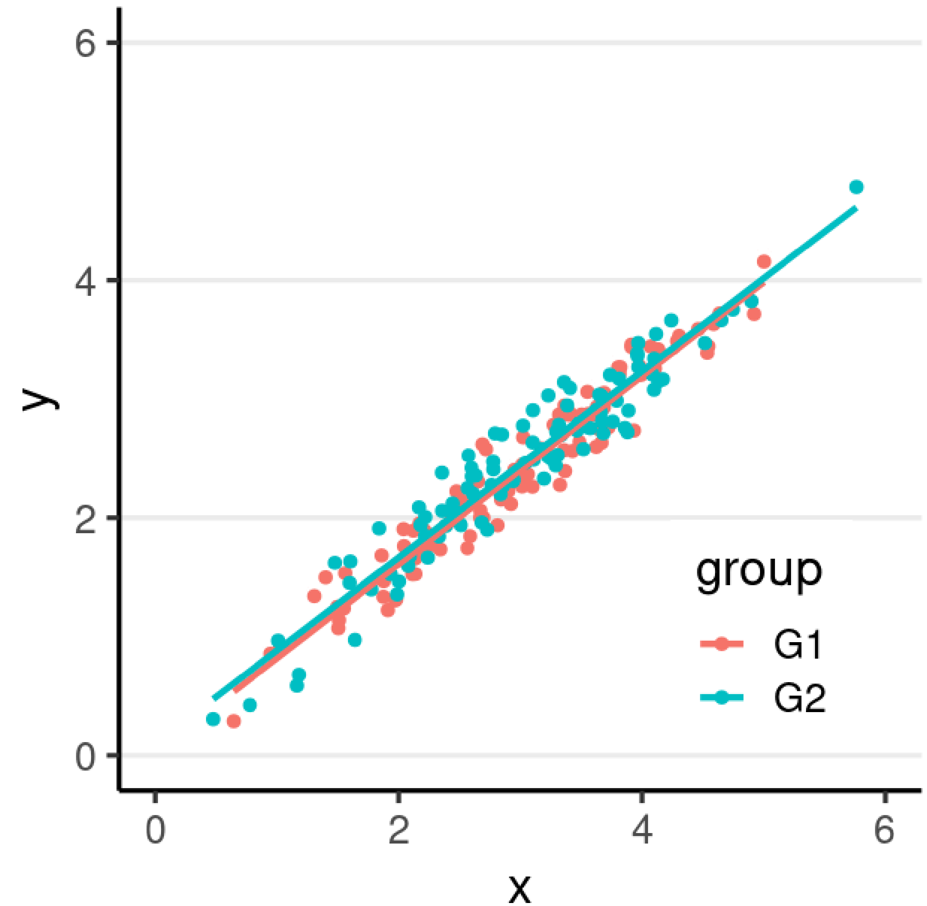
Unreliability

- Bias regression slopes
  - Unpredictable directions in complex models (Cole & Preacher, 2014)



Noninvariance/Differential functioning

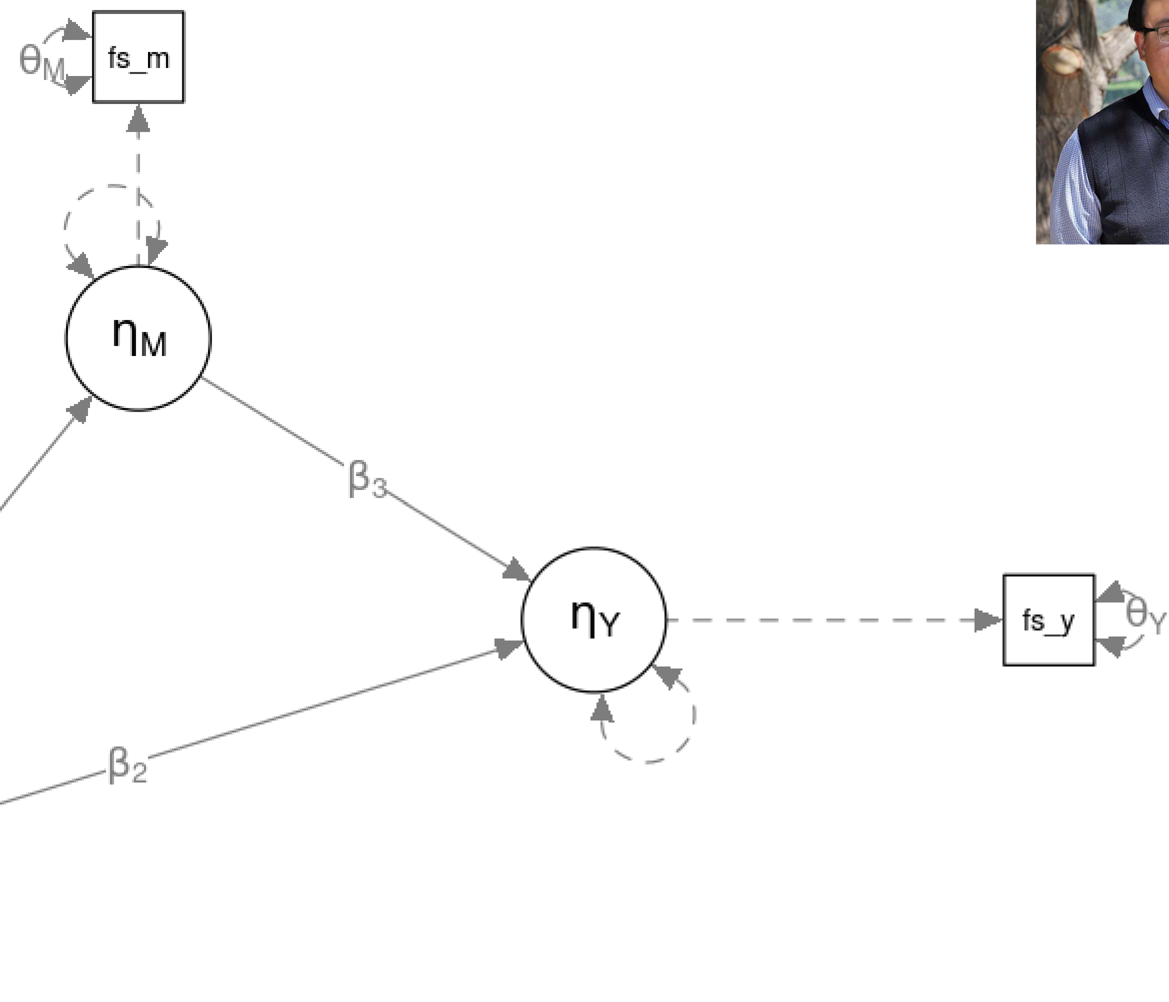
- Biased/Spurious group differences
- Biased/Spurious interactions (Hsiao & Lai, 2018)



