Generalized Linear Latent and Mixed Models

Sophia Rabe-Hesketh

University of California, Berkeley

& Institute of Education, University of London

Australian Statistical Conference
Adelaide, July 2012

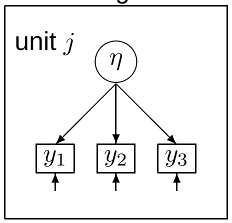
Outline

- Classical latent variable models
  - Factor analysis
  - Item response theory
  - Generalized linear mixed models
  - Latent class models
  - (Skip structural equation models)
- Generalized linear latent and mixed models (GLLAMMs)
- Examples of GLLAMMs
  - Several examples
  - Multilevel structural equation models

Basic idea of latent variable models

- What is a latent variable?
  - Random variable η_j for unit j whose realized values are hidden
  - (Hierarchical Bayes: η_j exchangeable, prior has free hyperparameters)
  - Properties must be indirectly inferred based on a statistical model connecting observed variables $\mathbf{y}_j' = (y_{1j}, y_{2j}, \ldots, y_{n_j j})$ to η_j
- Basic construction principle of latent variable models:
  Conditional independence of observed variables given latent variable

Path diagram



- Pr$(\mathbf{y}_j|\eta_j) = \prod_{i=1}^{n_j}$ Pr$(y_{ij}|\eta_j)$
- Dependence among y_{ij} for unit j induced by η_j
  $\Rightarrow$ Infer properties of η_j from Pr$(\mathbf{y}_j)$

Uses of latent variables

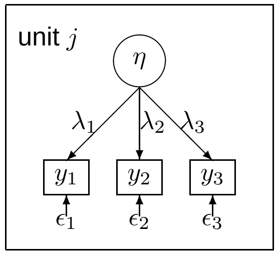
- Hypothetical constructs
- True variables (free from measurement error)
- Unobserved heterogeneity
- Missing data
- Device to induce dependence between different response types

Factor analysis

- Common factor η_j underlies different continuous variables y_{ij}
  - Spearman's (1904) factor g explains correlations between test scores in different subjects (French, English, Math, Music, etc.)
- Unidimensional factor model for ith variable for unit j:

$$y_{ij} = \beta_i + \lambda_i \eta_j + \epsilon_{ij} \quad \text{or} \quad E(y_{ij}|\eta_j) = \nu_{ij} = \beta_i + \lambda_i \eta_j$$

  - β_i is intercept
  - λ_i is factor loading
  - $\eta_j \sim \mathrm{N}(0, \psi)$ is common factor
  - $\epsilon_{ij} \sim \mathrm{N}(0, \theta_i)$ is unique factor
  - For identification, set $\psi = 1$ or $\lambda_1 = 1$



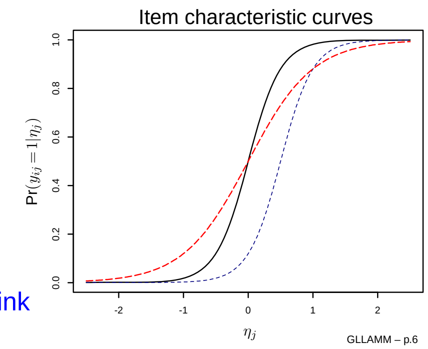
- Measurement model – It is often assumed that $\lambda_i = 1$ and $\theta_i = \theta$

Item response model

- Measurement model for binary data
  - In educational testing, ability η_j explains performance in test items ($y_{ij} = 1$: correct, $y_{ij} = 0$: incorrect)
- Two-parameter logistic (2-PL) model for ith item for person j

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

  - β_i is item intercept
  - λ_i is item discrimination
  - $\eta_j \sim \mathrm{N}(0, \psi)$ is person ability
  - $-\beta_i/\lambda_i$ is item difficulty $\Pr(y_{ij}=1|\eta_j=-\beta_i/\lambda_i) = 0.5$
  - Rasch model: $\lambda_i = 1$



- Common factor model with logit link

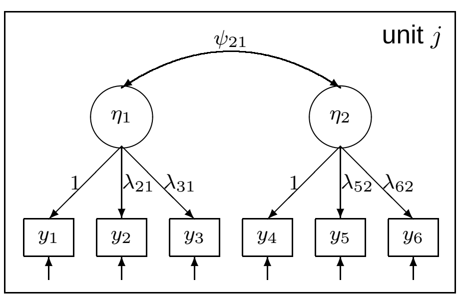
Multidimensional measurement model

- Several latent traits, e.g. verbal and spatial intelligence
- Model for vector of linear predictors $\boldsymbol{\nu}_j$

$$\boldsymbol{\nu}_j \;=\; \boldsymbol{\beta} + \boldsymbol{\Lambda}\boldsymbol{\eta}_j, \quad \boldsymbol{\eta}_j \sim N(\mathbf{0}, \boldsymbol{\Psi})$$

