Multilevel Selection Models using `gllamm`

Stata User Group Meeting in Maastricht, May 2002

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

Joint work with
Anders Skrondal, Norwegian Institute of Public Health
and
Andrew Pickles, The University of Manchester

Gllamm can be downloaded from
http://www.iop.kcl.ac.uk/iop/departments/biocomp/programs/gllamm.html

Outline

Brief introduction to GLLAMM

The Heckman model and extensions

The Hausman-Wise-Diggle-Kenward dropout model and extensions

Application: Cluster randomized study of sex education in Norwegian schools

Overview of GLLAMM models

- Response model: Generalised linear model conditional on latent variables
  - Linear predictor: latent variables as factors or random coefficients
  - Links and distributions
- Structural model: Equations for the latent variables
  - Regressions of latent variables on observed variables
  - Regressions of latent variables on other latent variables
- Distribution of the latent variables (disturbances)
  - Multivariate normal
  - Discrete

Linear Predictor in GLLAMM

$$\eta = \boldsymbol{\beta}'\mathbf{x} + \sum_{l=2}^{L}\sum_{m=1}^{M_l} u_m^{(l)} \boldsymbol{\lambda}_m^{(l)\prime}\mathbf{z}_m^{(l)} \quad \text{for identification, } \lambda_{m1}^{(l)} = 1$$

- Fixed part: $\boldsymbol{\beta}'\mathbf{x}$ as usual
- Random part:
  - $u_m^{(l)}$ is mth latent variable at level l, $m = 1, \cdots, M_l$, $l = 2, \cdots, L$
  - $u_m^{(l)}$ can be a **factor** or a **random coefficient**
  - $\mathbf{z}_m^{(l)}$ are variables and $\boldsymbol{\lambda}_m^{(l)}$ are parameters
  - Unless regressions for the latent variables are specified, latent variables at different levels are independent whereas latent variables at the same level may be correlated.

Random coefficient models in GLLAMM

- One covariate multiplies each latent variable,

$$u_m^{(l)} z_m^{(l)} \quad (\lambda_m^{(l)} = 1)$$

- e.g. Latent growth curve model for individuals j (level 2) observed at times t_{ij}, $i = 1, \cdots, n_j$ (level 1)

$$\eta_{ij} = \beta_1 + \beta_2 t_{ij} + u_{1j}^{(2)} + u_{2j}^{(2)} t_{ij}$$

β_1, β_2 : mean intercept and slope

$u_{1j}^{(2)}$, $u_{2j}^{(2)}$: random deviations of the subject-specific intercepts and slopes from their means

- The model can also be defined as

$$\begin{aligned} \eta_{ij} &= b_{1j} + b_{2j} t_{ij} \\ b_{1j} &= \beta_1 + u_{1j} \\ b_{2j} &= \beta_2 + u_{2j} \end{aligned}$$

Slide 5

Factor models in GLLAMM

- A linear combination of dummy variables for the items multiplies each latent variable,

$$u_m^{(l)} \boldsymbol{\lambda}_m^{(l)\prime} \mathbf{z}_m^{(l)}$$

- e.g. One-factor model for items i, $i = 1, \cdots, I$ (level 1) and subjects j (level 2)

$$\begin{aligned} \eta_{ij} &= \beta_1 \delta_{1i} + \cdots + \beta_I \delta_{Ii} + u_j^{(2)} (\delta_{1i} + \lambda_2^{(2)} \delta_{2i} \cdots + \lambda_I^{(2)} \delta_{Ii}) \\ &= \beta_i + u_j^{(2)} \lambda_i^{(2)} \end{aligned}, \quad \delta_{pi} = \begin{cases} 1 & \text{if } p = i \\ 0 & \text{otherwise} \end{cases}$$

β_i: intercept for item i

$u_j^{(2)}$: common factor

$\lambda_i^{(2)}$: factor loading for item i, $\lambda_1^{(2)} = 1$

unit j	item i	δ_{1i}	δ_{2i}	$\cdots$	δ_{Ii}	y_{ij}
1	1	1	0	$\cdots$	0	y_{11}
1	2	0	1	$\cdots$	0	y_{21}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
1	I	0	0	$\cdots$	1	y_{I1}

