Slide 1

Multilevel item response and structural equation modeling

Sophia Rabe-Hesketh
Department of Biostatistics and Computing
Institute of Psychiatry, London

joint work with
Anders Skrondal (Norwegian Institute of Public Health, Oslo)
and Andrew Pickles (The University of Manchester)

Seminar
Berkeley Evaluation & Assessment Research Center (BEAR)
University of California, Berkeley
1 October 2002

Slide 2

Outline

- Extending item response models
  - Two parameter logistic item response model
  - Including observed covariates
  - Multidimensional item response model
  - Multilevel item response model
  - Including latent covariates
- Generalized Linear Latent And Mixed Models (GLLAMM)
  - Response model
  - Structural model
- Estimation, prediction and simulation using Stata programs `gllamm`, `gllapred`, `gllasim`
- Application: Attitudes to Abortion

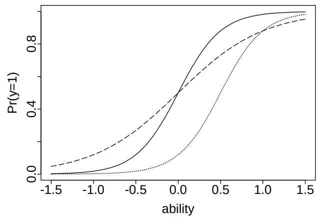
Two parameter logistic item response model

- Dichotomous response y_{ij} for item i and person j
- Two parameter logistic (2PL) model:

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

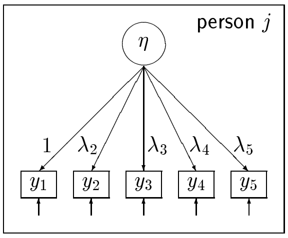
$$\lambda_1 = 1$$

$$\eta_j \sim N(0, \tau^2)$$



Slide 3

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



Including covariates

- Predictors of ability: w_j affects responses indirectly via η_j (MIMIC model)

$$\text{logit}[\Pr(y_{ij}=1|\eta_j)] = \beta_i + \eta_j\lambda_i, \quad \lambda_1 = 1$$

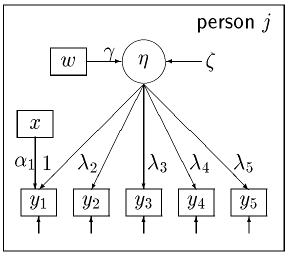
$$\eta_j = \gamma w_j + \zeta_j$$

$$\zeta_j \sim N(0, \tau^2)$$

Slide 4

- Differential item functioning: x_j affects responses directly

$$\text{logit}[\Pr(y_{ij}=1|\eta_j, x_j)] = \beta_i + \eta_j\lambda_i + \alpha_i x_j, \quad \lambda_1 = 1$$



Multidimensional item response model

- Different types of ability (e.g. verbal, quantitative):

$$\text{logit}[\Pr(y_{ij}=1|\boldsymbol{\eta}_j)] = \beta_i + \eta_{1j}\lambda_{i1} + \eta_{2j}\lambda_{i2} + \eta_{3j}\lambda_{i3},$$

$$\lambda_{11}=\lambda_{42}=\lambda_{73}=1$$

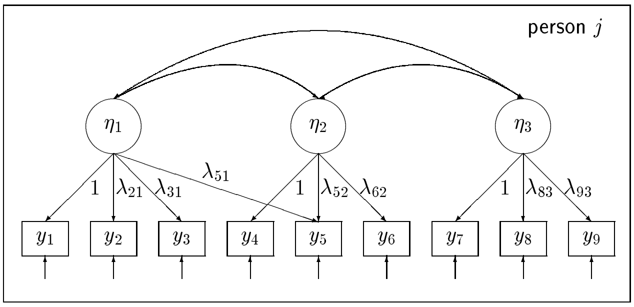
- Confirmatory model:

$$\lambda_{41}=\lambda_{61}=\lambda_{71}=\lambda_{81}=\lambda_{91}=0$$
$$\lambda_{12}=\lambda_{22}=\lambda_{32}=\lambda_{72}=\lambda_{82}=\lambda_{92}=0$$
$$\lambda_{13}=\lambda_{23}=\lambda_{33}=\lambda_{43}=\lambda_{53}=\lambda_{63}=0$$

- Correlated abilities:

Slide 5

$$\begin{bmatrix} \eta_{1j} \\ \eta_{2j} \\ \eta_{3j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_1^2 & & \\ \tau_{12} & \tau_2^2 & \\ \tau_{13} & \tau_{23} & \tau_3^2 \end{bmatrix} \right)$$

