The limitations of using school league tables to inform school choice

George Leckie and Harvey Goldstein
Introduction

• Each year the government publishes schools’ GCSE results and contextual value-added (CVA) performances in school league tables
  – They estimate value-added scores using multilevel models

• A principal justification for this is to inform parental choice of secondary schools

• A crucial limitation of these tables is that the most recent published information is based on a cohort of pupils who are 7 years ahead of the cohort of interest

• For choosing a school, it is the future performance of schools that is of interest

• The government make no adjustment for the statistical uncertainty that arises from making predictions into the future
### Secondary School achievement and attainment tables 2007

**LA: Oxfordshire**

#### Key Stage 2 to 4 Contextual Value Added

<table>
<thead>
<tr>
<th>GCSE and equivalent</th>
<th>Key Stage 2 to 4 CVA</th>
<th>Key Stage 3 to 4 CVA</th>
<th>Year on year comparisons</th>
<th>Absence</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Measure centred on 1000</td>
<td>Limit of Key Stage 2 to 4 CVA Confidence Intervals</td>
<td>Coverage</td>
<td>Number of qualifications</td>
<td></td>
</tr>
<tr>
<td>Banbury School</td>
<td>997.7</td>
<td>1005.8</td>
<td>989.7</td>
<td>97%</td>
<td>9.7</td>
</tr>
<tr>
<td>Bartholomew School</td>
<td>1009.1</td>
<td>1018.4</td>
<td>999.8</td>
<td>97%</td>
<td>10.0</td>
</tr>
<tr>
<td>Bicester Community College</td>
<td>994.3</td>
<td>1002.8</td>
<td>985.7</td>
<td>92%</td>
<td>10.8</td>
</tr>
<tr>
<td>Blessed George Napier Catholic School and College</td>
<td>1002.1</td>
<td>1011.8</td>
<td>992.4</td>
<td>96%</td>
<td>10.2</td>
</tr>
</tbody>
</table>
Seven years out of date

- During October 2008 parents will choose which secondary schools to send their children to

- These pupils will start secondary schooling in September 2009 and will take their GCSE examinations in 2014

- When choosing their secondary schools, the most recent published information will be for the cohort of pupils who take their GCSEs in 2007

- These two cohorts are seven years apart
Stability of school effects

• Previous literature has shown that whilst simple school averages are strongly correlated over time, value-added estimates of school effects are only moderately correlated

• Correlations of 0.5 - 0.6 for value-added estimates five years apart

• This limits the extent to which current school performance can be used as a guide to future performance
Data

• National Pupil Database (NPD)
  – Census of all state school pupils in England
  – Pupils test scores data at ages 11 and 16
  – Same data as is used to produce government school league tables

• Pupil Level Annual School Census (PLASC)
  – Provides data on pupil background characteristics
  – These are included in the CVA model specification

• We use data on the cohort of pupils that took their GCSEs in 2007

• We analyse a 10% random sample of all English secondary schools
  – 274 schools, approximately 190 pupils per school
Two-level multilevel model

- The traditional school effectiveness model is

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

\[ u_j \sim N\left(0, \sigma_u^2\right), \quad e_{ij} \sim N\left(0, \sigma_e^2\right) \]

- \( y_{ij} \) is the total GCSE score for pupil \( i \) in secondary school \( j \)
- \( x_{ij} \) is their achievement at age 11 intake
- \( u_j \) is the value-added school effect for secondary school \( j \)
- \( e_{ij} \) is the pupil level random effect
Predictor variables

• At the pupil level (level 1) we adjust for
  – Achievement at age 11
  – Month of birth
  – Gender
  – Free school meals
  – Special educational needs
  – English as an additional language
  – Ethnicity
  – Local neighbourhood deprivation

• We do not adjust for any school level (level 2) variables
School effects for the 2007 cohort

- Posterior estimates of the school effects and their associated variances are given by

\[ \hat{u}_j = \frac{n_j \sigma_u^2}{n_j \sigma_u^2 + \sigma_e^2} \tilde{y}_j, \quad \text{var}(\hat{u}_j - u_j) = \frac{\sigma_u^2 \sigma_e^2}{n_j \sigma_u^2 + \sigma_e^2} \]

- Assuming normality, standard 95% confidence intervals are calculated as

\[ \hat{u}_j \pm 1.96 \sqrt{\text{var}(\hat{u}_j - u_j)} \]

- These school effects are published in the DCSF school league tables
School effects for the 2007 cohort

~60% of schools are significantly different from the overall average
School effects for the 2014 cohort

- The previous school effects allow us to make inferences about how schools performed for the cohort that took their GCSEs in 2007.

- However, they do not allow us to make inferences about the likely performance of schools for future cohorts.

- We want to know whether the same significant differences remain in 2014.

- To do this, we need to adjust the estimates and standard errors of the 2007 school effects to reflect the additional uncertainty that arises from predicting into the future.

