Multilevel Structural Equation Modeling

Sophia Rabe-Hesketh

University of California, Berkeley

& Institute of Education, University of London

Joint work with Anders Skrondal, London School of Economics

& Xiaohui Zheng, University of California, Berkeley

Winemiller Conference

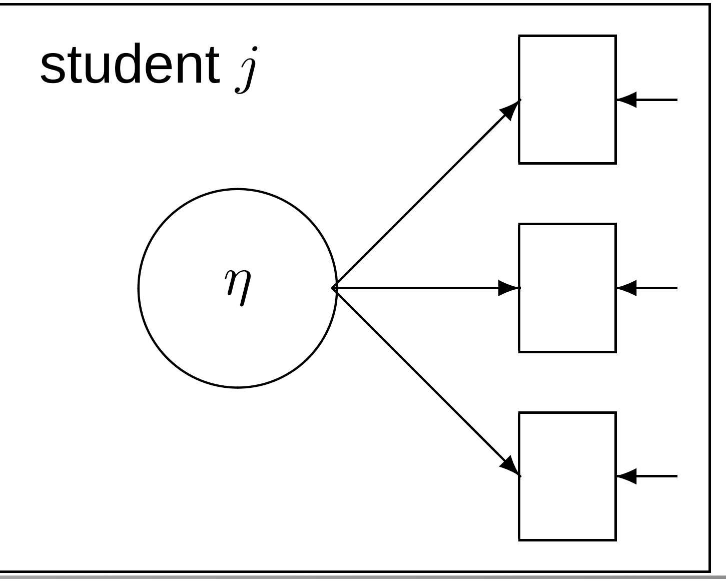
Columbia, MO, October, 2006

Outline

- IRT (Item Response Theory)
  - **Non-continuous indicators**
- Multilevel MIMIC (Multiple Indicator Multiple Cause) model
  - **Cross-level effects**
  - GLLAMM (Generalized Linear Latent and Mixed Model) specification
  - Traditional specification via within and between models
- Multilevel SEM (Structural Equation Model)
  - **Higher-level items** (or responses)
  - GLLAMM specification

Item response theory (IRT) models

- Also known as categorical factor analysis
- Typically, dichotomous response y_{ij} for item i and person j
- Canonical example:
  Ability testing, $y_{ij} = 1$ if person j answers item i correctly, 0 otherwise
- Assume that response probabilities depend on latent ability η_j
- Responses y_{ij} conditionally independent given η_j



One-parameter logistic (1-PL) model

- One-parameter logistic (1-PL) model:

$$\text{logit}[\Pr(y_{ij} = 1 \mid \eta_j)] \;=\; \nu_{ij} \;=\; \beta_i + \eta_j$$

$$\eta_j \sim \mathrm{N}(0, \psi)$$

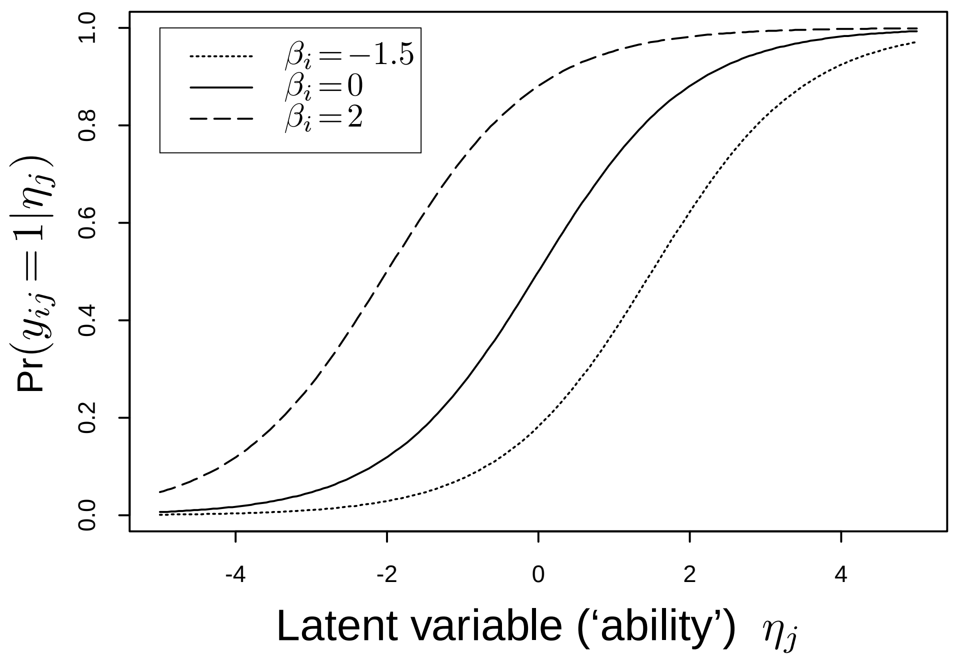
  - η_j is the 'ability' of person j (latent trait or common factor)
  - One parameter β_i (an intercept) for each item
    - $-\beta_i$ is 'difficulty' of item i

$$\Pr(y_{ij} = 1 | \eta_j = -\beta_i) = 0.5$$

- A logistic random intercept model without observed covariates
- Called the Rasch model when η_j treated as fixed parameters

Item characteristic curve for 1-PL

$$\Pr(y_{ij} = 1 \mid \eta_j) = \frac{\exp(\beta_i + \eta_j)}{1 + \exp(\beta_i + \eta_j)}$$



- Double-monotonicity:
  - for each ability, performance decreases with difficulty
  - for each difficulty, performance increases with ability

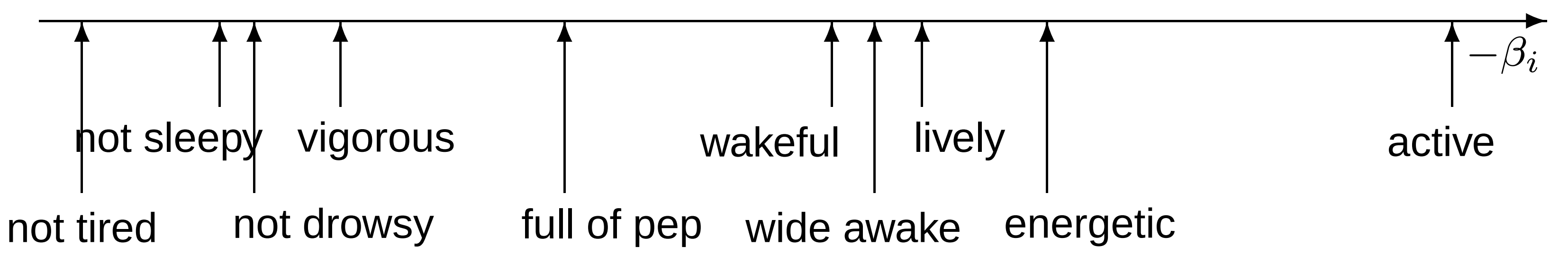
Items and persons placed on common scale in 1-PL

- Log odds, and hence probability, of correct response only depends on location η_j of person j relative to location $-\beta_i$ of item i on same scale