Alternative 2: Using Factor Scores

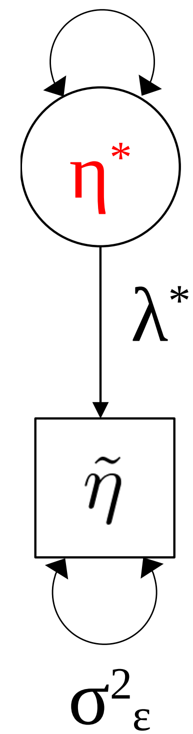
- Factor scores are also not perfectly reliable
- Some factor scores (e.g., regression scores, EAP scores) are not measurement invariant
  - Even when accounting for DIF in the model (Lai and Tse 2023)

Alternative 3: Two-Stage Path Analysis



2S-PA

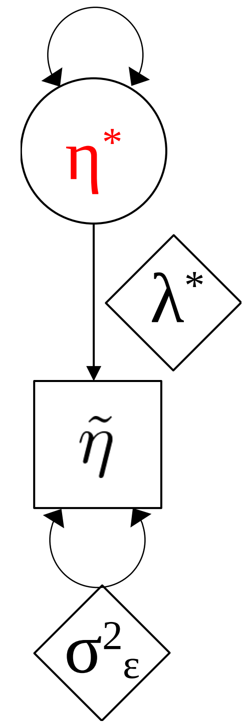
- First stage: Obtain one indicator ($\tilde{\eta}$) per latent construct (η)
  - E.g., regression/Bartlett/sum scores; EAP scores
  - Adjust for **noninvariance**
  - Estimate λ^* and ${\sigma^*}^2_\varepsilon$
- Second stage: Single-indicator model with known loading and error variance



2S-PA With Discrete Items

- Non-constant measurement error variance across observations[1]

- Definition variables
  - Available in OpenMx and Mplus
  - Also Bayesian estimation (e.g., Stan)



2S-PA With Definition Variables

$$\text{Measurement: } \tilde{\boldsymbol{\eta}}_i = \boldsymbol{\Lambda}_{i}^{*} \boldsymbol{\eta}_i^{*} + \boldsymbol{\varepsilon}_i^{*}$$

$$\boldsymbol{\varepsilon}_i^{*} \sim N(\mathbf{0}, \boldsymbol{\Theta}_{i}^{*})$$

$$\text{Structural: } \boldsymbol{\eta}_i^{*} = \boldsymbol{\alpha}^{*} + \mathbf{B}^{*} \boldsymbol{\eta}_i^{*} + \boldsymbol{\zeta}_i^{*}$$