Slide 6

Heckman selection model

- Selection equation (probit regression):

$$\begin{aligned} y_{1j}^* &= \boldsymbol{\gamma}' \mathbf{z}_j + \epsilon_{1j}, \quad \epsilon_{1j} \sim N(0,1), \\ y_{1j} &= \mathrm{I}(y_{i1}^* > 0) \end{aligned}$$

- Substantive equation (linear regression):

$$y_{2j} = \begin{cases} \boldsymbol{\alpha}' \mathbf{w}_j + \epsilon_{2j}, \ \epsilon_{2j} \sim N(0, \sigma^2) & \text{if } y_{i1} = 1 \\ \text{missing} & \text{if } y_{i1} = 0 \end{cases},$$

- Correlation

$$\mathrm{cor}(\epsilon_{1j}, \epsilon_{2j}) = \rho$$

- Missingness or non-selection is
  - Completely at random if $\boldsymbol{\gamma} = 0$ and $\rho = 0$
  - At random if $\rho = 0$
  - Informative if $\rho \neq 0$

Slide 7

Heckman selection model as a GLLAMM model

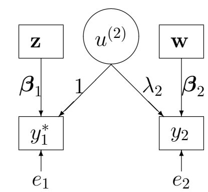
- Parameterize random part as

$$\epsilon_{ij} = u_j^{(2)} \lambda_i + e_{ij}, \quad \mathrm{var}(u_j^{(2)}) = 1, \ \lambda_1 = 1, \ \mathrm{var}(e_{ij}) = \nu^2, \ \mathrm{cov}(e_{1j}, e_{2j}) = 0$$

- Write as a GLLAMM model

$$\eta_{ij} = \boldsymbol{\beta}_1' \mathbf{z}_{ij} \delta_{1i} + \boldsymbol{\beta}_2' \mathbf{w}_{ij} \delta_{2i} + u_j^{(2)} (\delta_{1i} + \lambda_2 \delta_{2i}), \quad \mathrm{var}(u_j^{(2)}) = 1,$$

$$y_{ij} | \eta_{ij} \sim \begin{cases} \text{Bernouilli}(\Phi(\eta_{ij}/\nu)) \text{ if } i = 1 & \text{(Binomial with scaled probit link)} \\ N(\eta_{ij}, \nu^2) \text{ if } i = 2 & \text{(Gaussian with identity link)} \end{cases}$$



- Equivalences:

$$\sigma^2 = \lambda_2^2 + \nu^2, \ \rho = \frac{\lambda_2}{\sqrt{(\lambda_2^2 + \nu^2)(1 + \nu^2)}}, \quad \boldsymbol{\gamma} = \frac{\boldsymbol{\beta}_1}{\sqrt{1 + \nu^2}}, \quad \boldsymbol{\alpha} = \boldsymbol{\beta}_2$$

Slide 8

Syntax for linear predictor in `gllamm`

gllamm [*varlist*] [if *exp*] [in *range*] , i(*varlist*) [nrf(*numlist*)
    eqs(*eqnames*) noconstant offset(*varname*) constraints(*numlist*)
    . . .

`i(`*varlist*`)` $L-1$ variables identifying the hierarchical, nested clusters, from level 2 to L, e.g., `i(pupil class school)`.

`nrf(`*numlist*`)` $L-1$ numbers specifying the numbers of latent variables M_l at each level.

`eqs(`*eqnames*`)` $M=\sum M_l$ equations for the $\boldsymbol{\lambda}_m^{(l)}\mathbf{z}_m^{(l)}$ multiplying each latent variable. No constant is assumed unless explicitly included in the equation definition.

`noconstant` no constant in the fixed part $\boldsymbol{\beta}'\mathbf{x}$.

`offset(`*varname*`)` variable in fixed part with regression coefficient set to 1.

`constraints(`*numlist*`)` list of linear parameter constraints defined using the `constraint define` command.