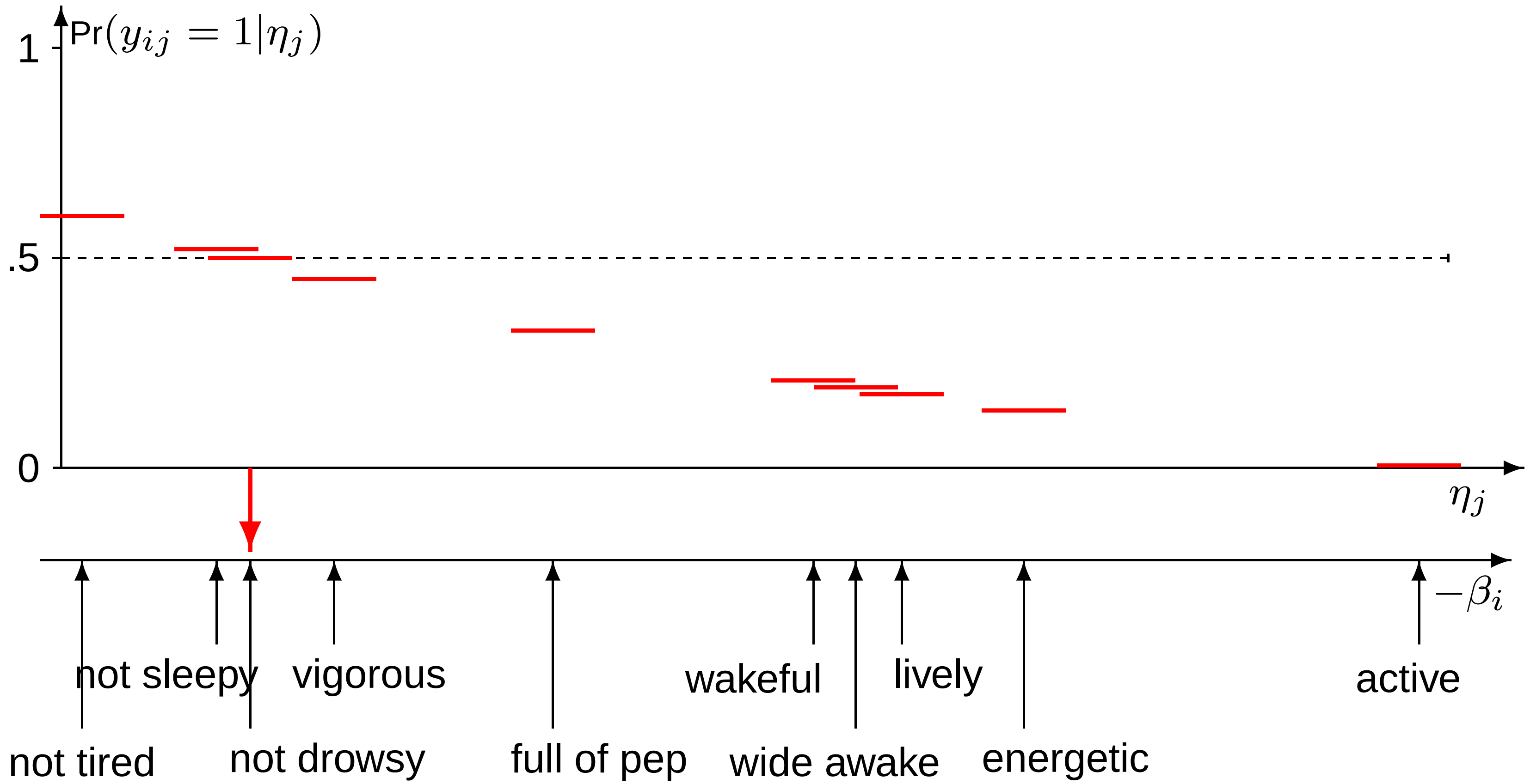
$$\nu_{ij} = \eta_j + \beta_i = \eta_j - (-\beta_i)$$

  - ν_{ij} is 'dominance': how much ability η_j exceeds difficulty $-\beta_i$
- Example: Energetic arousal measured by 10 items

[Embretson & Reise, 2000]:

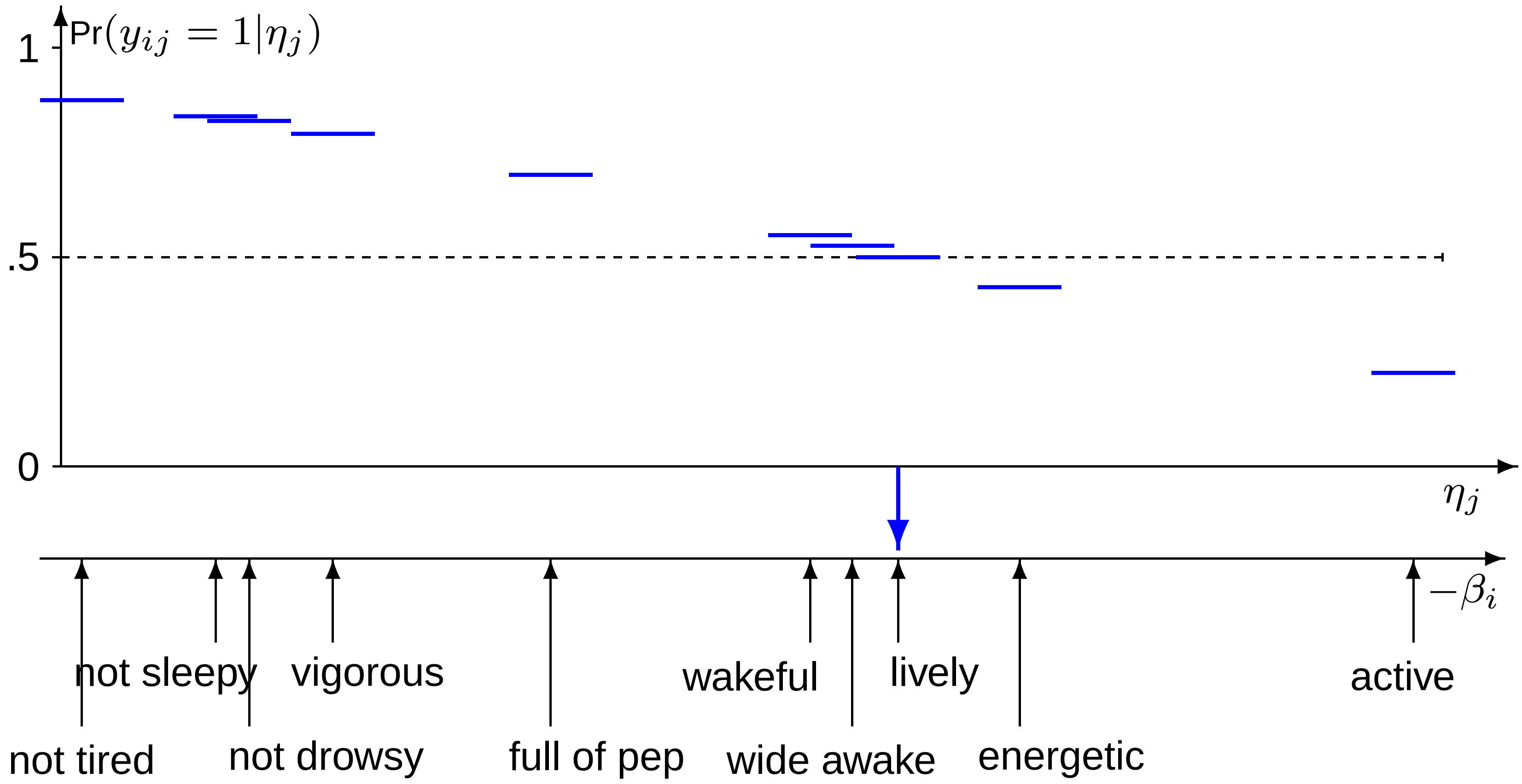


Person characteristic 'curve' in 1-PL



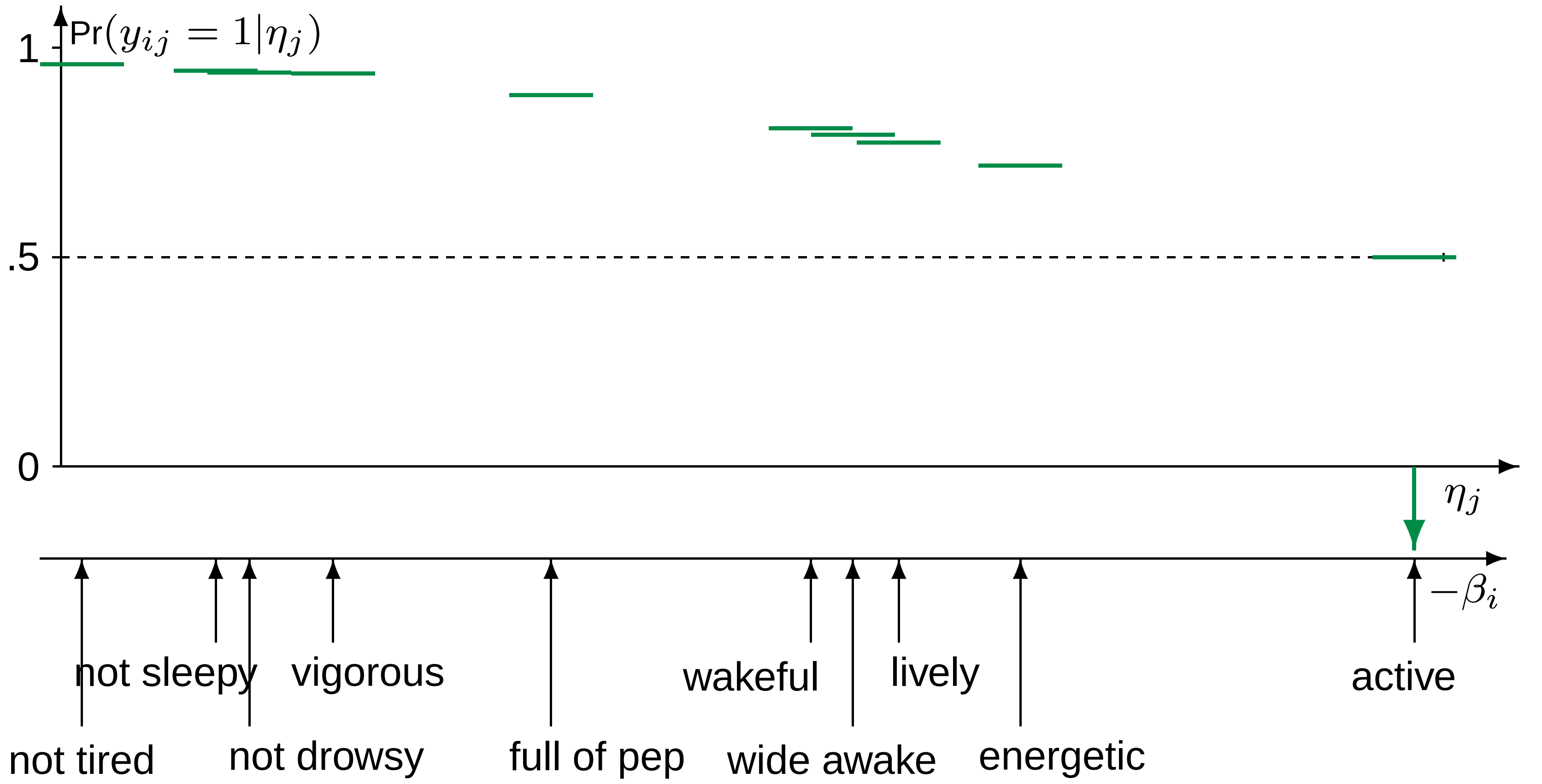
- Pr$(y_{ij} = 1|\eta_j) = 0.5$ when $\eta_j = -\beta_i$

