Centre for Market and
Public Organisation

4th ESRC Research Methods Festival
Oxford, 5-8 July 2010

School Markets & Correlated Random Effects

Paul Clarke

(with Fiona Steele, Harvey Goldstein and Rich Harris)

Outline

- School markets: competition & sorting
- Impact on multilevel modelling
- Methodology: correlated random effects
- ALSPAC data analysis
- Further work

School Markets

- Britain after 1944
  - Local Education Authority (LEA) control
  - ‘Catchment area’-based pupil allocation
- Education Reform Act (1988)
  - Reduced influence of LEA/catchment area
  - ‘Quasi-market’ = ‘parental choice’
  - Performance tables (GCSE, Key-stage, etc.)

2-level Random Intercepts Model

- Standard notation

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij}$$

- Drop j to emphasize selection mechanism

$$y_i = \beta_0 + \beta_1 x_i + \mathbf{z}_i' \mathbf{u} + e_i$$

$$\mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_q \end{pmatrix} \qquad \mathbf{z}_i = \begin{pmatrix} z_{i1} \\ \vdots \\ z_{iq} \end{pmatrix} \quad z_{ij} = \begin{cases} 1 & \text{if pupil } i \text{ in school } j \\ 0 & \text{otherwise} \end{cases}$$

Interpretation

- Ideally interpret u_j as ‘school effects’
  - e.g. teachers, ethos, size, special needs provision

$$s_j \lambda = u_j$$

Standard Assumptions

- School effects distribution $\mathbf{u} \sim N(\mathbf{0}, I\sigma_u^2)$
- No competition
  - Schools set s_j independently (e.g. nationally)
- School competition
  - e.g. For successive cohorts 1999 and 2000:

$$u_j^{2000} = w_j u_j^{1999} + \sum_{k \neq j} w_k u_k^{1999} + \varepsilon_j^{2000}$$

Selection/Sorting

- Parents' choice of school non-random
- Determined by selection mechanism

$$\pi_{ij}(\mathbf{u}) \equiv \Pr(z_{ij} = 1 | x_i, \mathbf{u})$$

i.e. multinomial

$$\Rightarrow \pi_i(\mathbf{u}) \equiv \Pr(\mathbf{z}_i | x_i, \mathbf{u}) = \prod_j \pi_{ij}(\mathbf{u})^{z_{ij}}$$

‘Random Effects’ Assumption

- Ideally school residual = school effect

- But only under this condition

$$\mathrm{Cov}(\mathbf{z}_i'\mathbf{u}, x_i) = 0$$

- What if selection depends on school effects?

Impact of Selection

- Under weak assumptions[1]

$$p(\mathbf{u} | \text{observed}) \propto p(\mathbf{u}) \prod_i \pi_i(\mathbf{u})$$

- If selection independent of $\mathbf{u}$ then

$$p(\mathbf{u} | \text{observed}) = p(\mathbf{u})$$

  – i.e. r. effects assumption & uncorrelated

---

[1]Schools respond to drivers of selection but population itself remains fixed

Impact of Selection (cont.)

- If selecting school j depends on u_j then

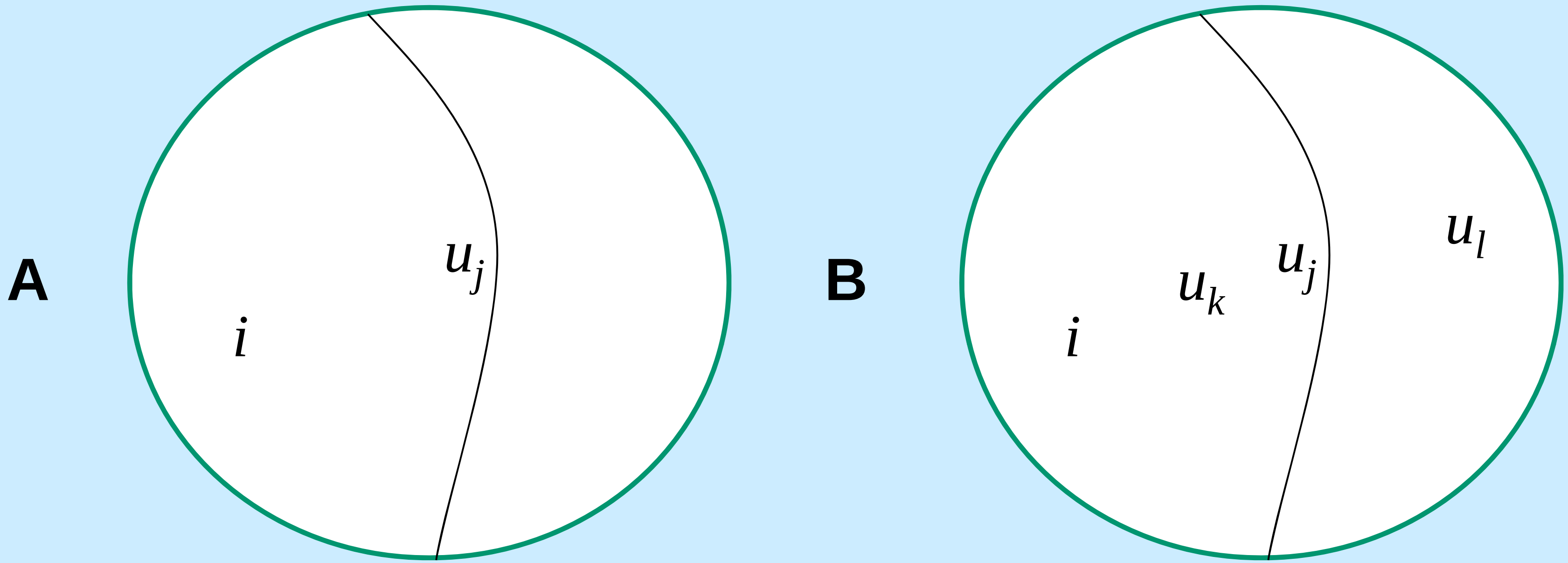
$$p(\mathbf{u}|\text{obs}) \propto \prod_j \left[\; p(u_j)\prod_i \pi_{ij}(u_j)^{z_{ij}} \; \right]$$