- The bivariate response version of the school effectiveness model provides a way to do this.
Bivariate response model

- The traditional school effectiveness model for two cohorts of pupils is

\[ y_{ij}^{(2007)} = \beta_0^{(2007)} + \beta_1^{(2007)} x_{ij}^{(2007)} + u_j^{(2007)} + e_{ij}^{(2007)} \]

\[ y_{ij}^{(2014)} = \beta_0^{(2014)} + \beta_1^{(2014)} x_{ij}^{(2014)} + u_j^{(2014)} + e_{ij}^{(2014)} \]

\[
\begin{bmatrix}
    u_j^{(2007)} \\
    u_j^{(2014)}
\end{bmatrix}
\sim N\left(\begin{bmatrix}
    0 \\
    0
\end{bmatrix}, \begin{bmatrix}
    \sigma_{u2007}^2 & 0 \\
    0 & \sigma_{u2014}^2
\end{bmatrix}\right),
\begin{bmatrix}
    e_{ij}^{(2007)} \\
    e_{ij}^{(2014)}
\end{bmatrix}
\sim N\left(\begin{bmatrix}
    0 \\
    0
\end{bmatrix}, \begin{bmatrix}
    \sigma_{e2007}^2 & 0 \\
    0 & \sigma_{e2014}^2
\end{bmatrix}\right)
\]

- The level 2 residuals are allowed to be correlated. The correlation measures the stability of school effects between the two cohorts.

- The level 1 residuals are modelled as independent as a pupil can only belong to one cohort.
School effects for the 2014 cohort

- It can be shown that the posterior estimates and variances of the school effects for the second cohort, given data only on the first cohort, are

\[
\hat{u}_{j}^{(2014)} = \frac{\rho \ n_{j}^{(2007)} \sigma_{u}^{2}}{n_{j}^{(2007)} \sigma_{u}^{2} + \sigma_{e2007}^{2}} \tilde{y}_{j}^{(2007)}, \quad \text{var} \left( \hat{u}_{j}^{(2014)} - u_{j}^{(2014)} \right) = \frac{n_{j}^{(2007)} \sigma_{u}^{4} (1 - \rho^{2}) + \sigma_{u}^{2} \sigma_{e2007}^{2}}{n_{j}^{(2007)} \sigma_{u}^{2} + \sigma_{e2007}^{2}}
\]

- Where, for simplicity, we have assumed that the school level variance is constant across cohorts

\[
\sigma_{u2007}^{2} = \sigma_{u2014}^{2} = \sigma_{u}^{2}
\]

- The two equations are the same as before, except for the addition of the terms in red

- The only term we don’t know is \( \rho \) the correlation between the two sets of school effects
Stability of school effects

- We want to estimate $\rho$ the 7 year apart correlation

- Ideally, we would estimate the bivariate response model based on two cohorts of pupils 7 years apart to obtain an estimate of $\rho$
  - Note, we assume that $\rho$ remains stable over time

- However, we only have data for cohorts five years apart (2002 and 2007)
  - The estimated correlation is 0.69
  - This is an overestimate of the 7 year apart correlation

- We can now adjust the estimates and standard errors of the 2007 school effects
School effects for the 2014 cohort

Only ~5% of schools are significantly different from the overall average
Comparison of the school effects for the 2007 and 2014 cohorts

- Note, these caterpillar plots only allow schools to be compared at the 5% level to the average school

- Different confidence intervals are required for pairwise comparisons
  - These are the types of comparisons which parents are interested in
Adjusting and not adjusting for school compositional variables

- The CVA model adjusts for two school level compositional variables
  - School mean of intake achievement
  - School spread of intake achievement

- This lowers the rankings of grammar schools

- Grammar schools admission policies lead them to have a high mean and narrow a spread of achievement at intake

- However, parents are interested in which schools will produce better subsequent achievement irrespective of whether this is due to school composition, policies or practices

\[ \rho = 0.76 \]
Conclusions

• School league tables make no adjustment for the statistical uncertainty that arises when current school performance is used to predict future school performance.

• Our main result is that, when we adjust for this uncertainty, the number of schools that can be separated from the average school drops from 60% to almost none.

• We also argue that, for the purpose of school choice, value-added measures should not adjust for school-level factors, since this is part of the very thing that parents are interested in.

• We show that adjusting for the school-level intake composition substantially alters the rank order of school effects:
  – Grammar schools drop down the rankings.
Conclusions (cont.)

• We do not propose our approach as a new means of producing league tables

• What we focus on is just one of a long list of statistical concerns that have been expressed about using results as indicators of school performance
  – Other concerns include the side effects and perverse incentives generated by the use of league tables

• However, we do feel that there is an accountability role for performance indicators as monitoring and screening devices to identify schools for further investigation
  – In which case, estimates for the 2007 cohort are the most appropriate
  – However, it is not clear whether to adjust for school compositional variables
  – Performance indicators will be of most use if combined with other sources of school information
Conclusions (cont.)

• Whilst we have focused on secondary school league tables, the issues we have discussed are relevant for other stages of schooling.

• Indeed, for primary schools our main result will be even more dramatic, since the small size of primary schools makes their estimated schools effects particularly imprecise.

• Scotland, Wales and Northern Ireland no longer publish school league tables, perhaps now is the time for England to stop.