Person characteristic ‘curve’ in 1-PL



- $\Pr(y_{ij} = 1|\eta_j) = 0.5$ when $\eta_j = -\beta_i$

Person characteristic 'curve' in 1-PL



- Pr$(y_{ij} = 1|\eta_j) = 0.5$ when $\eta_j = -\beta_i$

Two-parameter logistic (2-PL) model

- Two-parameter logistic (2-PL) model

$$\text{logit}[\text{Pr}(y_{ij} = 1 \mid \eta_j)] \; = \; \nu_{ij} \; = \; \beta_i + \lambda_i \eta_j$$

$$\eta_j \sim \text{N}(0, 1)$$

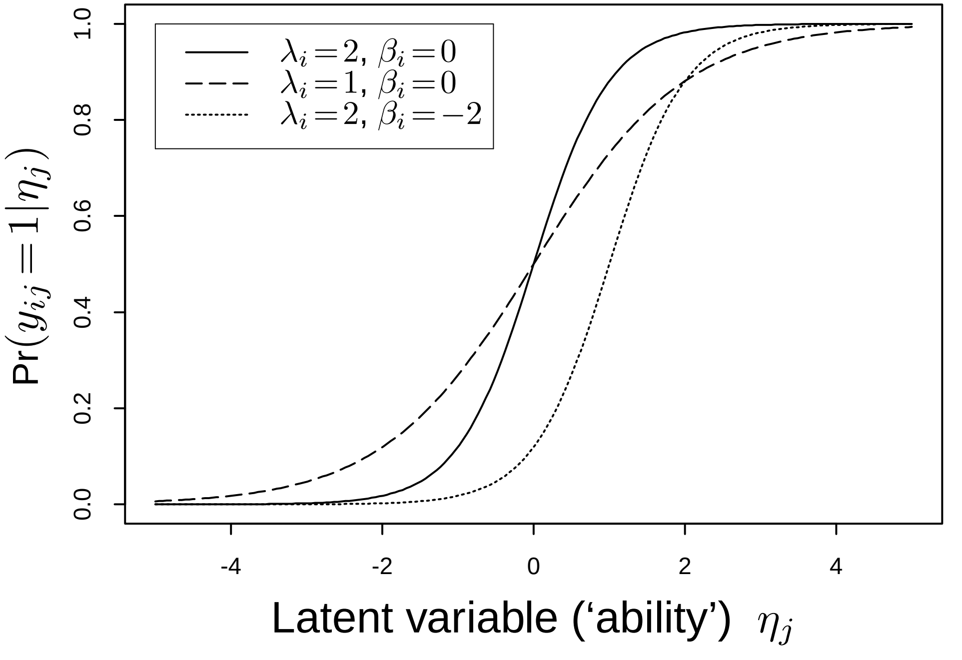
  - λ_i is the 'discrimination parameter' for item i (factor loading)
  - $-\beta_i/\lambda_i$ is the 'difficulty' of item i (β_i is an intercept)

$$\text{Pr}(y_{ij} = 1 | \eta_j = -\beta_i/\lambda_i) = 0.5$$

- One-factor model with logit link
- Usual IRT parameterization: $\nu_{ij} = a_i(\theta_j - b_i)$

Item characteristic curve for 2-PL

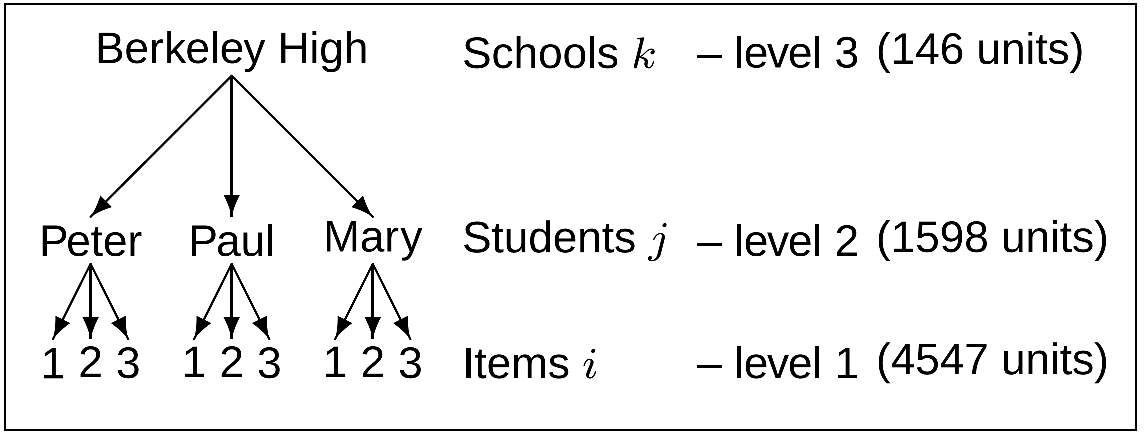
$$\Pr(y_{ij} = 1 \mid \eta_j) = \frac{\exp(\beta_i + \lambda_i \eta_j)}{1 + \exp(\beta_i + \lambda_i \eta_j)}$$



- No double monotonicity!

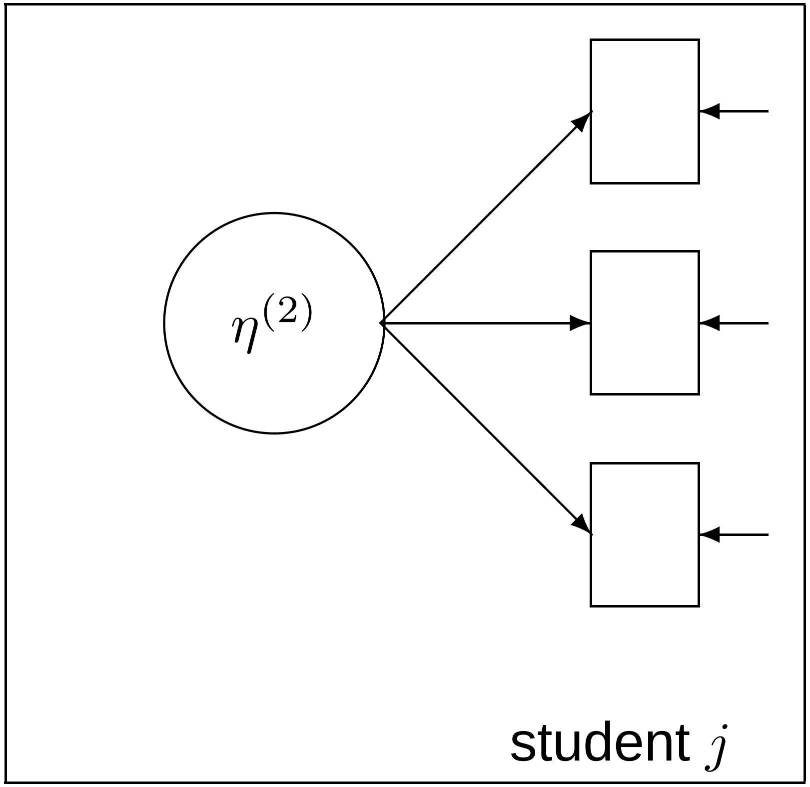
Analysis of U.S. PISA 2000 data

- Three-level data:
  (ignore PSUs here)