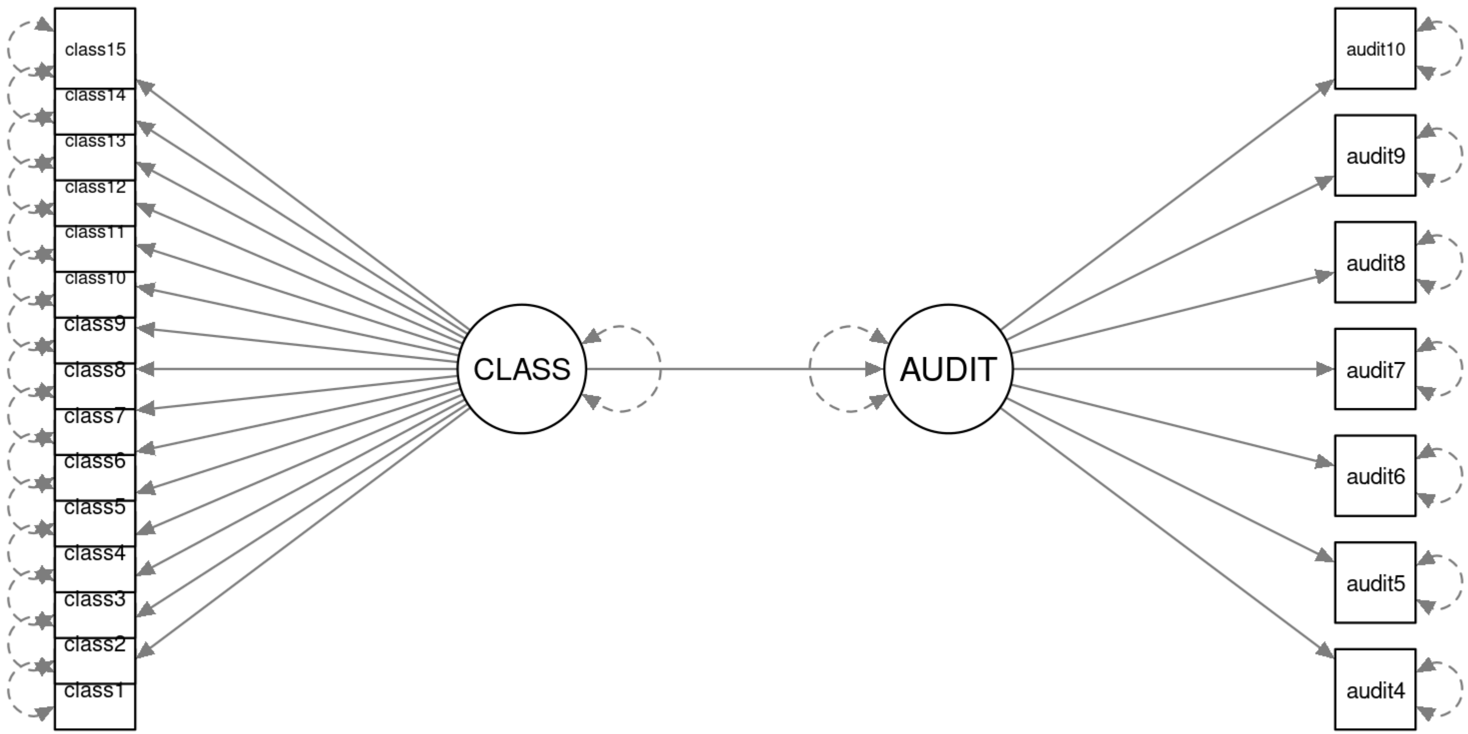
Note

Lai and Hsiao (2022) and Lai et al. (2023) found that, with categorical indicators, 2S-PA yielded

- better convergence rates, less SE bias, better Type I error rate control in small samples, compared to joint SEM modeling (with weighted least squares estimation)

Example 1: Latent Regression

Multiple-group latent regression



- Items on 3-to-5 point scales
- Across 4 ethnic groups (White, Asian, Black, Hispanic)
- Partial scalar invariance for Item 14 in CLASS

Challenges with JM

- One multiple-group model with many invariance constraints for both latent variables
  - 424 measurement parameters
- Two-dimensional numerical integration (with ML)
- DWLS cannot handle missing data

2S-PA

- With separate measurement models and EAP Scores

```
# AUDIT
m1a <- mirt::mirt(
    dat[, paste0("audit",
    verbose = FALSE)
fs_audit <- mirt::fscores(
    m1a, full.scores.SE =
head(fs_audit)
```

```
             F1     SE_F1
[1,] -0.9050792 0.6705626
[2,]  0.1212936 0.4302879
[3,] -0.9050792 0.6705626
[4,]  0.2327338 0.4089029
[5,]  0.2327338 0.4089029
[6,]  1.4184295 0.3354633
```

- EAP scores are shrinkage scores
  - $\tilde{\eta}_i = \lambda_i^* \eta_i + \varepsilon_i^*$
- λ_i^* = shrinkage factor = reliability of $\tilde{\eta}_i$, and
- $\mathrm{SE}^2(\tilde{\eta}_i) = (1 - \lambda_i^*)V(\eta)$

We set $V(\eta) = 1$. As inputs for 2S-PA, we need to obtain λ_i^* and $\tilde{\theta}_i^*$ as

- $\lambda_i^* = 1 - \mathrm{SE}^2(\tilde{\eta}_i)$
- $\theta_i^* = \mathrm{SE}^2(\tilde{\eta}_i)[1 - \mathrm{SE}^2(\tilde{\eta}_i)]$

	F1	SE_F1	loading_i	errorvar_i
1	-0.905	0.671	0.550	0.247
2	0.121	0.430	0.815	0.151
3	-0.905	0.671	0.550	0.247
4	0.233	0.409	0.833	0.139
5	0.233	0.409	0.833	0.139
6	1.418	0.335	0.887	0.100