- Ex: Independent clusters (between-item) two-dimensional model

$$\underbrace{\begin{bmatrix} \nu_{1j} \\ \nu_{2j} \\ \nu_{3j} \\ \nu_{4j} \\ \nu_{5j} \\ \nu_{6j} \end{bmatrix}}_{\boldsymbol{\nu}_j} = \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & 1 \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix}}_{\boldsymbol{\Lambda}} \underbrace{\begin{bmatrix} \eta_{1j} \\ \eta_{2j} \end{bmatrix}}_{\boldsymbol{\eta}_j}$$



Generalized linear mixed model (GLMM)

- Regression model for clustered data where latent variables $\boldsymbol{\eta}_j$ induce within-cluster dependence of y_{ij} given covariates
  - Longitudinal: i is occasion or time-point (level 1), j is unit (level 2)
  - Cross-sectional: i is unit (level 1), j is cluster (level 2)
  - Say 'level-1 unit i' and 'level-2 unit j'
- Generalized linear mixed model (GLMM)

$$\boldsymbol{\nu}_j \;=\; \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\boldsymbol{\eta}_j, \quad \boldsymbol{\eta}_j \sim N(\mathbf{0}, \boldsymbol{\Psi})$$

- Ex: $n_j = 3$, $\mathbf{Z}_j = \mathbf{X}_j$ $(n_j \times 2)$

$$\underbrace{\begin{bmatrix} \nu_{1j} \\ \nu_{2j} \\ \nu_{3j} \end{bmatrix}}_{\mathbf{y}_j} = \underbrace{\begin{bmatrix} 1 & x_{1j} \\ 1 & x_{2j} \\ 1 & x_{3j} \end{bmatrix}}_{\mathbf{X}_j} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} 1 & x_{1j} \\ 1 & x_{2j} \\ 1 & x_{3j} \end{bmatrix}}_{\mathbf{Z}_j} \underbrace{\begin{bmatrix} \eta_{1j} \\ \eta_{2j} \end{bmatrix}}_{\boldsymbol{\eta}_j}$$

Similarity of measurement model and GLMM

- Models for $\boldsymbol{\nu}_j$

$$\begin{aligned} \text{Measurement model:} \quad \boldsymbol{\nu}_j &= \mathbf{I}\,\boldsymbol{\beta} + \boldsymbol{\Lambda}\,\boldsymbol{\eta}_j \\ \text{Generalized linear mixed model:} \quad \boldsymbol{\nu}_j &= \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\boldsymbol{\eta}_j \end{aligned}$$

- Measurement model resembles GLMM where:
  - Items correspond to level-1 units and persons to level-2 units
  - Identity matrix $\mathbf{I}$ replaces covariate matrix $\mathbf{X}_j$
  - $\boldsymbol{\Lambda}$ replaces $\mathbf{Z}_j$
    - $\boldsymbol{\Lambda}$ does not vary between persons; elements are parameters
    - $\mathbf{Z}_j$ varies between level-2 units; elements are fixed constants
- Will see that GLLAMMs unify measurement models and GLMMs

Exploratory latent class model

- Unit j falls into group (latent class) c with probability π_c $(c = 1, \ldots, C)$
- Each latent class characterized by different set of response probabilities for binary variables y_{ij} $(i = 1, \ldots, n)$
  - In medical diagnosis, latent classes are diseases and variables are diagnostic test results
- Exploratory latent class model
  - n-dimensional discrete latent variable $\boldsymbol{\eta}_j$ discrete values $\{\mathbf{e}_1, \ldots, \mathbf{e}_C\}$ with probabilities $\{\pi_1, \ldots, \pi_C\}$
  - Conditional response probabilities for class c:

$$\begin{aligned} \text{logit}[\Pr(y_{ij}=1|\boldsymbol{\eta}_j = \mathbf{e}_c)] &= e_{ic} \\ \text{logit}[\Pr(y_{ij}=1|\boldsymbol{\eta}_j)] &= \eta_{ij} \end{aligned}$$