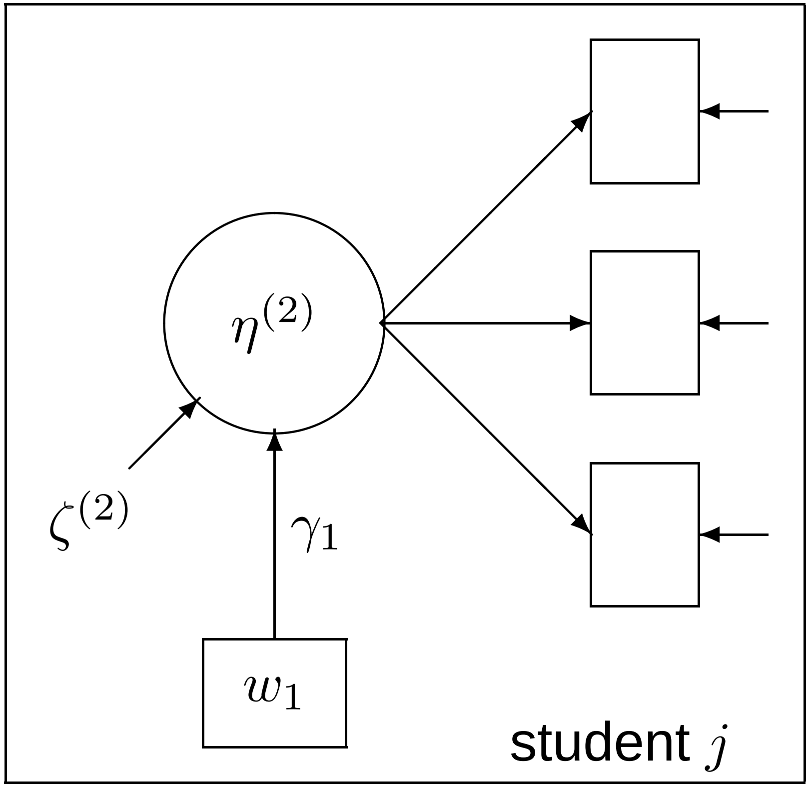


- Student-level covariates
  - [Female]: Student is female (dummy)
  - [ISEI]: International socioeconomic index
  - [Highschool]: Highest education level by either parent is high school (dummy)
  - [College]: Highest education level by either parent is college (dummy)
  - [English]: Test language (English) spoken at home (dummy)
- School-level covariate
  - [MnISEI]: School mean ISEI

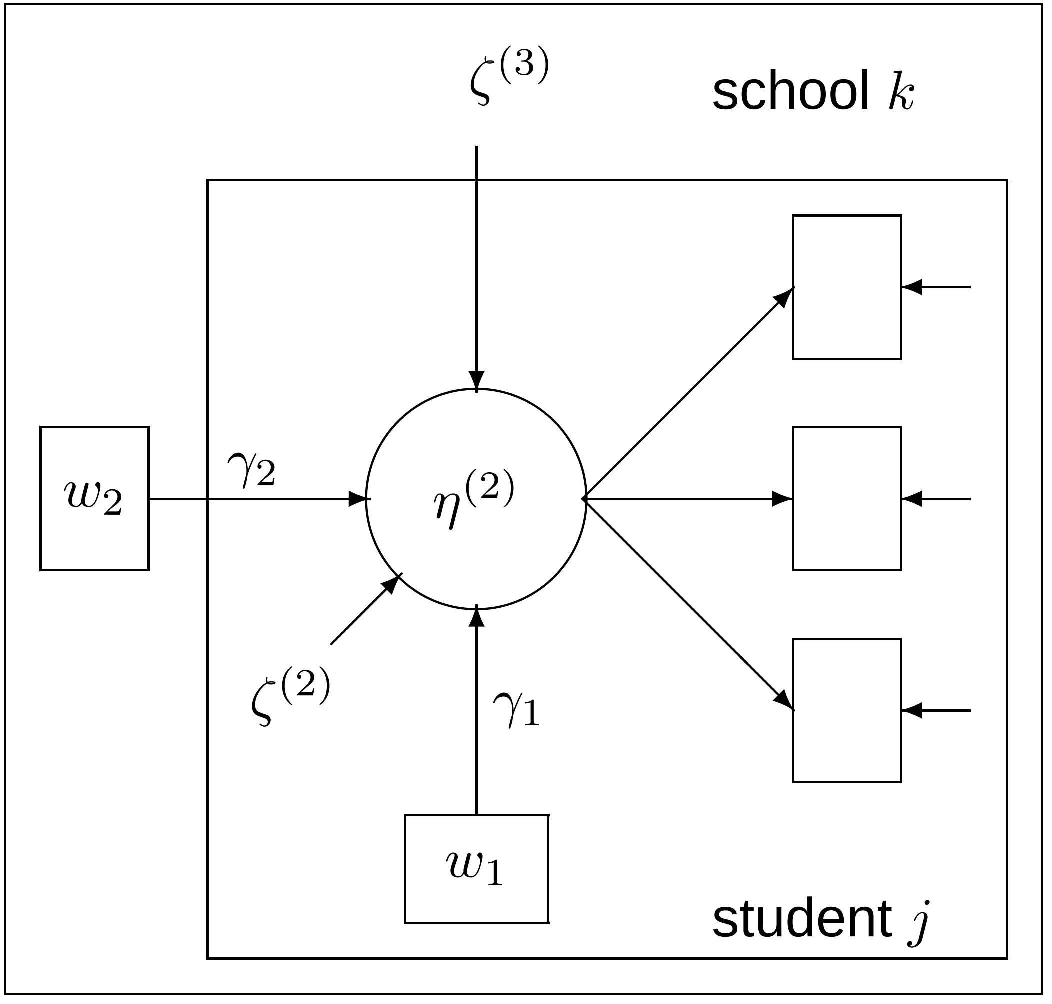
From item response model
to multilevel structural equation or MIMIC model



From item response model to multilevel structural equation or MIMIC model



From item response model to multilevel structural equation or MIMIC model



Multilevel MIMIC model

- **Response model:**
  Two-parameter logistic item response model for item i ($i = 1, \ldots, 7$):

$$\nu_{ijk} \;=\; \beta_i + \lambda_i \eta_{jk}^{(2)}$$

- **Structural model:**
  Two-level linear random intercept model for latent ability of student j in school k:

$$\eta_{jk}^{(2)} \;=\; \mathbf{w}'_{1jk}\gamma_1 + \mathbf{w}'_{2k}\gamma_2 + \zeta_{jk}^{(2)} + \zeta_k^{(3)}$$

- Disturbances at levels 2 (students j) and 3 (schools k) independent across levels,

$$\zeta_{jk}^{(2)} \;\sim\; \mathrm{N}(0, \psi^{(2)})$$

$$\zeta_{k}^{(3)} \;\sim\; \mathrm{N}(0, \psi^{(3)})$$

Taking into account PSUs and survey weights

- Three-stage survey
  - Stage 1 (Primary Sampling Units): Geographic areas
  - Stage 2 Schools k, sampled with probabilities π_j, $w_k = 1/\pi_k$
  - Stage 3 Students j, sampled with probabilities $\pi_{j|k}$, $w_{j|k} = 1/\pi_{j|k}$
- Log likelihood for three-level model

$$\sum_{k=1}^{n^{(3)}} \log \int \exp\left\{\sum_{j=1}^{n_k^{(2)}} \log \int \exp\left[\sum_{i=1}^{n_{jk}^{(1)}} \mathcal{L}(y_{ijk}|\zeta_{jk}^{(2)}, \zeta_k^{(3)})\right] g(\zeta_{jk}^{(2)}) \mathrm{d}\zeta_{jk}^{(2)}\right\} g(\zeta_k^{(3)}) \mathrm{d}\zeta_k^{(3)}$$

- Log Pseudolikelihood for three-level model

$$\sum_{k=1}^{n^{(3)}} w_k \log \int \exp\left\{\sum_{j=1}^{n_k^{(2)}} w_{j|k} \log \int \exp\left[\sum_{i=1}^{n_{jk}^{(1)}} \mathcal{L}(y_{ijk}|\zeta_{jk}^{(2)}, \zeta_k^{(3)})\right] g(\zeta_{jk}^{(2)}) \mathrm{d}\zeta_{jk}^{(2)}\right\} g(\zeta_k^{(3)}) \mathrm{d}\zeta_k^{(3)}$$