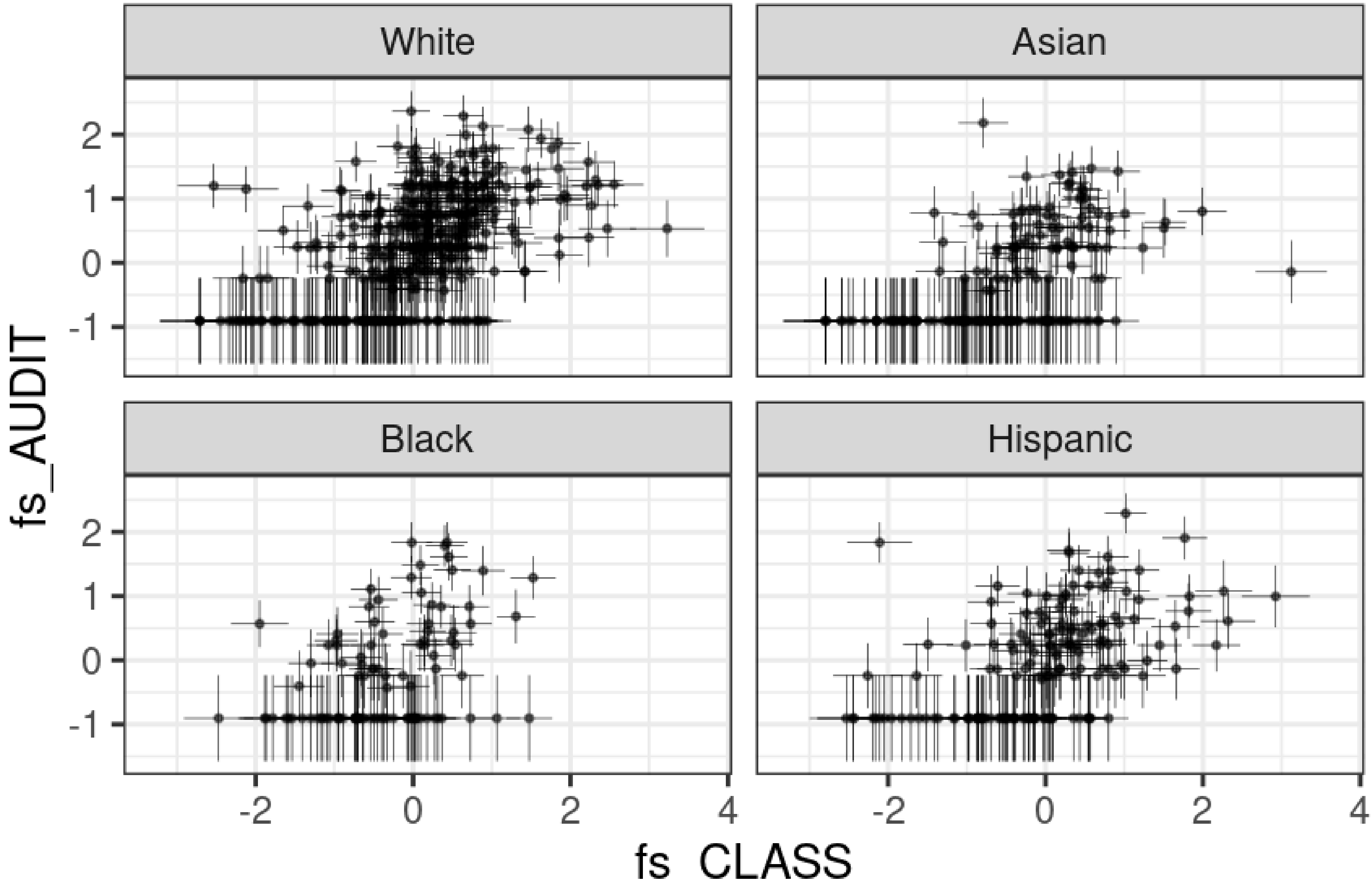
Generalizing to multidimensional measurement models

Software usually gives $\mathrm{ACOV}(\tilde{\boldsymbol{\eta}}_i)$ as output

- $\boldsymbol{\Lambda}_i^* = \mathbf{I} - \mathrm{ACOV}(\tilde{\boldsymbol{\eta}}_i)V(\boldsymbol{\eta})$
- $\boldsymbol{\Theta}_i^* = \boldsymbol{\Lambda}_i^* \mathrm{ACOV}(\tilde{\boldsymbol{\eta}}_i)$

Implementation in R package R2spa

```r
# Prepare data
fs_dat <- fs_dat |>
    within(expr = {
        rel_class <- 1 - class_se^2
        rel_audit <- 1 - audit_se^2
        ev_class <- class_se^2 * (1 - class_se^2)
        ev_audit <- audit_se^2 * (1 - audit_se^2)
    })
# Define model
latreg_umx <- umxLav2RAM(
    "
      fs_audit ~ fs_class
      fs_audit + fs_class ~ 1
    ",
    printTab = FALSE
)
# lambda (reliability)
cross_load <- matrix(c("rel_audit", NA, NA, "rel_class"), nrow = 2) |>
    `dimnames<-`(rep(list(c("fs_audit", "fs_class")), 2))
# Error of factor scores
err_cov <- matrix(c("ev_audit", NA, NA, "ev_class"), nrow = 2) |>
    `dimnames<-`(rep(list(c("fs_audit", "fs_class")), 2))
# Create model in Mx
tspa_mx <- tspa_mx_model(latreg_umx,
    data = fs_dat,
    mat_ld = cross_load, mat_vc = err_cov
)
```


White
Asian
Black
Hispanic
fs_AUDIT
fs_CLASS
2
1
0
-1
-2
0
2
4

Comparison of **standardized coefficients** (CLASS → AUDIT)

	est	se	ci
Joint Modeling[1]	0.614	0.030	[0.556, 0.672]
Factor score regression[2]	0.543	0.024	[0.495, 0.590]
2S-PA[2]	0.669	0.027	[0.617, 0.722]

