A review of multilevel value-added models in education

Leonardo Grilli & Carla Rampichini

Department of Statistics, University of Florence
email: grilli@ds.unifi.it rampichini@ds.unifi.it

Methods for comparing educational institutions

Methodology developed in different fields: educational statistics, psychometrics, sociology, econometrics …

In this presentation we focus on the methodological challenges connected with statistical modelling and data analysis:
- definition of effectiveness in education
- multilevel models and their role in assessing effectiveness
- statistical issues arising in effectiveness evaluation
- use of model results

Effectiveness

The effectiveness of an organization is the degree of achievement of its institutional targets
- **absolute** (absolute effectiveness or impact analysis): evaluation of interventions, e.g. a specific vocational training course
- **relative** (relative or comparative effectiveness): comparison among many institutions

For *educational institutions (schools, universities)* the effectiveness cannot be defined in absolute terms, but only with respect to the effects on the students

In economic terms, the customers (students) are also inputs of the production function of the educational institution

The effects on the students are affected by the features of the students themselves: how to make a fair assessment?

Special issue of the Journal of Econometrics (2004): The econometrics of higher education
The analysis of the educational process is difficult → the quality of educational institutions is usually measured via an input/output approach:

- the process is a black-box
- the output (outcome) is evaluated in the light of the input → effectiveness = value added by the school

**VALUE-ADDED = ACTUAL OUTCOME minus EXPECTED OUTCOME GIVEN THE INPUT**


**Internal/external effectiveness**

The educational process leads to multiple outcomes → many measures of effectiveness

- **Internal effectiveness:**
  - Dropout (1=Yes, 0=No)
  - Duration of studies (time to the degree)
  - Number of credits after a given period

- **External effectiveness:**
  - Occupational status after degree (1=Yes, 0=No)
  - Duration of unemployment (time to first job)
  - Wage or job satisfaction

**Type A and B effectiveness**

- **Type A:** performance of the institution adjusted for the features of the students, irrespective of the context → to inform school choice

- **Type B:** performance of the institution adjusted also for the context (e.g. resources, local labour market, socio-economic composition of enrolled students) → for accountability


**Statistical issues**

The statistical models for assessing the relative effectiveness of educational institutions must face two main issues:

- **Adjustment:** the measures must be adjusted at least for the features of the students (necessary for a fair comparison)
- **Quantification of uncertainty:** the measures must be accompanied by error bars (necessary to make assessments properly supported by empirical evidence)

The raw rankings (so called 'League Tables') ignore both issues:
Statistical issues

Adjustment & Quantification of uncertainty

Regression models

But standard models are not suitable!

- Standard models make unsuitable assumptions on the variance-covariance structure (independence among observations, while the results of the students of the same school usually are positively correlated) → poor quantification of uncertainty
- Standard models are unable to represent some key features, e.g. varying slopes

Multilevel models

- Multilevel (mixed, random effects) models overcome the main limitations of standard models and are well suited for assessing the relative effectiveness of schools
  - The effectiveness of a school is explicitly represented by the random effects

Random intercept model: definition

\[
Y_{ij} = \alpha + \beta x_{ij} + \gamma w_j + u_j + e_{ij}
\]

\[
= (\alpha + u_j) + \beta x_{ij} + \gamma w_j + e_{ij}
\]

Intercept of j-th school

Random intercept model: value added

\[
Y_{ij} - \left( \alpha + \beta x_{ij} + \gamma w_j \right) = u_j + e_{ij}
\]

The difference between actual and expected outcome is decomposed in two parts:

- School-level component (random effect) \( u_j \)
- Student-level component \( e_{ij} \)

The random effect \( u_j \) is the school value added, or effectiveness. It is a residual term → its meaning depends on which covariates are in the model.
Random intercept model: Type A and B effects

\[ Y_{ij} = \alpha + \beta x_{ij} + \gamma w_j + u_j + e_{ij} \]

- Type A effect of school \( j \)
- Type B effect of school \( j \)

Both effects are uniform (same effect for all the students)

- Constant slopes \( \rightarrow \) parallel regression lines
- Unique ranking of the schools
  - ranking on Type A effects to inform potential students
  - ranking on Type B effects for accountability

From uniform to varying school effects

- Uniform effects are often a restrictive assumption
- Typically a given school practice has more or less impact on student learning depending on the kind of student under consideration:
  - **Egalitarian schools** try to reduce the gap in the prior achievement
  - **Competitive schools** tend to boost the initial differences
- In statistical terms: competitive schools have a higher slope on prior achievement

Random slope model

\[ Y_{ij} = (\alpha + u_{0j}) + (\beta + u_{1j}) x_{ij} + e_{ij} \]

- Random intercepts of school \( j \)
- Random slopes of school \( j \)

- Random slopes \( \rightarrow \) crossing regression lines
- Varying effects \( \rightarrow \) different school effects, depending on student characteristics
- No unique ranking of the schools \( \rightarrow \) different rankings conditionally on student characteristics

Uncertainty about the school rankings

To inform school choice we need future rather than past effectiveness \( \rightarrow \) larger error bars \( \rightarrow \) comparisons are even more inconclusive

Models for non-hierarchical structures

- **cross-classified**, e.g. pupils are classified by primary and secondary school

<table>
<thead>
<tr>
<th>Secondary1</th>
<th>Secondary2</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **multiple membership**, e.g. pupils change their school

\[ i \in \{ \text{school A for 4/5, school B for 1/5} \} \]


Achievement progress and measurement error

- Value-added models are based on measures of student achievement usually obtained through standardized tests
- The score of a test is a fallible measure of the true achievement (measurement error depends on reliability)
- The prior score is often used as a covariate in value-added models, causing measurement error bias (attenuation)
  - the school ranking may change: the effect of the prior achievement in not fully controlled for to schools with disadvantaged students are penalized


Ferrao ME, Goldstein H (2009) Adjusting for measurement error in the value added model: evidence from Portugal. Quality and Quantity (forthcoming)

Volumes from Italian research projects on the evaluation of universities


... where the present review is going to appear: ask me a copy at grilli@ds.unifi.it