- GLMM with discrete random effects:

$$\boldsymbol{\nu}_j = \mathbf{I}\boldsymbol{\eta}_j$$

GLLAMM response model: Unifying latent variable models

- Level-1 units i nested in level-2 units j, etc. up to level L

$$\boldsymbol{\nu} = \mathbf{X}\boldsymbol{\beta} + \sum_{l=2}^{L}\sum_{m=1}^{M_l} \eta_m^{(l)}\mathbf{Z}_m^{(l)}\boldsymbol{\lambda}_m^{(l)}$$

  - Measurement model:

$$\underbrace{\begin{bmatrix}\nu_{1j}\\ \nu_{2j}\\ \nu_{3j}\end{bmatrix}}_{\boldsymbol{\nu}_j} = \underbrace{\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}}_{\mathbf{X}_j}\underbrace{\begin{bmatrix}\beta_1\\ \beta_2\\ \beta_3\end{bmatrix}}_{\boldsymbol{\beta}} + \eta_{1j}^{(2)}\underbrace{\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}}_{\mathbf{Z}_{1j}^{(2)}}\underbrace{\begin{bmatrix}\lambda_1\\ \lambda_2\\ \lambda_3\end{bmatrix}}_{\boldsymbol{\lambda}_1^{(2)}} = \begin{bmatrix}\beta_1+\eta_{1j}^{(2)}\lambda_1\\ \beta_2+\eta_{1j}^{(2)}\lambda_2\\ \beta_3+\eta_{1j}^{(2)}\lambda_3\end{bmatrix}$$

  - Generalized linear mixed model:

$$\underbrace{\begin{bmatrix}\nu_{1j}\\ \nu_{2j}\\ \nu_{3j}\end{bmatrix}}_{\boldsymbol{\nu}_j} = \underbrace{\begin{bmatrix}1&x_{1j}\\1&x_{2j}\\1&x_{3j}\end{bmatrix}}_{\mathbf{X}_j}\underbrace{\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix}}_{\boldsymbol{\beta}} + \eta_{1j}^{(2)}\underbrace{\begin{bmatrix}1\\1\\1\end{bmatrix}}_{\mathbf{Z}_{1j}^{(2)}} + \eta_{2j}^{(2)}\underbrace{\begin{bmatrix}x_{1j}\\ x_{2j}\\ x_{3j}\end{bmatrix}}_{\mathbf{Z}_{2j}^{(2)}} = \begin{bmatrix}\beta_1+\eta_{1j}^{(2)}+(\beta_2+\eta_{2j}^{(2)})x_{1j}\\ \beta_1+\eta_{1j}^{(2)}+(\beta_2+\eta_{2j}^{(2)})x_{2j}\\ \beta_1+\eta_{1j}^{(2)}+(\beta_2+\eta_{2j}^{(2)})x_{3j}\end{bmatrix}$$

Response types in GLLAMM

Ordinary GLMs:

Links
identity
reciprocal
logarithm
logit
probit
compl. log-log

Families
Gaussian
gamma
Poisson
binomial

Extensions:

Ordinal responses
ordinal logit
ordinal probit
ordinal compl. log-log
scaled ordinal probit

Nominal & Rankings
multinomial logit

- Mixed responses: different links and families for different units
- Heteroscedasticity: Variance or dispersion parameter θ can be modelled as $\log\theta = \mathbf{z}^{(1)\prime}\boldsymbol{\alpha}$

Structural model for continuous latent variables

- Regressions of latent variables on latent and observed variables at the same or higher levels

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

- $\boldsymbol{\eta} = (\overbrace{\eta_1^{(2)}, \eta_2^{(2)}, \ldots, \eta_{M_2}^{(2)}}^{\text{Level 2}}, \ldots, \overbrace{\eta_1^{(l)}, \eta_2^{(l)}, \ldots, \eta_{M_l}^{(l)}}^{\text{Level } l}, \ldots, \overbrace{\eta_1^{(L)}, \eta_2^{(L)}, \ldots, \eta_{M_L}^{(L)}}^{\text{Level } L})'$
  - $M = \sum_l M_l$ latent variables:
- $\mathbf{B}$ is an upper triangular $M \times M$ matrix of regression coefficients
- $\boldsymbol{\Gamma}$ is an $M \times p$ matrix of regression coefficients
- $\mathbf{w}$ is a p dimensional vector of explanatory variables
- $\boldsymbol{\zeta}$ is an M dimensional vector of disturbances