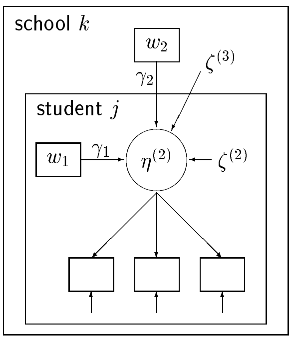


Multilevel item response model

- Students j (level 2) nested in schools k (level 3)
  (items i are at level 1)
- Predictors of ability: observed heterogeneity
  - Student-level predictor w_{1jk}, e.g. socio-economic status
  - School-level predictor w_{2k}, e.g. student-teacher ratio
- Omitted predictors: unobserved heterogeneity
  - Student-level residual $\zeta_{jk}^{(2)}$
  - School-level residual $\zeta_k^{(3)}$

Slide 6

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



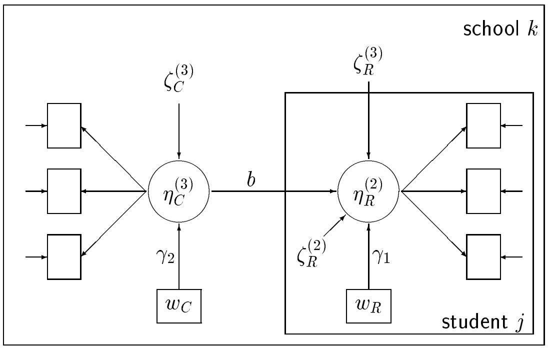
Slide 7

Including latent covariates

- Multilevel structural equation model
- Student ability $\eta^{(2)}_{Rjk}$
- Latent covariate at the school level $\eta^{(3)}_{Ck}$
  e.g. principal's attitude

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

$$\eta^{(3)}_{ck} = \gamma_2 w_{Ck} + \zeta^{(3)}_{Ck}$$



Slide 8

GLLAMM: Response model

The response model is a generalized linear model conditional on the latent variables

- 'Linear' predictor ν
  - for the 2PL model:
    $\nu_{ij} \;=\; \beta_i + \eta_j\lambda_i$
- Link function $g(\cdot)$, $\quad g(\mathrm{E}[y|\nu]) = \nu$
  - for the 2PL model:
    $\mathrm{logit}(\mathrm{E}[y_{ij}|\nu_{ij}]) \;=\; \nu_{ij}, \quad \mathrm{E}[y_{ij}|\nu_{ij}] \;=\; \Pr(y_{ij}=1|\nu_{ij})$
- Conditional distribution $f(y|\nu)$ from the exponential family
  - for the 2PL model:
    independent Bernouilli conditional on latent variable (ability)
    $\Longrightarrow$ conditional independence (or local independence)

Slide 9

GLLAMM: Form of 'linear' predictor

- Multilevel data with L levels:
  units i (level 1) nested in clusters j (level 2), etc. up to level L
  - for 2PL model, $L = 2$:
    items i (level 1) nested in persons j (level 2)
- M_l latent variables at level $l = 1, \cdots, L$
  - for 2PL model, $M_2 = 1$
- Model for vector of linear predictors for level 2 unit j

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

  - for 2PL model:

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

Slide 10

Further examples of 'linear' predictor

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

- Differential item functioning

$$\underbrace{\begin{bmatrix} \nu_{1j} \\ \nu_{2j} \\ \nu_{3j} \end{bmatrix}}_{\boldsymbol{\nu}_j} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & x_j \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\mathbf{X}_j} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \alpha_1 \end{bmatrix}}_{\boldsymbol{\beta}} + \eta_{1j}^{(2)} \mathbf{Z}_{1j}^{(2)} \boldsymbol{\lambda}_1^{(2)} = \begin{bmatrix} \beta_1 + \eta_{1j}^{(2)}\lambda_1 + \alpha_1 x_j \\ \beta_2 + \eta_{1j}^{(2)}\lambda_2 \\ \beta_3 + \eta_{1j}^{(2)}\lambda_3 \end{bmatrix}$$

- Response model for multilevel structural equation model

$$\underbrace{\begin{bmatrix} \nu_{R1j} \\ \nu_{R2j} \\ \nu_{R3j} \\ \nu_{C1j} \\ \nu_{C2j} \\ \nu_{C3j} \end{bmatrix}}_{\boldsymbol{\nu}_j} = \mathbf{X}_j\boldsymbol{\beta} + \eta_{Rjk}^{(2)} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\mathbf{Z}_{1j}^{(2)}} \underbrace{\begin{bmatrix} \lambda_{R1} \\ \lambda_{R2} \\ \lambda_{R3} \end{bmatrix}}_{\boldsymbol{\lambda}_1^{(2)}} + \eta_{Ck}^{(3)} \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{Z}_{1j}^{(3)}} \underbrace{\begin{bmatrix} \lambda_{C1} \\ \lambda_{C2} \\ \lambda_{C3} \end{bmatrix}}_{\boldsymbol{\lambda}_1^{(3)}}$$