Slide 9

Heckman selection model - linear predictor in `gllamm`

- `heckman` command in Stata

```
heckman y2 w, select(y1 = z w)
```

- Linear predictor in `gllamm`

$$\eta_{ij}=\beta_1 z_{ij}\delta_{1i}+\beta_2 w_{ij}\delta_{1i}+\beta_3\delta_{1i}+\beta_4 w_{ij}\delta_{2i}+\beta_5\delta_{2i}+u_j^{(2)}(\delta_{1i}+\lambda_2\delta_{2i})$$

- Data manipulation

```
gen id = _n
reshape long y, i(id) j(var)
tab var, gen(i)                  /* i1 = δ1i, i2 = δ2i */
gen z_i1 = z*i1
gen w_i1 = w*i1
gen w_i2 = w*i2
```

- `gllamm` command

```
eq load: i1 i2                   /* for u_j^(2)(δ1i + λ2δ2i) */
constraint define 1 [id1]i1 = 1  /* sets var(u_j^(2)) = 1 */
gllamm y z_i1 w_i1 i1 w_i2 i2, i(id) eqs(load) nocons constr(1) /*
    */  << more options >>
```

Slide 10

Links and families in GLLAMM

- The conditional expectation of the response is 'linked' to the linear predictor

$$g(E[y|\mathbf{x},\mathbf{u},\mathbf{z}])=\eta$$

- The conditional distribution of the response is from the exponential family
- The response variables may be of mixed type - requiring mixed links and families:

Links
identity
reciprocal
logarithm
logit
probit
scaled probit
compl. log-log

Families
Gaussian
gamma
Poisson
binomial

Polytomous responses
ordinal logit
ordinal probit
ordinal compl. log-log
scaled ord. probit
multinomial logit

- Heteroscedasticity: The dispersion parameter for the Gauss and gamma families and the scale for the scaled probit link can depend on covariates: $\log\phi=\boldsymbol{\alpha}'\mathbf{z}^{(1)}$

Slide 11

Options for links and families in `gllamm`

[· · · family(*families*) fv(*varname*) link(*links*) lv(*varname*) nats
    s(*eqname*) · · ·]

`family(`*families*`)` family or families to be used.

`fv(`*varname*`)` variable whose values indicate which family applies to which observation.

`link(`*links*`)` and `lv(`*varname*`)` analogous to `family(`*families*`)` and `fv(`*varname*`)`.

`nats` option to estimate the scale parameter directly instead of its logarithm.

`s(`*eqname*`)` equation for (log) scale parameter.

- Heckman selection model: links and families

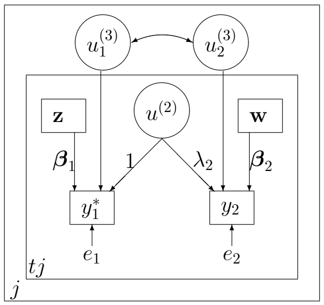
```
reshape long y, i(id) j(var)   /* var = 1 for y=y1 and 2 for y=y2 */
tab var, gen(i)
     << more data manipulation >>
gllamm y z_i1 w_i1 i1 w_i2 i2, nocons i(id) eq(load) constr(1) /*
  */ family(binom gauss) fv(var) link(sprobit ident) lv(var)   /*
  */ nip(10) adapt
```

Slide 12

Multilevel Heckman selection model

- Example: longitudinal data with observations at times t on subjects j where data are missing intermittently
- Add correlated subject level random effects $u_{j1}^{(3)}$ for the selection model and $u_{j2}^{(3)}$ for the substantive model:

$$\eta_{itj} = \boldsymbol{\beta}_1{}'\mathbf{z}_{tj}\delta_{1i} + \boldsymbol{\beta}_2{}'\mathbf{w}_{tj}\delta_{2i} + u_{tj}^{(2)}(\delta_{1i} + \lambda\delta_{2i}) + u_{j1}^{(3)}\delta_{1i} + u_{j2}^{(3)}\delta_{2i}, \quad \text{var}(u_{tj}^{(2)}) = 1$$



- The variances of $u_1^{(3)}$ and $u_2^{(3)}$ are identified through the intraclass correlations in the selection and substantive models respectively.