$\boldsymbol{\zeta} = (\overbrace{\zeta_1^{(2)}, \zeta_2^{(2)}, \ldots, \zeta_{M_2}^{(2)}}, \ldots, \overbrace{\zeta_1^{(l)}, \zeta_2^{(l)}, \ldots, \zeta_{M_l}^{(l)}}, \ldots, \overbrace{\zeta_1^{(L)}, \zeta_2^{(L)}, \ldots, \zeta_{M_L}^{(L)}})'$

$\boldsymbol{\zeta}^{(l)} \sim N(\mathbf{0}, \boldsymbol{\Psi}^{(l)})$, independent across levels

Structural model for discrete latent variables

- Probability that unit j belongs to class $c = 1, \ldots, C$ may depend on covariates $\mathbf{v}_j$ through a multinomial logit model

$$\pi_{jc} \;=\; \frac{\exp(\mathbf{v}_j'\boldsymbol{\varrho}^c)}{\sum_d \exp(\mathbf{v}_j'\boldsymbol{\varrho}^d)},$$

where $\boldsymbol{\varrho}^c$ are parameters with $\boldsymbol{\varrho}^C = \mathbf{0}$ for identification

- Nonparametric maximum likelihood (NPML) estimator if number of classes increased to maximize likelihood

Estimation and prediction in Stata program `gllamm`

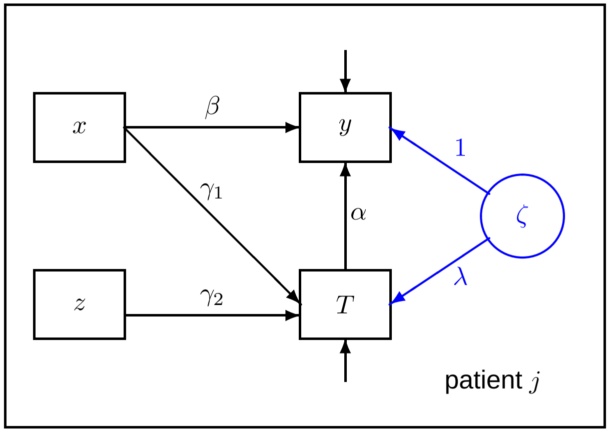
- To obtain the likelihood of GLLAMM's, the latent variables must be integrated out
  - Sequentially integrate over latent variables, starting with the lowest level using a recursive algorithm
  - Use Gauss-Hermite quadrature to replace integrals by sums
  - Scale and translate quadrature locations to match the peak of the integrand using **adaptive quadrature**
- Maximum likelihood estimates obtained using Newton-Raphson
- Model-based and robust standard errors (sandwich estimator)
- Empirical Bayes (EB) predictions of latent variables and EB standard errors obtained as byproducts of adaptive quadrature

Examples of GLLAMMs

- Mixed response types
  - Ex 1: Random coefficient dependent dropout
  - Ex 2: Endogenous treatment model
- Continuous latent variables regressed on observed and latent var.
  - Ex 1: Item response model with differential item functioning (DIF)
  - Ex 2: Covariate measurement error model
  - Ex 3: Missing covariate model
- Discrete latent variables regressed on observed variables
  - Ex 1: Complier average causal effect (CACE) models
- Multilevel latent variables
  - Ex 1: Multilevel structural equation models (SEMs)