Taking into account PSUs and survey weights (cont'd)

- Conventional standard errors not appropriate with sampling weights
- **Sandwich estimator** of standard errors (Taylor linearization)

$$\widehat{\text{Cov}}(\widehat{\boldsymbol{\vartheta}}) = \widehat{\mathcal{I}}^{-1}\widehat{\mathcal{J}}\widehat{\mathcal{I}}^{-1}$$

  - $\mathcal{J}$: Expectation of outer product of gradients, approximated using PSU contributions to gradients
  - $\mathcal{I}$: Expected information, approximated by observed information
- Sandwich estimator accounts for
  - Sampling weights
  - Clustering at levels 'above' highest level of multilevel model
  - (Stratification at stage 1)
- Adaptive quadrature, pseudolikelihood, and sandwich estimator implemented in `gllamm`

Estimates for multilevel MIMIC model:

Structural model

	Unweighted Maximum likelihood		Weighted Pseudo maximum likelihood		
Parameter	Est	(SE)	Est	(SE_R)	(SE_R^{PSU})
γ_1: [Female]	0.146	(0.122)	0.107	(0.201)	(0.241)
γ_2: [ISEI]	0.012	(0.004)	0.021	(0.008)	(0.007)
γ_3: [Highschool]	0.138	(0.249)	0.056	(0.472)	(0.357)
γ_4: [College]	0.411	(0.263)	0.101	(0.449)	(0.413)
γ_5: [English]	0.555	(0.227)	0.568	(0.230)	(0.252)
γ_6: [MnISEI]	0.039	(0.012)	0.020	(0.014)	(0.016)
$\psi^{(2)}$	1.244	(0.642)	1.201	(0.835)	(0.761)
$\psi^{(3)}$	0.111	(0.136)	0.051	(0.129)	(0.111)

ICC=0.08 (ICC=0.14, not controlling for [MnISEI])

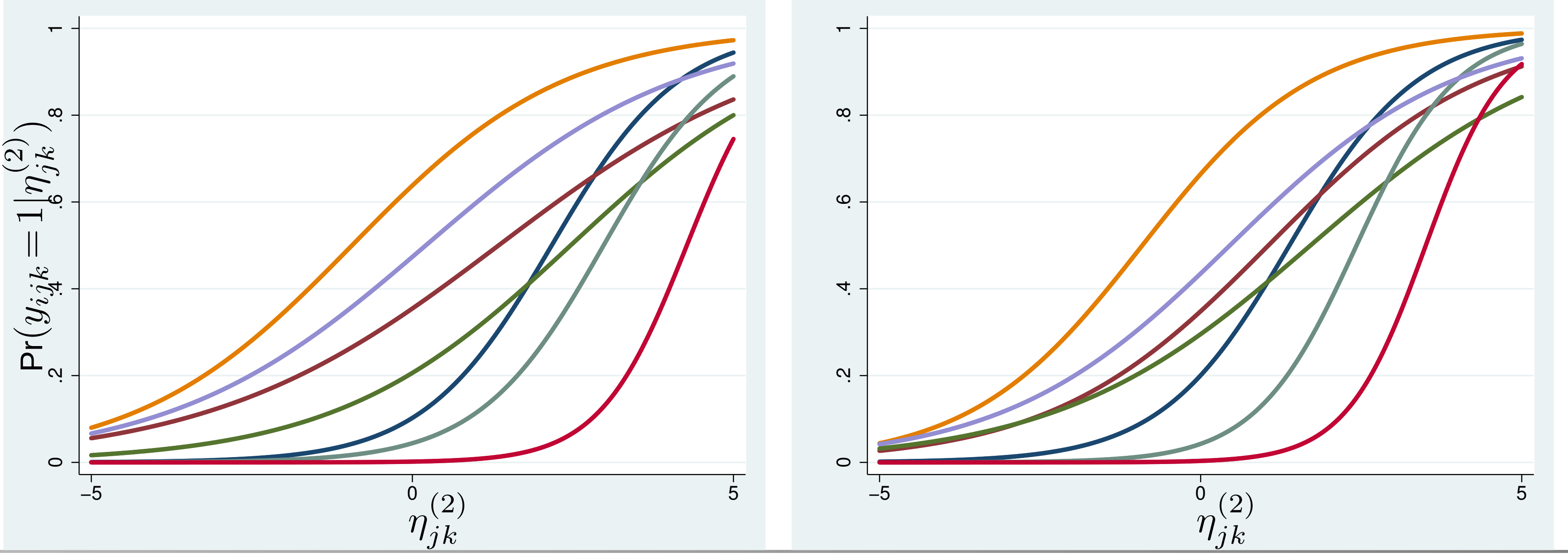
Estimates for multilevel MIMIC model:

Measurement model

- Item characteristic curves for two-parameter logistic model

$$\text{Pr}(y_{ijk}=1|\eta_{jk}^{(2)}) = \frac{\exp(\beta_i + \lambda_i \eta_{jk}^{(2)})}{1+\exp(\beta_i + \lambda_i \eta_{jk}^{(2)})}$$

Unweighted | Weighted



Cross-level effects: Direct paths from higher-level to lower-level latent variables

- GLLAMM structural model:

$$\boldsymbol{\eta} \;=\; \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\mathbf{w} + \boldsymbol{\zeta}$$

  - $\boldsymbol{\eta}$ is the vector of all latent variables
  - $\mathbf{B}$ is a an upper triangular matrix of regression coefficients
  - $\boldsymbol{\Gamma}$ is a matrix of regression coefficients
  - $\mathbf{w}$ is a vector of observed covariates
  - $\boldsymbol{\zeta}$ is a vector of disturbances

$$\begin{aligned}
\boldsymbol{\eta} &= (\overbrace{\eta_1^{(2)}, \eta_2^{(2)}, \ldots, \eta_{M_2}^{(2)}}^{\text{Level 2}}, \ldots, \overbrace{\eta_1^{(l)}, \ldots, \eta_{M_l}^{(l)}}^{\text{Level } l}, \ldots, \eta_{M_L}^{(L)}) \\
\boldsymbol{\zeta} &= (\zeta_1^{(2)}, \zeta_2^{(2)}, \ldots, \zeta_{M_2}^{(2)}, \ldots, \zeta_1^{(l)}, \ldots, \zeta_{M_l}^{(l)}, \ldots, \zeta_{M_L}^{(L)})
\end{aligned}$$

GLLAMM specification of MIMIC structural model

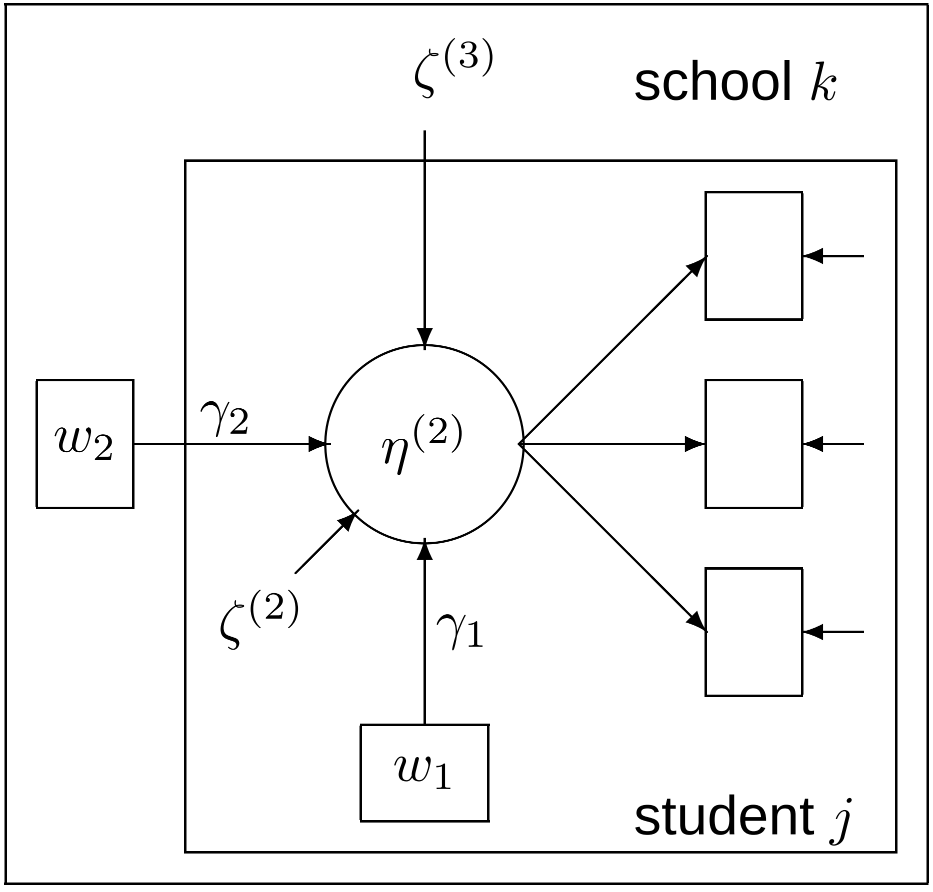
$$\boldsymbol{\eta} \;=\; \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\mathbf{w} + \boldsymbol{\zeta}$$