Slide 13

Multilevel Heckman selection model in `gllamm`

- Linear predictor

$$\eta_{itj} = \beta_1 z_{tj}\delta_{1i} + \beta_2 w_{tj}\delta_{1i} + \beta_3\delta_{1i} + \beta_4 w_{tj}\delta_{2i} + \beta_5\delta_{2i} + u_{tj}^{(2)}(\delta_{1i} + \lambda\delta_{2i}) + u_{j1}^{(3)}\delta_{1i} + u_{j2}^{(3)}\delta_{2i}$$

- `gllamm` command

`eq load: i1 i2` `/* for` $u_{tj}^{(2)}(\delta_{1i} + \lambda_2\delta_{1i})$ `*/`

`eq i1: i1` `/* for` $u_{j1}^{(3)}$ `*/`

`eq i2: i2` `/* for` $u_{j2}^{(3)}$ `*/`

`constraint define 1 [t1]i1 = 1` `/* sets` $\text{var}(u_{tj}^{(2)}) = 1$ `*/`

```
gllamm y z_i1 w_i1 i1 w_i2 i2, nocons i(t id) nrf(1 2) eq(load i1 i2) /*
  */ constr(1) family(binom gauss) fv(var) link(sprobit ident) lv(var) /*
  */ nip(19 15) ip(m) adapt
```

Slide 14

Hausman-Wise-Diggle-Kenward dropout model

- Longitudinal data at times $t = 1, 2, 3$ for subjects j. Subjects drop out at some time $t > 1$ and never return
- Substantive model (without autocorrelated errors)

$$y_{tj} = \boldsymbol{\beta}'\mathbf{x}_{tj} + u_j + \epsilon_{tj}, \quad \epsilon_{tj} \sim N(0, \sigma^2), \quad u_j \sim N(0, \tau^2)$$

- Dropout model ($d_{tj} = 1$ if subject j drops out at time $t > 1$)

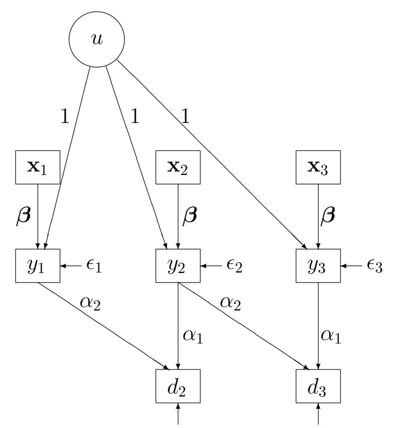
$$\text{logit}(\Pr(d_{tj} = 1)) = \alpha_0 + \alpha_1 y_{tj}^* + \alpha_2 y_{t-1,j}, \quad y_{tj}^* = \begin{cases} \text{observed } y_{tj} & \text{if } d_{tj} = 0 \\ \text{unobserved } y_{tj} & \text{if } d_{tj} = 1 \end{cases}$$

- Dropout is
  - Completely at random if $\alpha_1 = \alpha_2 = 0$
  - At random if $\alpha_1 = 0$
  - Informative if $\alpha_1 \neq 0$

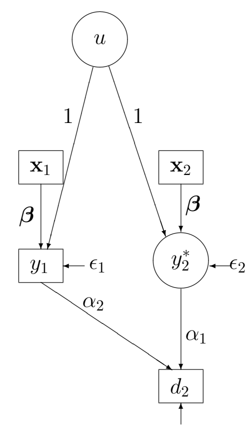
Slide 15

Diagram for dropout model

Complete data ($d_2 = d_3 = 0$)



Dropout at time 2 ($d_2 = 1$)



Slide 16

Structural model in GLLAMM

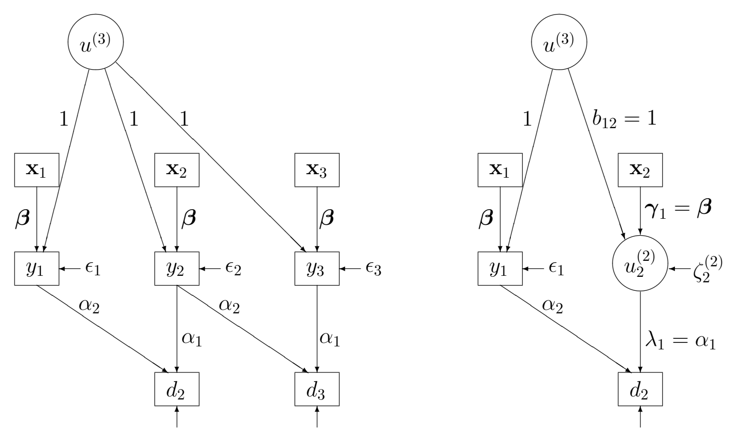
Regressions of latent variables on other latent and explanatory variables