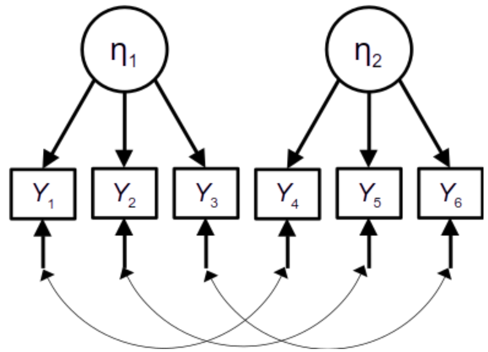
2S-PA is Flexible

- I used MG-IRT for CLASS to model partial invariance, and single-group IRT for AUDIT to assume invariance

But Choices Needed To Be Made . . .

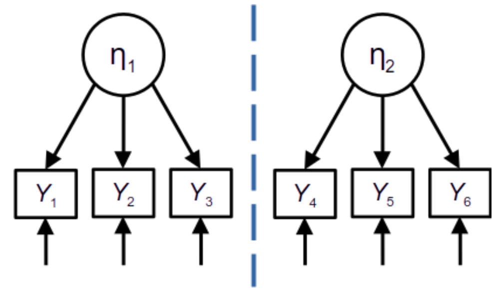
- Joint vs. separate measurement models
- Types of factor scores
- Frequentist vs. Bayesian estimation[1]

Joint: Multidimensional model



- Same complexity as joint modeling
- Needed when there are
  - longitudinal invariance
  - Cross-loadings/error covariances
- Assumes correct measurement model

Separate: Several unidimensional models



- Can use different software for different components
- Less complexity, but less efficiency
- Biased when ignoring misspecification
  - May have some robustness
- Can have separate multidimensional/unidimensional models

Types of Factor Scores

- Sum scores (or mean scores)
- Shrinkage scores
  - Regression scores, EAP scores, MAP scores
- Maximum likelihood (ML) scores
  - Bartlett scores, ML scores in IRT

Simulation Results in Lai et al. (2023)

- All three types of scores performed reasonably well (as long as the right λ^* and Θ^* are used)
- Using sum scores give better RMSE, SE bias, and coverage in small samples/low reliability conditions

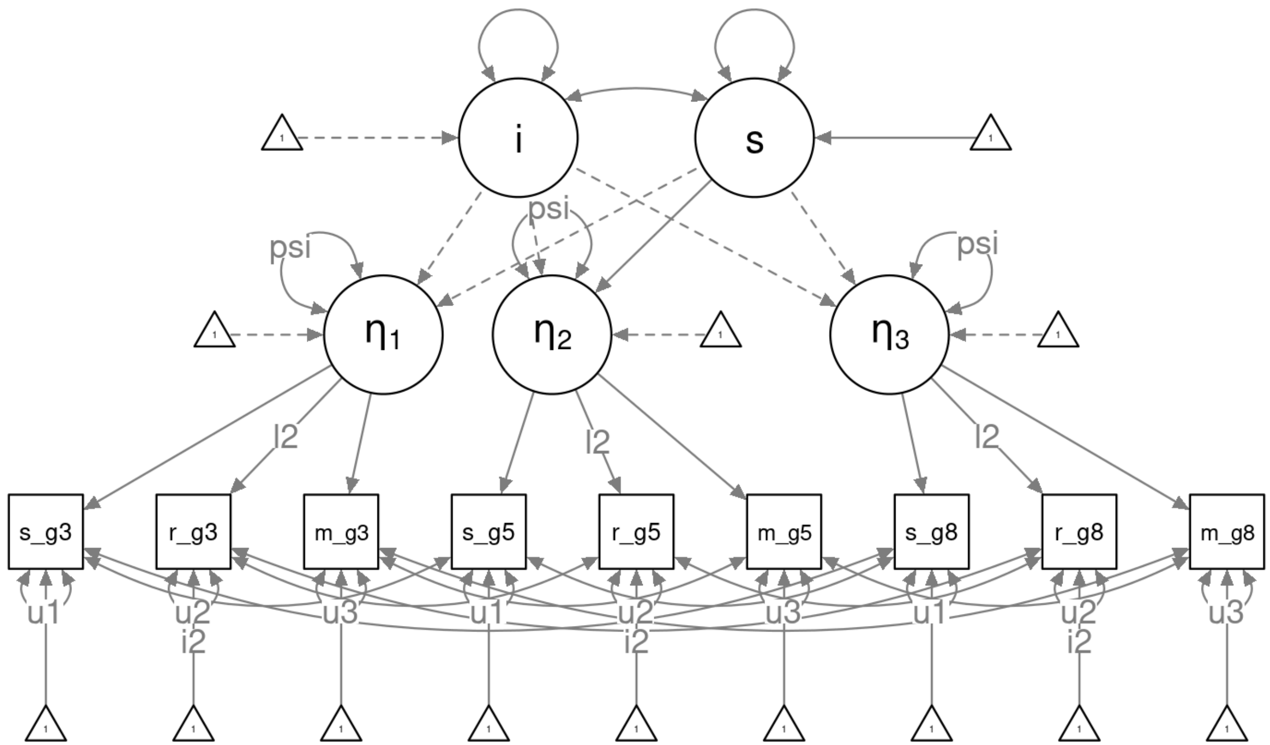
cf. Lai et al. (2023)