  - i.e. r. effects assumption fails
  - Heteroskedastic but uncorrelated residuals

- Otherwise: … plus correlated residuals

Plausibility

- Yes: e.g. if spatial selection element



$$\pi_{ij}^{A}(\mathbf{u}) \neq \pi_{ij}^{B}(\mathbf{u})$$

MCMC Methodology

- From Browne & Goldstein (2010)[1]
  – Adaptive Gibbs/Metropolis-Hastings hybrid

- Level-two covariance matrices of form:

$$\mathrm{Cov}(\mathbf{u} \mid \mathrm{obs}) = \begin{pmatrix} \sigma_u^2 & & \{\sigma_u^2 \rho_{jk}\} \\ & \ddots & \\ \{\sigma_u^2 \rho_{jk}\} & & \sigma_u^2 \end{pmatrix}$$

[1] "MCMC sampling for a random intercepts model with non-independent residuals within & between cluster units", *J. Educational & Behavioral Statistics* (in press)

ALSPAC Application

- Avon Longitudinal Study of Parents & Children
  - Followed up all births in Avon 1991-1992
  - 14000 children followed up

- Analyse primary schools (key-stage 2)
  - Children tested 10-11y
  - Mathematics and English test scores

Correlation Model

- Link function is $\tanh^{-1}$

$$\rho_{jk} = \frac{\exp(g_{jk}) - 1}{\exp(g_{jk}) + 1} \qquad g_{jk} = \alpha(d_{jk})$$

- Core catchment areas (CCAs)
  - ‘Distance’ d_{jk} is proportional CCA overlap[1]
  - School’s CCA is area containing 50% of its pupils
  - Zero overlap $\Rightarrow$ zero correlation

---

[1] Harris & Johnston (2008), “Primary schools, markets and choice: studying polarization and the Core Catchment Areas of Schools”, *Appl. Spatial Analysis* **1**, 59-84.

Results

2-level Model for KS2 Mathematics Scores

Parameters	Uncorrelated	Linear	Quadratic	Piecewise[1]
$\hat{\beta}_0$	62.8	63.0	63.0	63.0
$\hat{\sigma}_e$	20.0	20.0	20.0	20.0
$\hat{\sigma}_u$	6.4	7.5	7.5	7.5
$\hat{\alpha}_1$	0	–0.07	1.64	–0.11
		(–1.4,1.3)	(–3.6,2.9)	(–0.7,0.5)
$\hat{\alpha}_2$	0		3.14	0.15
			(–5.6,6.9)	(–0.7,1.0)
$\hat{\alpha}_3$	0			0.11
				(–0.7,1.0)
$\hat{\alpha}_4$	0			0.11
				(–0.7,1.0)
Correlation				
10% overlap	0	–0.03	0.23	–0.05
50% overlap	0	–0.02	0.83	0.00
90% overlap	0	0.00	0.97	0.00

[1] Piecewise for percentiles: 10% α_1; 10-50% $\alpha_1 + \alpha_2$; 50-90% $\alpha_1 + \alpha_2$; 90-100% $\alpha_1 + \alpha_3$

Further Work

- ALSPAC example:
  - No evidence of correlation in primary schools
  - Robustness to CCA definition
  - Analyse secondary schools
- Possible that
  - Two sources cancel out?
  - Possible that markets entrench difference