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

\[ y_i = \beta_0 + \beta_1 x_i + z_i' u + e_i \]

\[
\begin{align*}
\mathbf{u} &= \begin{pmatrix} u_1 \\ \vdots \\ u_q \end{pmatrix} \\
\mathbf{z}_i &= \begin{pmatrix} z_{i1} \\ \vdots \\ z_{iq} \end{pmatrix} \\
\end{align*}
\]

\[ z_{ij} = \begin{cases} 1 & \text{if pupil } i \text{ in school } j \\ 0 & \text{otherwise} \end{cases} \]
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:

\[
\mathbf{u}_{j,2000} = \mathbf{w}_j \mathbf{u}_{j,1999} + \sum_{k \neq j} \mathbf{w}_k \mathbf{u}_{k,1999} + \mathbf{e}_{j,2000}
\]
Selection/Sorting

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

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

i.e. multinomial

\[
\Rightarrow \pi_i(u) \equiv \Pr(z_i|x_i, u) = \prod_j \pi_{ij}(u)^{z_{ij}}
\]
‘Random Effects’ Assumption

• Ideally school residual = school effect

• But only under this condition

\[ \text{Cov}(z'_i u, x_i) = 0 \]

• What if selection depends on school effects?
Impact of Selection

• Under weak assumptions¹

\[ p(u|\text{observed}) \propto p(u) \prod_i \pi_i(u) \]

• If selection independent of \( u \) then

\[ p(u|\text{observed}) = p(u) \]

- i.e. r. effects assumption & uncorrelated

¹Schools 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

\[ \pi_{ij}^A (u) \neq \pi_{ij}^B (u) \]
MCMC Methodology

- From Browne & Goldstein (2010)\(^1\)
  - Adaptive Gibbs/Metropolis-Hastings hybrid

- Level-two covariance matrices of form:

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

\(^1\) “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 $\tanh^{-1}$

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

• 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

---

## Results

### 2-level Model for KS2 Mathematics Scores

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Uncorrelated</th>
<th>Linear</th>
<th>Quadratic</th>
<th>Piecewise$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_0$</td>
<td>62.8</td>
<td>63.0</td>
<td>63.0</td>
<td>63.0</td>
</tr>
<tr>
<td>$\hat{\sigma}_e$</td>
<td>20.0</td>
<td>20.0</td>
<td>20.0</td>
<td>20.0</td>
</tr>
<tr>
<td>$\hat{\sigma}_u$</td>
<td>6.4</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>0</td>
<td>$-0.07$</td>
<td>1.64</td>
<td>$-0.11$</td>
</tr>
<tr>
<td>&amp;</td>
<td>$(-1.4,1.3)$</td>
<td>$(-3.6,2.9)$</td>
<td>$(-0.7,0.5)$</td>
<td></td>
</tr>
<tr>
<td>$\hat{\alpha}_2$</td>
<td>0</td>
<td>3.14</td>
<td>0.15</td>
<td>$(-5.6,6.9)$</td>
</tr>
<tr>
<td>$\hat{\alpha}_3$</td>
<td>0</td>
<td></td>
<td></td>
<td>$(-0.7,1.0)$</td>
</tr>
<tr>
<td>$\hat{\alpha}_4$</td>
<td>0</td>
<td></td>
<td></td>
<td>0.11</td>
</tr>
</tbody>
</table>

### Correlation

<table>
<thead>
<tr>
<th>Overlap</th>
<th>Uncorrelated</th>
<th>Linear</th>
<th>Quadratic</th>
<th>Piecewise$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10% overlap</td>
<td>0</td>
<td>$-0.03$</td>
<td>0.23</td>
<td>$-0.05$</td>
</tr>
<tr>
<td>50% overlap</td>
<td>0</td>
<td>$-0.02$</td>
<td>0.83</td>
<td>0.00</td>
</tr>
<tr>
<td>90% overlap</td>
<td>0</td>
<td>0.00</td>
<td>0.97</td>
<td>0.00</td>
</tr>
</tbody>
</table>

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