	Composite scores	Regression scores[1]	Bartlett scores[2]
Observed variance	$\mathbf{1}^\top \mathbf{\Sigma}_X \mathbf{1}$	$\psi^2 \boldsymbol{\lambda}^\top \mathbf{\Sigma}_X^{-1} \boldsymbol{\lambda}$	$\psi + (\boldsymbol{\lambda}^\top \mathbf{\Theta}^{-1} \boldsymbol{\lambda})^{-1}$
λ^*	$\sum_j \lambda_j$	$\psi \boldsymbol{\lambda}^\top \mathbf{\Sigma}_X^{-1} \boldsymbol{\lambda}$	1
Reliability	$\frac{(\sum_j \lambda_j)^2 \psi}{\mathbf{1}^\top \mathbf{\Sigma}_X \mathbf{1}}$	$\psi \boldsymbol{\lambda}^\top \mathbf{\Sigma}_X^{-1} \boldsymbol{\lambda}$	$\frac{\psi}{\psi + (\boldsymbol{\lambda}^\top \mathbf{\Theta}^{-1} \boldsymbol{\lambda})^{-1}}$

$\mathbf{\Sigma}_X$ = covariance matrix of indicators; ψ = latent variance; λ_j = loading of indicator j; $\mathbf{\Theta}$ = error covariance of indicators

Example 2: Longitudinal Model

ECLS-K: Achievement (Science, Reading, Math) across Grades 3, 5, and 8



Interpretational Confounding

A challenge of joint modeling is that the definition of latent variables can change across models

	Latent Basis	**No Growth**	**Measurement Only**
Science	14.87	18.57	14.83
Reading	21.47	28.19	21.39
Math	20.20	25.93	20.11

↑ Note the loadings change across different models

Longitudinal Model With 2S-PA

- Stage 1a: Longitudinal invariance model
  - configural → metric → scalar → strict (e.g., Widaman and Reise 1997)
  - alignment optimization (Asparouhov and Muthén 2014; Lai 2023)
- Stage 1b: Scoring and measurement properties
  - Regression scores, Bartlett scores, etc
- Stage 2: Growth model with q indicators (q = number of time points)

Note on Scoring

> With cross-loadings and/or correlated errors, scoring should be done with a joint multidimensional factor model

Mean structure

$$\tilde{\boldsymbol{\eta}}_i = \mathbf{b^*}_i + \boldsymbol{\Lambda}_i^* \boldsymbol{\eta}_i^* + \boldsymbol{\varepsilon}_i^*$$

- Bartlett scores are convenient, as generally we have
  - $\mathbf{b}^*$ = 0 and $\boldsymbol{\Lambda}_i^* = \mathbf{I}$
  - But they may be less reliable than regression scores

Sample Code

```r
# Get factor scores from partial scalar invariance model
fs_dat <- R2spa::get_fs(eclsk, model = pscalar_mod)

# Growth model
tspa_growth_mod <- "
i =~ 1 * eta1 + 1 * eta2 + 1 * eta3
s =~ 0 * eta1 + start(.5) * eta2 + 1 * eta3

# factor error variances (assume homogeneity)
eta1 ~~ psi * eta1
eta2 ~~ psi * eta2
eta3 ~~ psi * eta3

i ~~ start(.8) * i
s ~~ start(.5) * s
i ~~ start(0) * s

i + s ~ 1
"
# Fit the growth model
tspa_growth_fit <- tspa(tspa_growth_mod, fs_dat,
                        fsT = attr(fs_dat, "fsT"),
                        fsL = attr(fs_dat, "fsL"),
                        fsb = attr(fs_dat, "fsb"),
                        estimator = "ML")
summary(tspa_growth_fit)
```

Parameter	Model	Est	SE	LRT χ^2
Mean slope	JSEM	1.873	0.025	2223.513
	2S-PA (Reg)	1.874	0.018	2271.428
	2S-PA (Bart)	1.874	0.018	2271.428
	FS (Reg)	1.874	0.010	3282.137
	FS (Bart)	1.874	0.019	2248.001
Var slope	JSEM	0.099	0.017	
	2S-PA (Reg)	0.100	0.016	
	2S-PA (Bart)	0.100	0.016	
	FS (Reg)	0.065	0.004	
	FS (Bart)	0.141	0.016	

Further Adjustment

2S-PA treats $\mathbf{\Lambda}^*$ and $\mathbf{\Theta}^*$ as known

- When these are estimated, and their uncertainty is ignored,
    - SE maybe underestimated in the structural model

Solution 1: Bayesian estimation of factor scores (Lai and Hsiao 2022)

Solution 2: Incorporating SE of $\boldsymbol{\Lambda}^*$ and $\boldsymbol{\Theta}^*$ (Meijer, Oczkowski, and Wansbeek 2021)[1]

First-order correction for SE

$$\hat{V}_{\gamma,c} = \hat{V}_{\gamma} + \boldsymbol{J}_{\boldsymbol{\gamma}}(\hat{\boldsymbol{\theta}})\hat{V}_{\theta}\boldsymbol{J}_{\boldsymbol{\gamma}}(\hat{\boldsymbol{\theta}})^{\top},$$