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

- $\mathbf{u} = (u_1^{(2)}, u_2^{(2)}, \cdots, u_{M_2}^{(2)}, \cdots, u_1^{(l)}, \cdots, u_{M_l}^{(l)}, \cdots, u_{M_L}^{(L)})'$ (M elements)
  - factors
  - random coefficients
- $\mathbf{B}$ is an upper diagonal $M \times M$ matrix of regression coefficients
- $\mathbf{\Gamma}$ 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 errors/disturbances (same level as corresponding elements in $\mathbf{u}$).

Hausman-Wise-Diggle-Kenward dropout model in GLLAMM

Complete data ($d_2 = d_3 = 0$)   Dropout at time 2 ($d_2 = 1$)



Hausman-Wise-Diggle-Kenward dropout model in GLLAMM

- Linear predictor

$$\begin{aligned}\eta_{itj} &= \delta_{1i}\left(\boldsymbol{\beta}'\mathbf{x}_{tj} + u_j^{(3)}\right) + \delta_{2i}\left(\alpha_0 + \alpha_1 y_{tj}(1-d_{tj}) + \lambda_1 u_{tj}^{(2)} d_{tj} + \alpha_2 y_{i-1,j}\right)\\ &= \boldsymbol{\beta}'\mathbf{x}_{tj}\delta_{1i} + \alpha_0\delta_{2i} + \alpha_1 y_{tj}\delta_{2i}(1-d_{tj}) + \alpha_2 y_{i-1,j}\delta_{2i} + u_{tj}^{(2)}\lambda_1\delta_{2i}d_{tj} + u_j^{(3)}\delta_{1i}\end{aligned}$$

- Response process

$$y_{itj}|\eta_{itj} \sim \begin{cases} N(\eta_{itj}, \sigma^2) \text{ if } i = 1 & \text{(Gaussian with identity link)} \\ \text{Bernouilli}\left(\frac{\exp(\eta_{itj})}{1+\exp(\eta_{itj})}\right) \text{ if } i = 2 & \text{(Binomial with logit link)} \end{cases}$$

- Structural model

$$\begin{bmatrix} u_{tj}^{(2)} \\ u_j^{(3)} \end{bmatrix} = \begin{bmatrix} 0 & b_{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_{tj}^{(2)} \\ u_j^{(3)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\gamma}_1{}' \\ \mathbf{0}' \end{bmatrix} \mathbf{x}_{tj} + \begin{bmatrix} \zeta_{tj}^{(2)} \\ \zeta_j^{(3)} \end{bmatrix}$$

- Constraints

$$\lambda_1 = \alpha_1, \quad \boldsymbol{\gamma}_1 = \boldsymbol{\beta}, \quad \text{var}(\zeta_{tj}) = \sigma^2$$

Options for the structural model

[··· `bmatrix`(*matname*) `geqs`(*eqnames*) `frload`(*numlist*) ···]

`bmatrix`(*matrix*) $M \times M$ matrix of 1s and 0s. Elements equal to 0 indicate that the corresponding element in $\mathbf{B}$ is 0; elements equal to 1 that the corresponding element in $\mathbf{B}$ should be estimated.

`geqs`(*eqnames*) equations for regressions of latent variables on explanatory variables. The second character of each equation name indicates which latent variable is regressed on the predictors.

`frload`(*numlist*) frees first factor loading for latent variables corresponding to *numlist*.