$$\underbrace{\begin{bmatrix} \eta_{jk}^{(2)} \\ \eta_k^{(3)} \end{bmatrix}}_{\boldsymbol{\eta}} = \underbrace{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} \eta_{jk}^{(2)} \\ \eta_k^{(3)} \end{bmatrix}}_{\boldsymbol{\eta}} + \underbrace{\begin{bmatrix} \gamma_1 & \gamma_2 \\ 0 & 0 \end{bmatrix}}_{\boldsymbol{\Gamma}} \underbrace{\begin{bmatrix} w_{1jk} \\ w_{2k} \end{bmatrix}}_{\mathbf{w}} + \underbrace{\begin{bmatrix} \zeta_{jk}^{(2)} \\ \zeta_k^{(3)} \end{bmatrix}}_{\boldsymbol{\zeta}}$$

$$\eta_k^{(3)} = \zeta_k^{(3)}$$

$$\eta_{jk}^{(2)} = \gamma_1 w_{1jk} + \gamma_2 w_{2k} + \underbrace{\zeta_k^{(3)}}_{\eta_k^{(3)}} + \zeta_{jk}^{(2)}$$

Path diagram for GLLAMM formulation



Traditional formulation in terms of within-model and between-model (cont. responses)

- Within and between covariance matrices

$$\begin{aligned} \mathbf{y}_{jk} &\sim \mathrm{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_W) \\ \boldsymbol{\mu}_k &\sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_B). \end{aligned}$$

- Specify covariance structure models for $\boldsymbol{\Sigma}_W$ and $\boldsymbol{\Sigma}_B$
  - Covariance structures: two-level factor model (no covariates)

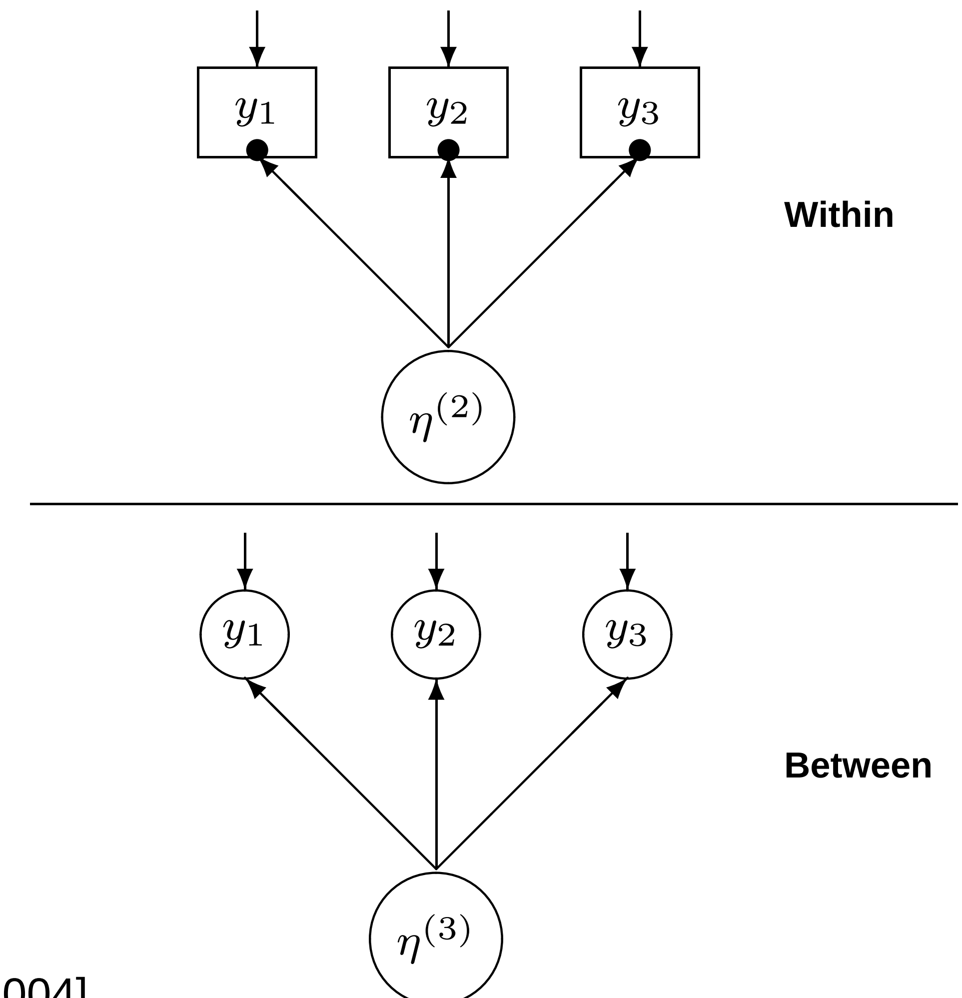
$$\begin{aligned} \boldsymbol{\Sigma}_W &= \boldsymbol{\Lambda}^{(2)}\boldsymbol{\Psi}^{(2)}\boldsymbol{\Lambda}^{(2)\prime} + \boldsymbol{\Theta}^{(2)} \\ \boldsymbol{\Sigma}_B &= \boldsymbol{\Lambda}^{(3)}\boldsymbol{\Psi}^{(3)}\boldsymbol{\Lambda}^{(3)\prime} + \boldsymbol{\Theta}^{(3)} \end{aligned}$$

  - Direct specification of two-level factor model

$$\begin{aligned} \mathbf{y}_{jk} &= \boldsymbol{\mu}_k + \boldsymbol{\Lambda}^{(2)}\boldsymbol{\eta}^{(2)}_{jk} + \boldsymbol{\epsilon}^{(2)}_{jk}, \quad \boldsymbol{\eta}^{(2)}_{jk} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Psi}^{(2)}), \quad \boldsymbol{\epsilon}^{(2)}_{jk} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Theta}^{(2)}) \\ \boldsymbol{\mu}_k &= \boldsymbol{\mu} + \boldsymbol{\Lambda}^{(3)}\boldsymbol{\eta}^{(3)}_{k} + \boldsymbol{\epsilon}^{(3)}_{k}, \quad \boldsymbol{\eta}^{(3)}_{k} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Psi}^{(3)}), \quad \boldsymbol{\epsilon}^{(3)}_{k} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Theta}^{(3)}) \end{aligned}$$

Path diagram for multilevel factor model: within-model and between-model formulation

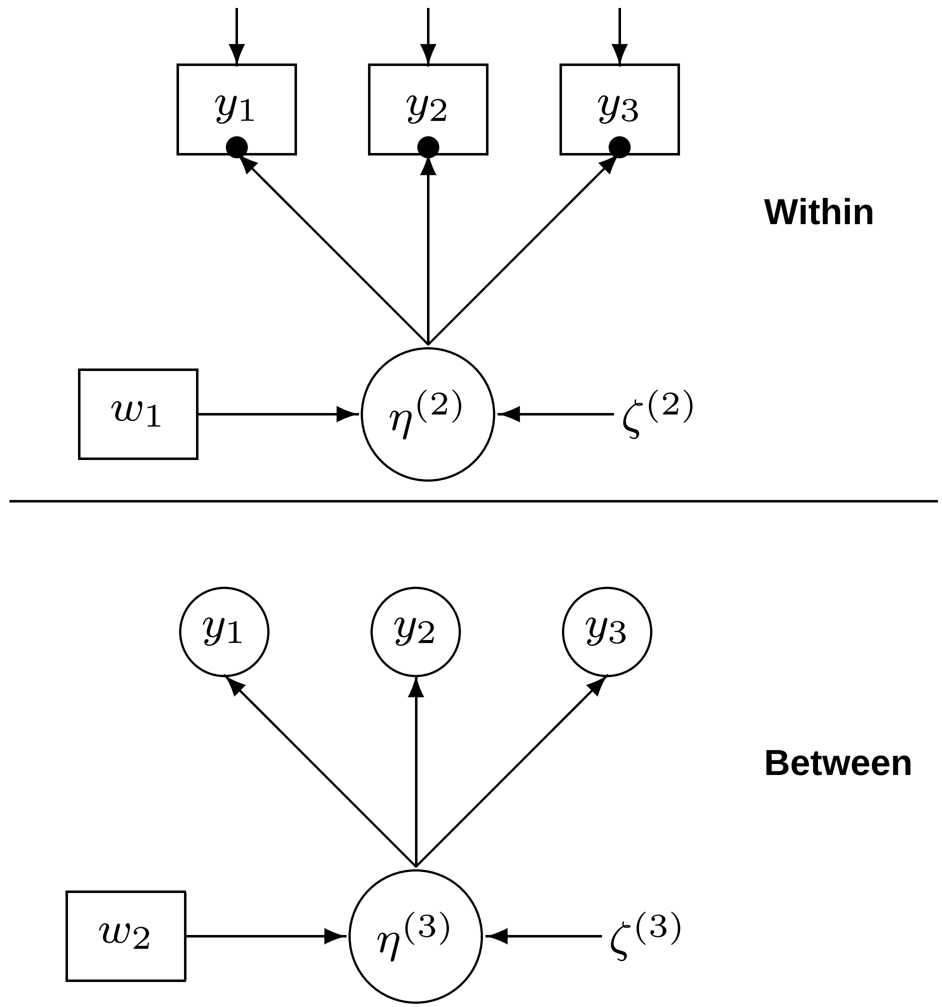
- Unidimensional two-level factor model



[Muthén & Muthén, 2004]