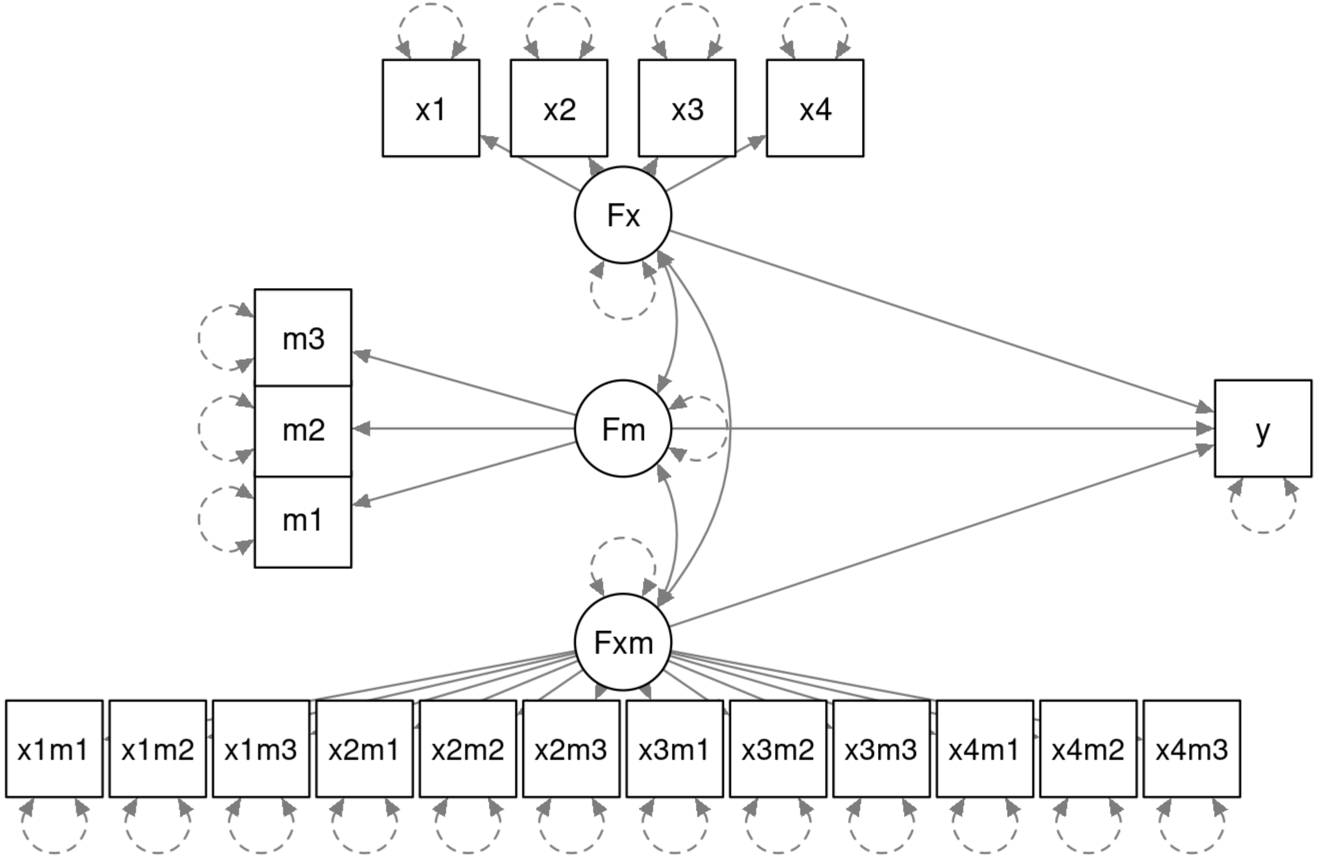
where $\boldsymbol{J}_{\boldsymbol{\gamma}}$ is the Jacobian matrix of $\hat{\boldsymbol{\gamma}}$ with respect to $\boldsymbol{\theta}$, or

$$\hat{V}_{\gamma,c} = \hat{V}_{\gamma} + (\boldsymbol{H}_{\gamma})^{-1}\left(\frac{\partial^2 \ell}{\partial\theta\partial\gamma^{\top}}\right)\hat{V}_{\theta}\left(\frac{\partial^2 \ell}{\partial\theta\partial\gamma^{\top}}\right)^{\top}(\boldsymbol{H}_{\gamma})^{-1},$$

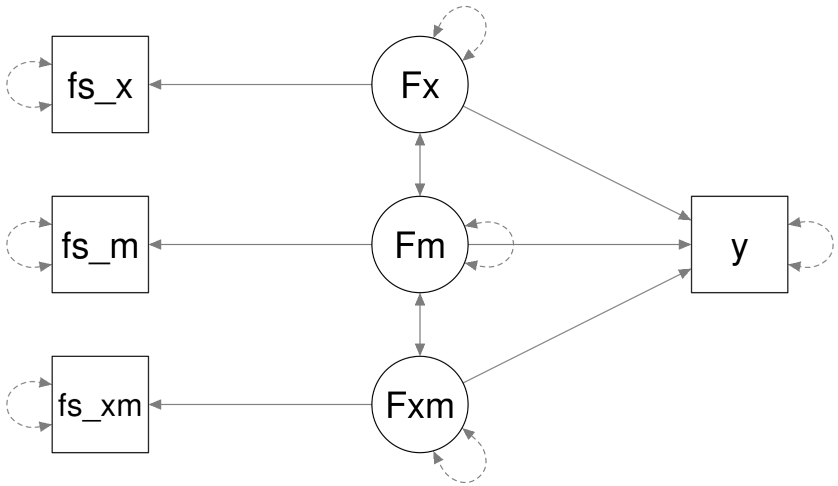
where V_{γ} is the naive covariance matrix of the structural parameter estimates $\hat{\boldsymbol{\gamma}}$ assuming the measurement error variance parameter, $\boldsymbol{\theta}$, is known, $\boldsymbol{H}_{\gamma}$ is the Hessian matrix of the log-likelihood ℓ with respect to $\hat{\boldsymbol{\gamma}}$, and V_{θ} can be obtained in the first-stage measurement model analysis.