continue

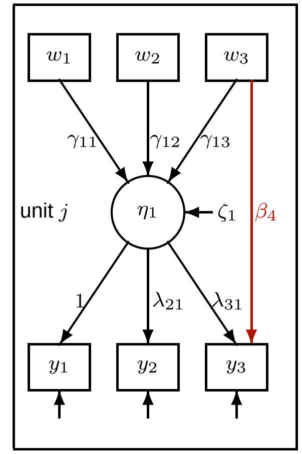
Endogenous treatment model



Treatment endogenous

back

Item response model with DIF

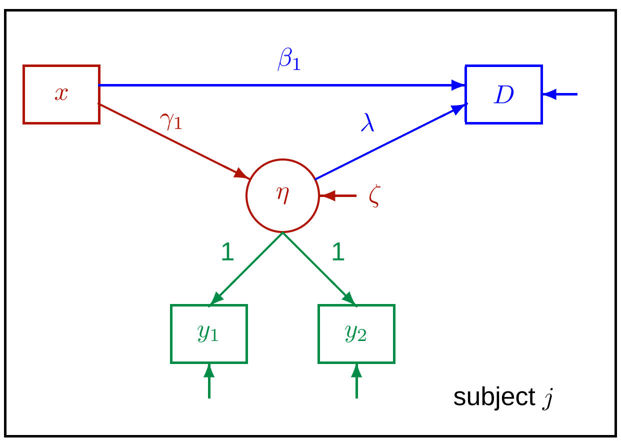


MIMIC model with direct effect

back

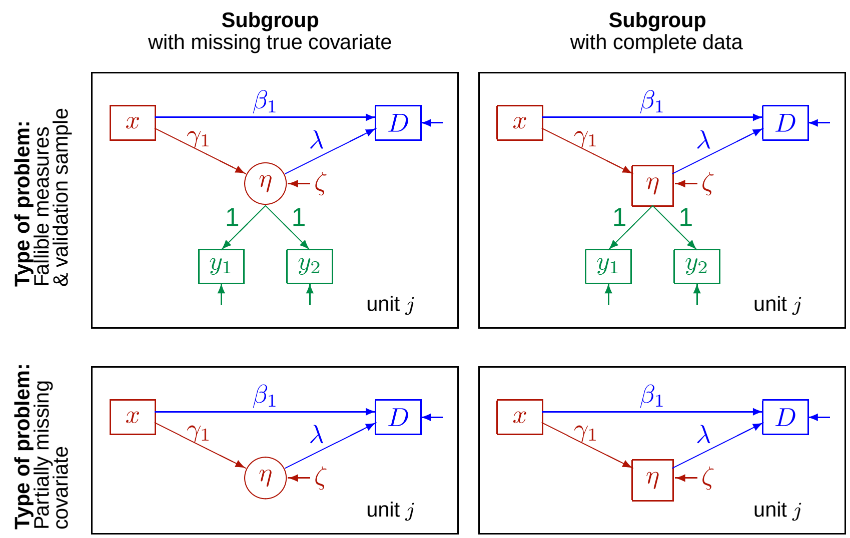
Covariate measurement error

- Exposure model + Measurement model + Disease model



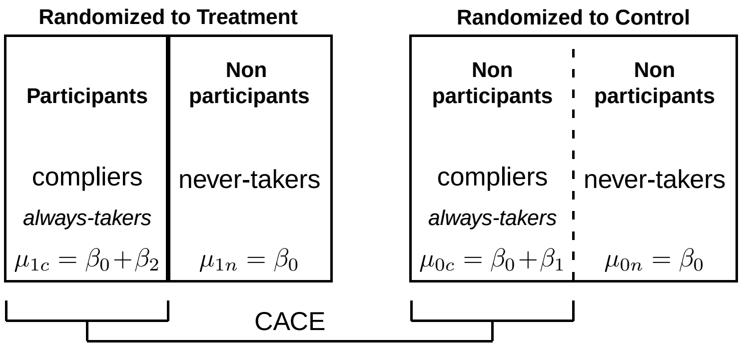
back

Missing covariate model

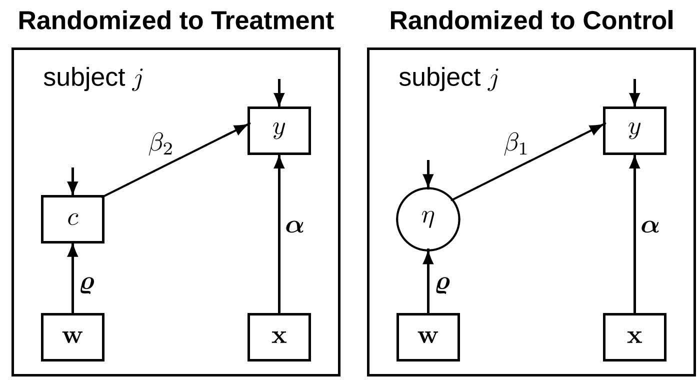


back

Complier average causal model



Complier average causal model (cont'd)