Path diagram for multilevel MIMIC model: within-model and between-model formulation

- Multilevel MIMIC model – must constrain factor loadings



Disadvantage of formulation in terms of within-model and between-model

- No cross-level effects (only indirectly with constraints)
- No higher-level items
- Traditional implementation
  - only for continuous responses
  - only for complete data

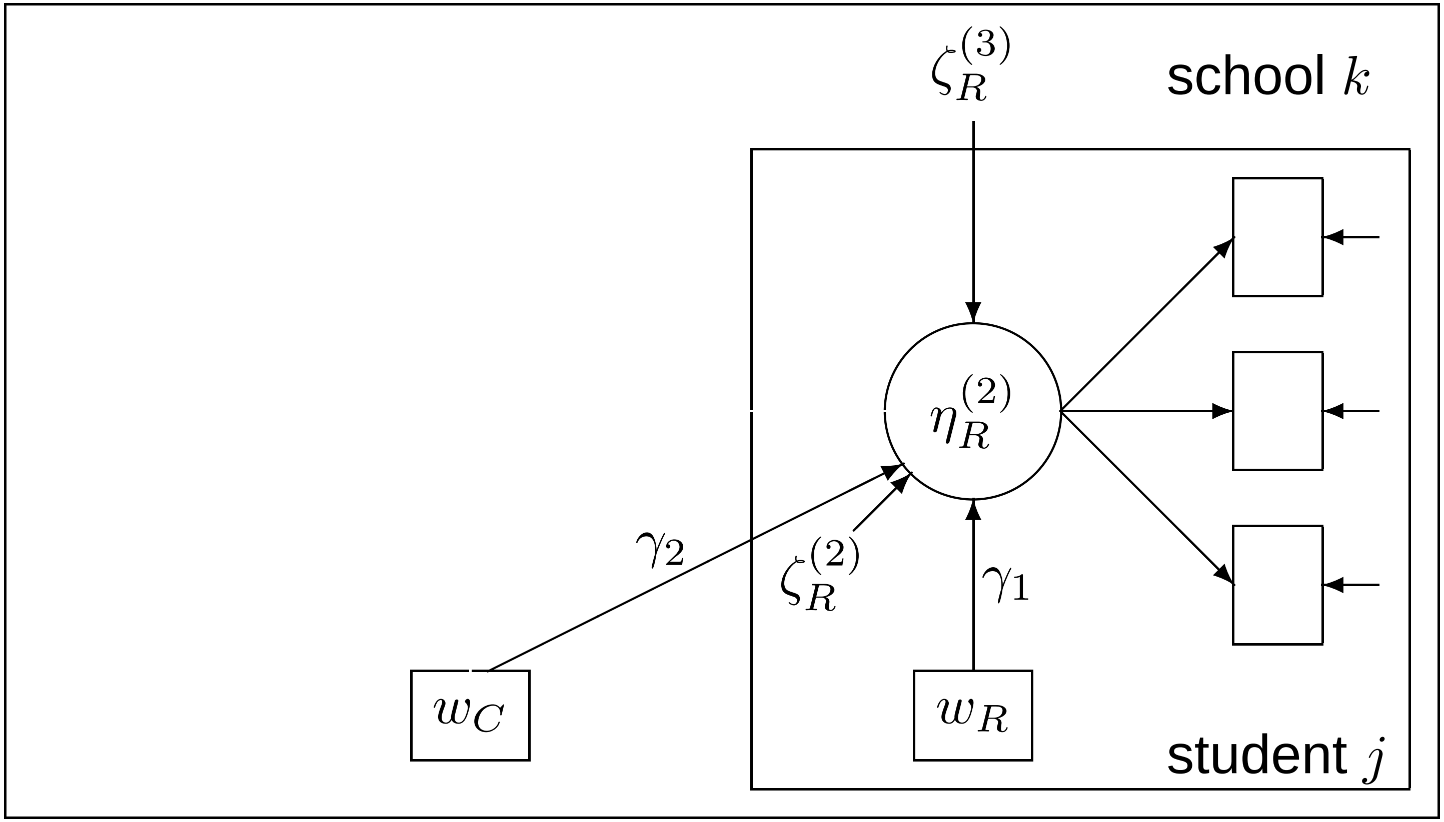
PISA 2000 data: School-level items for school-level latent covariate

- Latent covariate: Teacher excellence
- Responses from school principal: ordinal items with three categories ("satisfied", "somewhat satisfied" and "dissatisfied")
- Questions about teacher excellence:
  1. teacher expectations
  2. student-teacher relations
  3. teacher turnover
  4. teachers meeting individual students' needs
  5. teacher absenteeism
  6. teachers' strictness with students
  7. teacher morale
  8. teachers' enthusiasm
  9. teachers taking pride in the school
  10. teachers valuing academic achievement
- Use cumulative logit (graded-response) model for ordinal items

Multilevel structural equation model with school-level items

- Latent student-level response variable $\eta^{(2)}_{Rjk}$
- Latent school-level covariate $\eta^{(3)}_{Ck}$

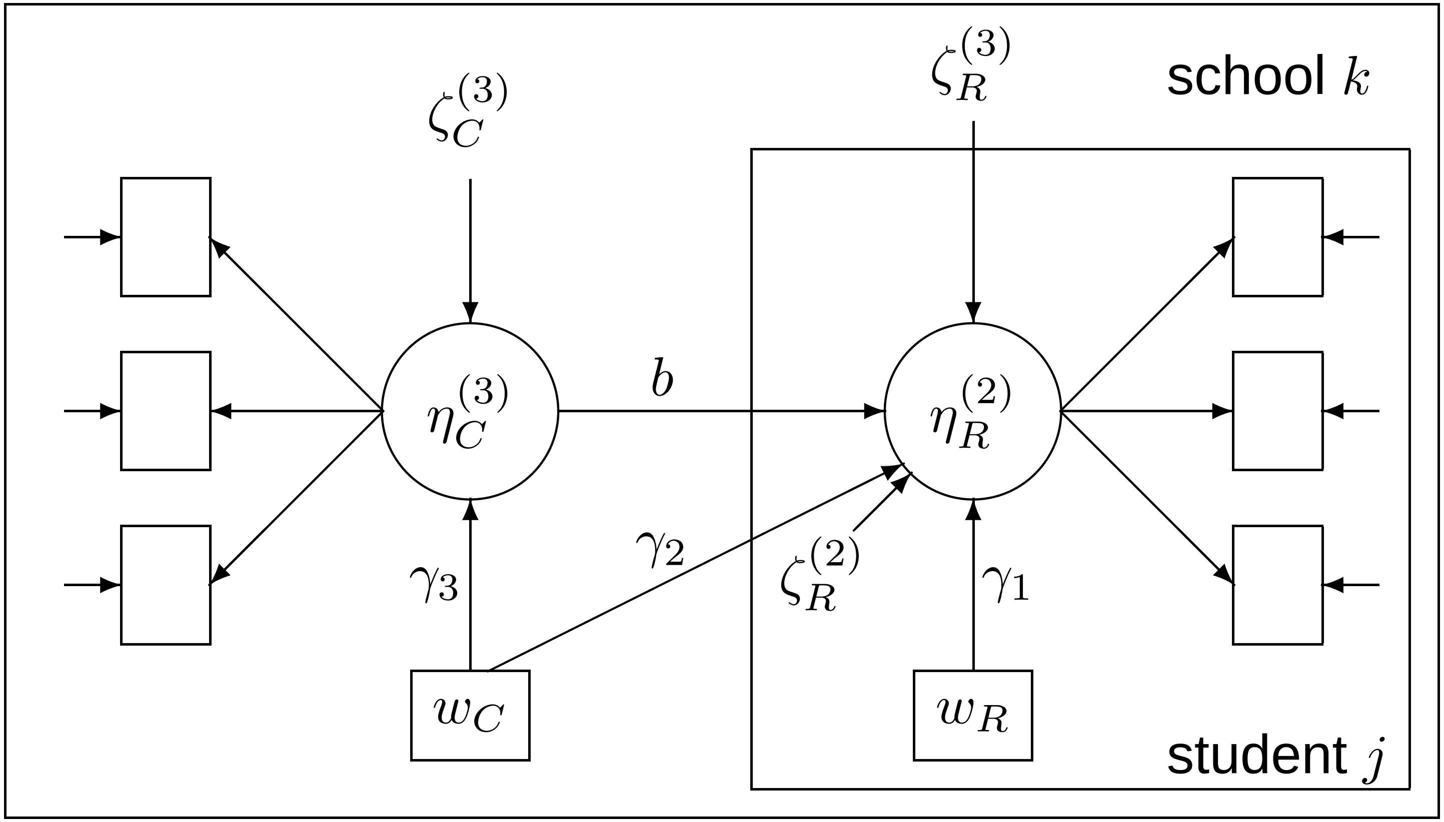
$$\eta^{(2)}_{Rjk} = b\eta^{(3)}_{Ck} + \gamma_1 w_{Rjk} + \gamma_2 w_{Cjk} + \zeta^{(3)}_{Rk} + \zeta^{(2)}_{Rjk}, \qquad \eta^{(3)}_{Ck} = \gamma_3 w_{Ck} + \zeta^{(3)}_{Ck}$$