Extension: Latent Interactions

Tedious to do product indicators



With 2S-PA, just one product factor score indicator



- Bias and SE bias for 2S-PA-Int was in acceptable range in all conditions
- Overall, better coverage and RMSE than product indicators

Extension: Location-Scale Modeling

With measurement error

- Predicting individual-specific mean (location) and fluctuation/variance (scale) over time

Estimates are virtually identical to those with joint modeling

Blozis & Lai (2024, IMPS Proceedings); code available at https://github.com/marklhc/2s-

Other Extensions Underway

- Latent interaction with categorical indicators
- Location scale model with partial invariance
- Random coefficients from multilevel models
  - E.g., individual-specific slope for self-efficacy → individual-specific slope for achievement
- Vector autoregressive modeling (Rein, Vermunt, & de Roover, preprint)

Limitations/Future Work

- Account for uncertainty is $\mathbf{\Lambda}_i^*$, $\mathbf{\Theta}_i^*$, and $\mathbf{b}_i^*$
- Requires error covariance matrix of factor scores
  - Or some estimates of reliability
- Incorporate auxiliary variables for missing data
  - And potentially applicable to multiply imputed data
- More simulation results

Acknowledgment

Undergraduate and Graduate students

- Yixiao Li
- Meltem Ozcan
- Wing-Yee (Winnie) Tse
- Gengrui (Jimmy) Zhang
- Yichi Zhang

Collaborators

- Shelley Blozis
- Yu-Yu Hsiao
- George B. Richardson
- Dave Raichlen

Thank You!

hokchiol@usc.edu

References

Asparouhov, Tihomir, and Bengt Muthén. 2014. "Multiple-Group Factor Analysis Alignment." *Structural Equation Modeling: A Multidisciplinary Journal* 21 (4): 495–508. https://doi.org/10.1080/10705511.2014.919210.

Croon, M.A. 2002. "Using Predicted Latent Scores in General Latent Structure Models." In *Latent Variable and Latent Structure Models*, edited by G.A. Marcoulides and I. Moustaki, 195–224. Mahwah, NJ: Lawrence Erlbaum.

Lai, Mark H. C. 2023. "Adjusting for Measurement Noninvariance with Alignment in Growth Modeling." *Multivariate Behavioral Research* 58 (1): 30–47. https://doi.org/10.1080/00273171.2021.1941730.

Lai, Mark H. C., and Yu-Yu Hsiao. 2022. "Two-Stage Path Analysis with Definition Variables: An Alternative Framework to Account for Measurement Error." *Psychological Methods* 27 (4): 568–88. https://doi.org/10.1037/met0000410.

Lai, Mark H. C., and Winnie Wing-Yee Tse. 2023. "Are Factor Scores Measurement Invariant?" Preprint. PsyArXiv. https://doi.org/10.31234/osf.io/uzrak.

Lai, Mark H. C., Winnie Wing-Yee Tse, Gengrui Zhang, Yixiao Li, and Yu-Yu Hsiao. 2023. "Correcting for Unreliability and Partial Invariance: A Two-Stage Path Analysis

Approach." *Structural Equation Modeling: A Multidisciplinary Journal* 30 (2): 258–71. https://doi.org/10.1080/10705511.2022.2125397.

Lui, P Priscilla. 2019. "College Alcohol Beliefs: Measurement Invariance, Mean Differences, and Correlations with Alcohol Use Outcomes Across Sociodemographic Groups." *Journal of Counseling Psychology* 66 (4): 487–95. https://doi.org/10.1037/cou0000338.

Meijer, Erik, Edward Oczkowski, and Tom Wansbeek. 2021. "How Measurement Error Affects Inference in Linear Regression." *Empirical Economics* 60 (1): 131–55. https://doi.org/10.1007/s00181-020-01942-z.

Thissen, David, and Anne Thissen-Roe. 2020. "Factor Score Estimation from the Perspective of Item Response Theory." In *Quantitative Psychology: 84th Annual Meeting of the Psychometric Society, Santiago, Chile, 2019*, edited by Marie Wiberg, Dylan Molenaar, Jorge González, Ulf Böckenholt, and Jee-Seon Kim, 322:171–84. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-43469-4_14.

Widaman, Keith F., and Steven P. Reise. 1997. "Exploring the Measurement Invariance of Psychological Instruments: Applications in the Substance Use Domain." In *The Science of Prevention: Methodological Advances from Alcohol and Substance Abuse Research.*, edited by Kendall J. Bryant, Michael Windle, and Stephen G. West, 281–324. Washington: American Psychological Association. https://doi.org/10.1037/10222-009.