Slide 11

Response processes

- Continuous
- Dichotomous
- Ordinal
  - Graded response model (cumulative) including models for thresholds or scale parameter
  - Partial credit model (adjacent odds)
  - Continuation ratio model
- Unordered categorical and rankings
- Counts
- Durations in continuous time
  - Proportional hazards model
  - Accelerated failure time model
- Durations in discrete time
  - Censored cumulative models
  - Continuation ratio model
  - Proportional hazards in continuous time
- Mixed responses

Slide 12

Multinomial logit for categorical responses

- Categorical responses with S_i response alternatives for item i.
- If person j responds to item i with response s, $y_{sij} = 1$, $y_{tij} = 0$ for all $t \neq s$.
  - Example $S_1 = 3$, $S_2 = 2$:

$$\begin{bmatrix} y_{11j} \\ y_{21j} \\ y_{31j} \\ y_{12j} \\ y_{22j} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

- Specify GLLAMM model for vector of linear predictors

$$\boldsymbol{\nu}_j = (\nu_{1ij}, \cdots \nu_{S_1,ij}, \cdots, \nu_{1Ij} \cdots \nu_{S_I,Ij})'$$

- Response probability

$$\Pr(y_{sij} = 1 | \boldsymbol{\nu}_{ij}) = \frac{\exp(\nu_{sij})}{\sum_{t=1}^{S_i} \exp(\nu_{tij})}$$

- This seems to correspond to the Multidimensional Random Coefficient Multinomial Logit (MRCML) model (Adams, Wilson & Wang, 1997)

Slide 13

GLLAMM: Structural model

Latent variables can be regressed on other latent and observed variables

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

- $\boldsymbol{\eta}$ is an M-dimensional vector of latent variables

$$\boldsymbol{\eta} = (\eta_1^{(2)}, \eta_2^{(2)}, \cdots, \eta_{M_2}^{(2)}, \cdots, \eta_1^{(l)}, \cdots, \eta_{M_l}^{(l)}, \cdots, \eta_{M_L}^{(L)})'$$

- $\boldsymbol{\Gamma}$ is an $M \times p$ matrix of regression coefficients
- $\mathbf{w}$ is a p dimensional vector of explanatory variables
- $\mathbf{B}$ is an upper diagonal $M \times M$ matrix of regression coefficients
  - cannot regress higher level latent variable on lower level latent variable
  - relations among latent variables are recursive
- $\boldsymbol{\zeta}$ is an M-dimensional vector of errors/disturbances (each element $\boldsymbol{\zeta}$ varies at the same level as corresponding elements in $\boldsymbol{\eta}$).

Slide 14

Examples of structural model

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

- Multilevel item response model

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

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

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

- Multilevel structural equation model

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

$$\eta_{Rk}^{(3)} = \zeta_{Rk}^{(3)}$$

$$\eta_{Ck}^{(3)} = \gamma_2 w_{Ck} + \zeta_{Ck}^{(3)}$$

$$\eta_{Rjk}^{(2)} = \gamma_1 w_{Rjk} + b\eta_{Ck}^{(3)} + \underbrace{\zeta_{Rk}^{(3)}}_{\eta_{Rk}^{(3)}} + \zeta_{Rjk}^{(2)}$$

Slide 15

Estimation using the 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 multivariate integrals by nested sums
  - Improve approximation using adaptive quadrature which scales and translates quadrature locations to match the peak of the integrand
- Maximize likelihood using Newton-Raphson

Slide 16

Adaptive quadrature: unidimensional case

- Ordinary quadrature:

$$\begin{aligned}\ell_j(\boldsymbol{\beta},\boldsymbol{\lambda},\tau) &= \int \phi(\eta_j;0,\tau^2)\prod_i f(y_{ij}|\eta_j;\boldsymbol{\beta},\boldsymbol{\lambda})\mathrm{d}\eta_j,\ [u_j=\eta_j/\tau] \\ &= \int \phi(u_j)\prod_i f(y_{ij}|\tau u_j)\mathrm{d}u_j \\ &\approx \sum_{r=1}^{R} W_r \prod_i f(y_{ij}|\tau A_r)\end{aligned}$$

  - W_r and A_r: quadrature weights and locations
- Adaptive quadrature:

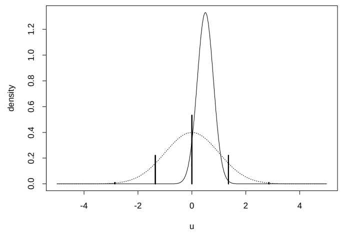
$$\begin{aligned}\ell_j(\boldsymbol{\beta},\boldsymbol{\lambda},\tau) &= \int \phi(u_j;\mu_j,\sigma_j^2)\left[\frac{\phi(u_j)\prod_i f(y_{ij}|\tau u_j)}{\phi(u_j;\mu_j,\sigma_j^2)}\right]\mathrm{d}u_j,\ [v_j=\frac{u_j-\mu_j}{\sigma_j}] \\ &= \int \frac{\phi(v_j)}{\sigma_j}\left[\frac{\phi(\sigma_j v_j+\mu_j)\prod_i f(y_{ij}|\tau(\sigma_j v_j+\mu_j))}{\frac{1}{\sqrt{2\pi}\sigma_j}\exp(-v_j^2/2)}\right]\sigma_j\mathrm{d}v_j \\ &\approx \sum_{r=1}^{R} W_r\left[\frac{\phi(\sigma_j A_r+\mu_j)\prod_i f(y_{ij}|\tau(\sigma_j A_r+\mu_j))}{\frac{1}{\sqrt{2\pi}\sigma_j}\exp(-A_r^2/2)}\right] \\ &= \sum_{r=1}^{R} \omega_{jr}\prod_i f(y_{ij}|\tau\alpha_{jr})\end{aligned}$$

  - $\phi(u_j;\mu_j,\sigma_j^2)$: normal density approximating posterior (approximately proportional to the numerator in [ ])
  - $\alpha_{jr}=\sigma_j A_r+\mu_j$
  - $\omega_{jr}=\sqrt{2\pi}\sigma_j\exp(A_r^2/2)\phi(\sigma_j A_r+\mu_j)W_r$

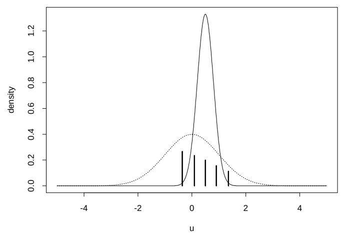
Slide 17

Adpative quadrature places quadrature locations under peak of integrand

Quadrature



Adaptive quadrature



Prior (dotted curve) and posterior (solid curve) densities

(Integrand is proportional to posterior)

Slide 18

Prediction using `gllapred`

- Use empirical Bayes (EB) prediction to derive scores
  - Posterior mean for person j

$$\mathrm{E}[\eta_j|\mathbf{y}_j] = \frac{\int \eta_j \phi(\eta_j; 0, \hat{\tau}^2) \prod_i f(y_{ij}|\eta_j; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\lambda}}) \mathrm{d}\eta_j}{\ell_j(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\lambda}}, \hat{\tau})}$$

  - Posterior variance for person j

$$\mathrm{var}[\eta_j|\mathbf{y}_j] = \frac{\int \eta_j^2 \phi(\eta_j; 0, \hat{\tau}^2) \prod_i f(y_{ij}|\eta_j; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\lambda}}) \mathrm{d}\eta_j}{\ell_j(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\lambda}}, \hat{\tau})} - \mathrm{E}[\eta_j|\mathbf{y}_j]^2$$

  - The integrals are evaluated using adaptive quadrature

Simulation using `gllasim`

Simulate the abilities and responses for a given model

- Model diagnostics
  - Derive sampling distribution of measures of lack of fit
  - Compare diagnostic plot for real data with equivalent plots for simulated data
- Model interpretation
- Investigate (and correct) bias of estimators
- Investigate effects of model misspecification
- Power calculations

Slide 19

Example syntax for 2PL

- Data:

person	item	i1	i2	i2	y
1	1	1	0	0	1
1	2	0	1	0	1
1	3	0	0	1	0
2	1	1	0	0	1
2	2	0	1	0	0
2	3	0	0	1	1

- Estimate model:

```
eq discrim: i1 i2 i3
gllamm y i1 i2 i3, i(person) eqs(discrim) /*
       */ link(logit) family(binom) adapt
```

- Obtain posterior means and standard deviations:

```
gllapred score, u
```

- Simulate abililties and responses:

```
gllasim abil, u
gllasim response
```

Slide 20

Application: Attitudes to Abortion

- British Social Attitudes Survey Panel 1983-1986
- Respondents were asked whether or not abortion should be allowed by law if:

[wom] The woman decides on her own she does not wish to have the child

[cou] The couple agree that they do not wish to have the child

[mar] The woman is not married and does not wish to marry the man

[fin] The couple cannot afford any more children

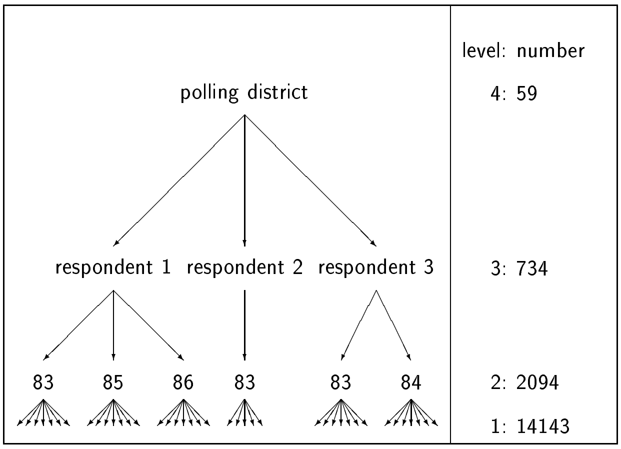
[def] There is a strong chance of a defect in the baby

[ris] The woman's health is seriously endangered by the pregnancy

[rap] The woman became pregnant as a result of rape

Slide 21

Data structure



- Unit non-response was common:

Number of panel waves participated in:	4	3	2	1
Percentage of respondents:	49	12	13	25

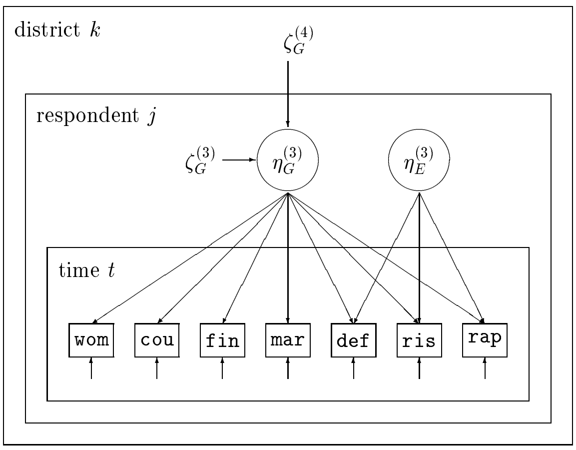
- Item non-response occurrred in only 7% of interviews

Slide 22

Models

- Only general attitude factor $\eta_G^{(3)}$
- Add extreme circumstance factor $\eta_E^{(3)}$, $X^2(3) = 207.7$
- (Add unique factors for respondents, $X^2(7) = 12.6$)
- Add district-level residual $\zeta_G^{(4)}$ for general attitude factor, $X^2(1) = 8.2$
- (Add district-level residual $\zeta_E^{(4)}$ for extreme circumstance factor, $X^2(1) = 3.2$)

Final Model



Slide 23

FIXED PART

Intercepts:		
[wom]	-0.83 (0.14)	
[cou]	-0.17 (0.15)	
[mar]	-0.28 (0.16)	
[fin]	-0.01 (0.14)	
[def]	3.79 (0.27)	
[ris]	5.90 (0.56)	
[rap]	4.82 (0.39)	
RANDOM PART: Respondent level		
Factor loadings:	General	Extreme
[wom]	1	0
[cou]	1.13 (0.08)	0
[mar]	1.21 (0.09)	0
[fin]	1.01 (0.08)	0
[def]	0.78 (0.09)	1
[ris]	0.73 (0.13)	1.53 (0.26)
[rap]	0.72 (0.11)	1.23 (0.21)
Common factor variance:	5.22 (0.67)	3.30 (0.80)
RANDOM PART: District level		
Common factor variance:	0.36 (0.17)	0

Slide 24

Some links and references

- The gllamm programs and manual can be downloaded from www.iop.kcl.ac.uk/iop/departments/biocomp/programs/gllamm.html

- Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2002). Generalized multilevel structural equation modelling. *Psychometrika*, conditionally accepted.

- Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal, 2*, 1–21.

- Skrondal, A. & Rabe-Hesketh, S. (2002). Multilevel logistic regression for polytomous data and rankings. *Psychometrika*, in press.

- Skrondal, A. & Rabe-Hesketh, S. (2003). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models.* Boca Raton, FL: Chapman & Hall/ CRC.

If you would like a copy of any of the papers, email me: spaksrh@iop.kcl.ac.uk