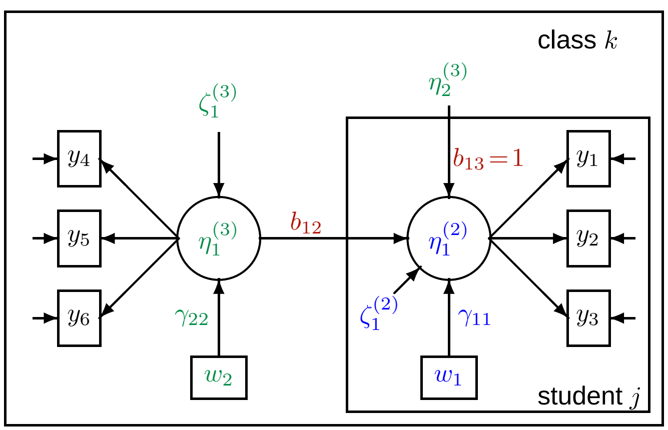
- c is dummy variable for complier (or always taker), versus never-taker
- $c = \eta$ is latent in control group (no access to treatment)
- Complier average treatment effect: $\beta_2 - \beta_1$

back

Multilevel SEM

- Structural model

$$\begin{bmatrix}\eta^{(2)}_{1jk}\\ \eta^{(3)}_{1k}\\ \eta^{(3)}_{2k}\end{bmatrix} = \begin{bmatrix}0 & b_{12} & 1\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\eta^{(2)}_{1jk}\\ \eta^{(3)}_{1k}\\ \eta^{(3)}_{2k}\end{bmatrix} + \begin{bmatrix}\gamma_{11} & 0\\ 0 & \gamma_{22}\\ 0 & 0\end{bmatrix}\begin{bmatrix}w_{1jk}\\ w_{2k}\end{bmatrix} + \begin{bmatrix}\zeta^{(2)}_{1jk}\\ \zeta^{(3)}_{1k}\\ \zeta^{(3)}_{2k}\end{bmatrix}$$



Multilevel SEM (cont'd)

- Response model (measurement model)

$$\underbrace{\begin{bmatrix}\nu_{1jk}\\ \nu_{2jk}\\ \nu_{3jk}\\ \nu_{4k}\\ \nu_{5k}\\ \nu_{6k}\end{bmatrix}}_{\boldsymbol{\nu}_k} = \underbrace{\mathbf{I}\boldsymbol{\beta}}_{\mathbf{X}_k\boldsymbol{\beta}} + \eta^{(2)}_{1jk}\underbrace{\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\\ 0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}}_{\mathbf{Z}^{(2)}_{1k}}\underbrace{\begin{bmatrix}1\\ \lambda^{(2)}_2\\ \lambda^{(2)}_3\end{bmatrix}}_{\boldsymbol{\lambda}^{(2)}_1} + \eta^{(3)}_{1k}\underbrace{\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\\ 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}}_{\mathbf{Z}^{(3)}_{1k}}\underbrace{\begin{bmatrix}1\\ \lambda^{(3)}_2\\ \lambda^{(3)}_3\end{bmatrix}}_{\boldsymbol{\lambda}^{(3)}_1} + \eta^{(3)}_{2k}\underbrace{\begin{bmatrix}0\\ 0\\ 0\\ 0\\ 0\\ 0\end{bmatrix}}_{\mathbf{Z}^{(3)}_{2k}}\underbrace{1}_{\boldsymbol{\lambda}^{(3)}_2}$$

- Column of zeros is necessary because $\eta^{(3)}_{2k}$ has no measures

back

Concluding remarks

- Many more possible models in GLLAMM framework
- Advantage of modeling framework
  - Helps recognize commonalities between models
  - One program to estimate all models
  - Facilitates formulation and estimation of new models
- Features not covered in GLLAMM framework
  - Interactions among latent variables
  - Models with both discrete and continuous latent variables
- Problems with complex models
  - Difficult to establish identification and equivalence
  - Estimation can take a long time
  - Models may look naive – but less naive than simpler (constrained) versions

Some references related to GLLAMM

- Generalized Linear Latent and Mixed Modeling (GLLAMM) Framework:
  - Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2004). Generalized multilevel structural equation modelling. *Psychometrika* **69** (2), 167–190.
  - Skrondal, A. & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Chapman & Hall/ CRC.
- Estimation and prediction:
  - Rabe-Hesketh, S., Skrondal, A. & 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. & Skrondal, A. (2006). Multilevel modelling of complex survey data. *Journal of the Royal Statistical Society, Series A* **169**, 805-827.
  - Skrondal, A. & Rabe-Hesketh (2009). Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society, Series A* **172**, 659-687.
- gllamm Software:

  - Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2004). *GLLAMM Manual*. U.C. Berkeley Division of Biostatistics Working Paper Series 160.
  - Rabe-Hesketh, S. & Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata (3rd Ed.)*. Stata Press.
- Check out `http://www.gllamm.org` for more information!