Estimating the Hausman-Wise-Diggle-Kenward model

- Data manipulation (data are in long form)

```
gen y0 = cond(y~=.,y,0)                       /* lag 0  y_tj  */
sort id t
qui by id: gen y1 = cond(_n>1,y[_n-1],0)  /* lag 1  y_t-1,j */
gen d = y == .                                /* d_tj */
sort id d t
qui by id d: drop if d==1&_n>1    /* drop records after first missing */
gen resp1 = y
gen resp2 = d
reshape long resp, i(id t) j(var) /* var = 1,2 if resp = y,d */
drop if var == 2 & t == 1         /* no dropout at time 1 */
tab var, gen(i)                   /* i1 = δ_1i, i2 = δ_1i */
gen x_i1 = x*i1
gen y0_i2d0 = y0*i2*(1-d)         /* y_tj δ_2i (1 - d_tj) */
gen y1_i2 = y1*i2
gen i2d1 = i2*d
```



```
. list id var t resp y0_i2d0 y1_i2 i2 in 1/12
```

	id	var	t	resp	y0_i2d0	y1_i2	i2d1
1.	1	1	1	2.657621	0	0	0
2.	1	2	2	1	0	2.657621	1
3.	2	1	1	1.423789	0	0	0
4.	2	2	2	1	0	1.423789	1
5.	3	1	1	-.7317839	0	0	0
6.	3	1	2	-.597519	0	0	0
7.	3	1	3	.8041697	0	0	0
8.	3	2	2	0	-.597519	-.7317839	0
9.	3	2	3	0	.8041697	-.597519	0
10.	4	1	1	.4663057	0	0	0
11.	4	1	2	1.797121	0	0	0
12.	4	2	2	0	1.797121	.4663057	0

- Linear predictor

$$\eta_{itj} = \boldsymbol{\beta}'\mathbf{x}_{tj}\delta_{1i} + \alpha_0\delta_{2i} + \alpha_1 y_{tj}\delta_{2i}(1-d_{tj}) + \alpha_2 y_{i-1,j}\delta_{2i} + u_{tj}^{(2)}\lambda_1\delta_{2i}d_{tj} + u_j^{(3)}\delta_{1i}$$

- Syntax for `gllamm`

```
eq u_2: i2d1            /* for eqs(): u_tj^(2) λ_1 δ_2i d_tj */
eq u_3: i1              /* for eqs(): u_j^(3) δ_1i */
matrix B=(0,1\0,0)      /* for bmatrix() */
gen one = 1
eq f1: x one             /* for geqs(): γ_2 x_tj  +  γ_1 */

constraint def 1 [b1_2]_cons = 1            /* set b_12 = 1 */
constraint def 2 [f1]one = [resp]i1         /* set γ_1 = β_1 */
constraint def 3 [f1]x = [resp]x            /* set γ_2 = β_2 */
constraint def 4 [t1]i2d1 = [s1]_cons       /* set var(ζ_jt^(2)) = var(ε_jt) */
constraint def 5 [t1l]i2d1 = [resp]y0_i2d0  /* set λ_1  =  α_1 */

gllamm resp x_i1 i1 y0_i2d0 y1_i2 i2, i(t id) eqs(u_2 u_3)        /*
  */ nocons family(gauss binom) fv(var) link(ident probit) lv(var) /*
  */ bmat(B) geqs(f1) frload(1) nats constr(1/5) nip(7) adapt
```


Extensions of the Hausman-Wise-Diggle-Kenward model

Application

- Cluster randomised study of sex education in Norway
- Schools were randomised to receive sex education or not
- Assessments pre randomisation, 6 months and 18 months post randomisation
- Three ordinal outcomes (5-point scale) measuring readiness to use contraception: "If my partner and I were about to have intercourse without either of us having mentioned contraception ...
  - I would have no problems saying that I have no contraception"
  - I would have no problems asking my partner whether he/she has contraception"
  - it would be easy for me to produce a condom (if I brought one)"
- 46 schools and 1183 pupils contributed to the analysis

Model

- Factor model with ordinal logit link for three outcomes i at time t for pupil j in school k

$$y^*_{itjk} = \beta_i + u^{(2)}_{tjk}\lambda_i + [u^{(3)}_{jk}0 + u^{(4)}_{k}0] + \epsilon_{itjk},\ \beta_1 = 0$$

$$y_{itjk} = s \text{ if } \kappa_{s-1} < y^*_{itjk} \le \kappa_s,\ s = 1, \cdots 5,\ \infty = \kappa_0 < \kappa_1 < \cdots < \kappa_5 = \infty$$