Multilevel structural equation model with school-level items

- Latent student-level response variable $\eta^{(2)}_{Rjk}$
- Latent school-level covariate $\eta^{(3)}_{Ck}$

$$\eta^{(2)}_{Rjk} = b\eta^{(3)}_{Ck} + \gamma_1 w_{Rjk} + \gamma_2 w_{Cjk} + \zeta^{(3)}_{Rk} + \zeta^{(2)}_{Rjk}, \qquad \eta^{(3)}_{Ck} = \gamma_3 w_{Ck} + \zeta^{(3)}_{Ck}$$



Maximum likelihood estimates of structural model

	MIMIC		SEM		
Parameter	Est	(SE)	Est	(SE_R)	(SE_R^{PSU})
Model for student ability					
b: [Teacher excellence]			0.109	(0.058)	(0.044)
γ_1: [Female]	0.146	(0.122)	0.148	(0.120)	(0.156)
γ_2: [ISEI]	0.012	(0.004)	0.012	(0.004)	(0.004)
γ_3: [Highschool]	0.138	(0.249)	0.133	(0.246)	(0.232)
γ_4: [College]	0.411	(0.263)	0.397	(0.259)	(0.238)
γ_5: [English]	0.555	(0.227)	0.541	(0.224)	(0.222)
γ_6: [MnISEI]	0.039	(0.012)	0.038	(0.012)	(0.014)
$\psi_R^{(2)}$	1.244	(0.642)	1.233	(0.521)	(0.606)
$\psi_R^{(3)}$	0.111	(0.136)	0.083	(0.085)	(0.121)
Model for teacher excellence					
γ_7: [MnISEI]			0.006	(0.020)	(0.019)
$\psi_C^{(3)}$			2.192	(0.427)	(0.432)

103 of the 146 schools had items on teacher excellence

Higher-level items in GLLAMM formulation

- GLLAMM response model

$$\nu = \mathbf{x}'\boldsymbol{\beta} + \sum_{l=2}^{L}\sum_{m=1}^{M_l} \eta_m^{(l)} \mathbf{z}_m^{(l)\prime} \boldsymbol{\lambda}_m^{(l)}$$

  - ν is linear predictor for response at level some level
  - $\mathbf{z}_m^{(2)}$ assigns factor loadings to student-level responses
  - $\mathbf{z}_m^{(3)}$ assigns factor loadings to school-level responses

GLLAMM specification of multilevel SEM:

Response model

$$\underbrace{\begin{bmatrix} \nu_{1jk} \\ \vdots \\ \nu_{7jk} \\ \hline \nu_{1k} \\ \vdots \\ \nu_{10,k} \end{bmatrix}}_{\boldsymbol{\nu}} = \underbrace{\begin{bmatrix} \mathbf{I}_{7\times 7} \\ \hline \mathbf{0}_{10\times 7} \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_7 \end{bmatrix}}_{\boldsymbol{\beta}} + \eta^{(2)}_{Rjk} \underbrace{\begin{bmatrix} \mathbf{I}_{7\times 7} \\ \hline \mathbf{0}_{10\times 7} \end{bmatrix}}_{\mathbf{Z}^{(2)}_{1k}} \underbrace{\begin{bmatrix} 1 \\ \lambda^{(2)}_2 \\ \vdots \\ \lambda^{(2)}_7 \end{bmatrix}}_{\boldsymbol{\lambda}^{(2)}_1} + \eta^{(3)}_{Ck} \underbrace{\begin{bmatrix} \mathbf{0}_{7\times 1} \\ \hline \mathbf{1}_{10\times 1} \end{bmatrix}}_{\mathbf{Z}^{(3)}_{1k}} \underbrace{1}_{\lambda^{(3)}_1}$$

$$+ \eta^{(3)}_{Rk} \underbrace{\begin{bmatrix} \mathbf{0}_{17\times 1} \end{bmatrix}}_{\mathbf{Z}^{(3)}_{2k}} \underbrace{1}_{\lambda^{(3)}_2}$$

& thresholds $\kappa_{i1}, \kappa_{i2},\ i = 1, \ldots, 10$ for school-level items

- $\eta^{(2)}_{Rjk}$: student-level latent variable (interpretation ability)
- $\eta^{(3)}_{Ck}$: school-level latent variable (teacher excellence)
- $\eta^{(3)}_{Rk}$: school-level random intercept for interpretation ability

GLLAMM specification of multilevel SEM:

Structural model

$$\boldsymbol{\eta} \ = \ \boldsymbol{\Gamma}\mathbf{w} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\zeta}$$

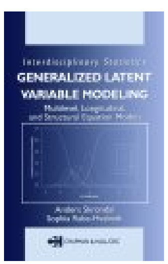
$$\underbrace{\begin{bmatrix} \eta^{(2)}_{Rjk} \\ \eta^{(3)}_{Ck} \\ \eta^{(3)}_{Rk} \end{bmatrix}}_{\boldsymbol{\eta}} = \underbrace{\begin{bmatrix} 0 & b & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} \eta^{(2)}_{Rjk} \\ \eta^{(3)}_{Ck} \\ \eta^{(3)}_{Rk} \end{bmatrix}}_{\boldsymbol{\eta}} + \underbrace{\begin{bmatrix} \gamma_1 & \gamma_2 \\ 0 & \gamma_3 \\ 0 & 0 \end{bmatrix}}_{\boldsymbol{\Gamma}} \underbrace{\begin{bmatrix} w_{Rjk} \\ w_{Ck} \end{bmatrix}}_{\mathbf{w}} + \underbrace{\begin{bmatrix} \zeta^{(2)}_{Rjk} \\ \zeta^{(3)}_{Ck} \\ \zeta^{(3)}_{Rk} \end{bmatrix}}_{\boldsymbol{\zeta}}$$

$$\eta^{(3)}_{Rk} \ = \ \zeta^{(3)}_{Rk}$$

$$\eta^{(3)}_{Ck} \ = \ \gamma_3 w_{Ck} + \zeta^{(3)}_{Ck}$$

$$\eta^{(2)}_{Rjk} \ = \ b\eta^{(3)}_{Ck} + \gamma_1 w_{Rjk} + \gamma_2 w_{Ck} + \underbrace{\zeta^{(3)}_{Rk}}_{\eta^{(3)}_{Rk}} + \zeta^{(2)}_{Rjk}$$

Some of our relevant publications

- Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika* **69**, 167-190.
- Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics* **128**, 301-323.
- Rabe-Hesketh, S. and Skrondal, A. (2006). Multilevel modeling of complex survey data. *Journal of the Royal Statistical Society, Series A* **169**, 805-827.
- Skrondal, A. and Rabe-Hesketh, S. (2003). Multilevel logistic regression for polytomous data and rankings. *Psychometrika* **68**, 267-287.
- 

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman & Hall/CRC.

- For `gllamm` software and more publications, see: `http://www.gllamm.org`