- Substantive model: structural model for latent outcome $u^{(2)}_{tjk}$

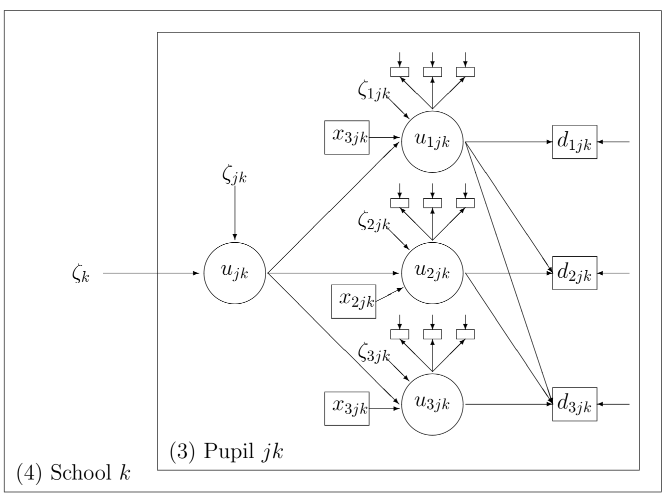
$$\begin{aligned} u^{(2)}_{tjk} &= \gamma_1 x_{Ttij} + \gamma_2 x_{Itij} + \gamma_3 x_{Ttij} x_{Itij} + u^{(3)}_{jk} + \zeta^{(2)}_{tjk} \\ u^{(3)}_{jk} &= u^{(4)}_{k} + \zeta^{(3)}_{jk} \end{aligned}$$

  where x_{Ttij} is time (0,1,3) and x_{Itij} is an indicator for the intervention group.

- Selection model

$$\text{logit}(\Pr(d_{tjk} = 1)) = \beta_6 + \alpha_0 u^{(2)}_{tjk} + \alpha_1 u^{(2)}_{t-1jk} + \alpha_2 u^{(2)}_{t-2jk}$$

Slide 25

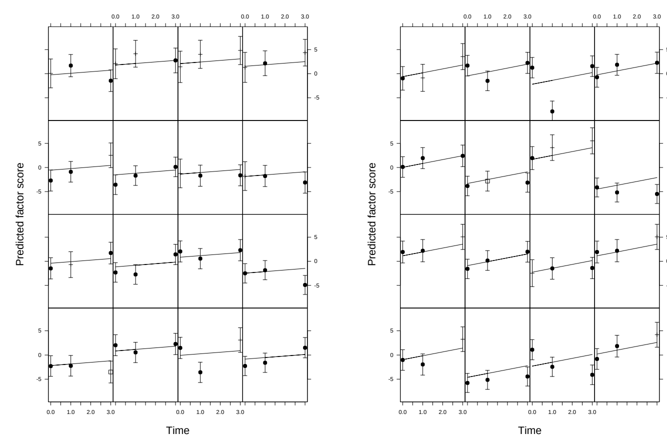


Slide 26

	Model 1		Model 2		Model 3		Model 4	
	estimate	se	estimate	se	estimate	se	estimate	se
Selection model								
β_6	-1.99	0.26	-2.10	0.29	-0.85	0.04	-1.09	0.06
α_0	0.77	0.09	0.64	0.09	–	–	–	–
α_1	-0.10	0.04	–	–	–	–	-0.07	0.03
α_2	-0.20	0.05	–	–	–	–	-0.25	0.04
Substantive model								
γ_1 (time)	0.32	0.10	0.45	0.09	-0.06	0.09	-0.39	0.09
γ_2 (interv.)	-0.91	0.26	-0.25	0.23	-0.28	0.24	-1.46	0.21
γ_3 (time by interv.)	**0.48**	**0.11**	**0.34**	**0.10**	**0.20**	**0.11**	**0.56**	**0.11**
var($\zeta^{(1)}_{tjk}$)	6.74	0.60	7.29	0.68	4.57	0.41	5.04	0.45
var($\zeta^{(2)}_{jk}$)	5.30	0.57	4.29	0.98	3.72	0.43	3.51	0.39
Measurement model				Not shown				
log-likelihood	-8624.49		-8631.63		-8680.35		-8657.93	

Slide 27

Empirical Bayes predictions for control (left) and intervention group (right) with standard errors



Slide 28