Quality of Upper Secondary Schools in Finland: Evaluating Persistent and Non-Persistent Differences in Value-Added

Mika Kortelainen*, Heikki Pursiainen, Jenni Pääkkönen

VATT Institute for Economic Research

25 May 2016

Outline

- Motivation
- Value-Added Methodology
- Data
- Results
- Discussion
- Ongoing Research: Elite Schools

Motivation 1/2

- Quality differences of **Finnish upper secondary schools** (below: USS, school) has been a matter of some debate in Finland.
- USS are selective: choose pupils based on comprehensive school success (GPA).
  - Best pupils cluster to "elite" institutions in larger cities.
- Graduation requires passing a nationwide matriculation examination (besides completing the required courses)
  - standardized exams for different subjects: Finnish, mathematics etc.
  - exams have to be passed (roughly speaking) in at least 4 subjects

Motivation 2/2

- Each year different rankings or league tables presented in the media.
  - e.g. Helsingin Sanomat, MTV3, STT
- Various rankings are based on school-level average results in matriculations exams in spring/fall
- Most rankings make no effort to control for the quality of student intake.
  - STT use ad hoc method to control for student quality that has several statistical problems

Summary of the results

- We use a value-added model to evaluate quality differences of USS in 2002-2013 and reliability of existing rankings
- We find significant cross-sectional differences in value-added between the top and bottom schools, but for most schools estimates are statistically indistiguishable.
- The results also imply that the ranking of the schools is highly unstable over time, making any yearly league tables dubious.
- We also find significant differences in value-added between Finnish- and Swedish-speaking schools.
  - Finnish schools outperform Swedish schools
  - This result seems to be mainly explained by lower performance of Swedish schools in Ostrobothnia

The Finnish School System

- Comprehensive school (grades 1-9) is compulsory. According to the PISA, school performance differences in Finland at this level are among the lowest in the world.
- After comprehensive school, around 50 % of students continue to upper secondary school (USS).
- USS are selective, selection based on comprehensive school grade average.
- Most popular schools in large cities are highly selective, only very successful students enter.
  - minimum required GPA might be 9.4 or even higher for some schools

Outcome Variables

- To graduate from an upper secondary school, a student must pass (roughly speaking) in at least four different subjects (e.g. Finnish, mathematics).
- Our main measure of student educational success is based on **the average score of the exams taken**
  - numerical values for grades: laudatur 7, eximia 6, ..., approbatur 2, improbatur 0
- We have also used the grade in English and mother tongue as an alternative outcome variables.
  - these are the most common exams ($>$ 95% students take them)

Valued-Added Methodology

- Our approach is to use a value-added model to estimate differences in school quality.
- Value-added simply means the effect of the school on students' educational success when the initial student quality is controlled for.
- In an ideal value-added model the difference between two schools' value-addeds is **the expected effect** of school on educational success for a group of students randomly assigned to two schools.
- There is now a large literature on school and teacher value-added models (see e.g. Kirabo Jackson et al. 2014 ARE).

Our model 1/2

- Our model is an adaptation of a value-added model from Chetty, Friedman and Rockoff (2014a, AER).
- Chetty et al. use the model for evaluating value-added of teachers in primary schools, while we apply it to upper secondary schools
- The measure of educational success we use is the average of the matriculation exam scores of each student.
  - the main results stay the same if other outcome variables are used

Our model 2/2

- The basic model is of the form

$$A^*_{i,t} = \mu_{j,t} + \beta' X_{i,t} + \varepsilon_{i,t}, \tag{1}$$

  where $A^*_{i,t}$ is the average score for student i in year t, $\mu_{j,t}$ is the school effect of school j (the school i is attending) in year t and $X_{i,t}$ include control variables.

- In our case $X_{i,t}$ include **a third degree polynomial in comprehensive school average**, mother tongue, gender, year dummies and **comprehensive school dummies**.

- Comprehensive school indicators are included for two reasons
  1. the grade scales differ between schools
  2. to limit selection bias generated by unobserved factors

Assumptions (adapted from Chetty et al.)

1. The school effect process $\mu_{j,t}$ for each school j follows the same zero-mean covariance stationary process.
2. The joint distribution of the zero-mean error terms $\varepsilon_{i,t}$ and the school effects is time-homogeneous.
3. The school effect $\mu_{j,t}$ is not correlated with the idiosyncratic shocks $\varepsilon_{i,s}$.
4. The idiosyncratic effects are uncorrelated with each other.

- When these assumptions hold, it can be shown that

$$\hat{\mu}_{j,t} = \hat{E}(\overline{A}_{j,t} \,|\overline{A}_{j,t-1}, ..., \overline{A}_{j,t-s}) = \hat{E}(\mu_{j,t} \,|\overline{A}_{j,t-1}, ..., \overline{A}_{j,t-s})$$

i.e. the value-added estimator is **the best linear estimator** for the school effect $\mu_{j,t}$ given the information on average scores in previous years.

Estimation 1/2

- The basic idea is to estimate first β from model (1) and then construct a residual score $A_{i,t} = A^*_{i,t} - \hat{\beta}' X_{i,t}$ for each student.
- We then construct value-added measures from the mean residualized scores $\overline{A}_{j,t}$.
- The basic idea is
  1. to model the persistence in time of the school effect by looking at how the means change over time (nonparametric time-series model)
  2. to take into account the amount of random variation included in the mean residualized scores by *shrinkage*.
- Shrinkage means that in value-added estimation the means are shrunk towards the overall mean (standardized to be 0)

- Following Chetty et al. (2014a), we use the best linear prediction of the mean residualized score of a school conditional on all previous / other mean residualized scores:

$$\hat{\mu}_{j,t} = \hat{E}(\overline{A}_{j,t} \,|\overline{A}_{j,t-1}, ..., \overline{A}_{j,t-s}\,).$$

- The time-series model allows for value-added estimates to vary over time, but also account for persistence in quality.
- In addition, the model accounts for uncertainty concerning school quality estimates.
    - not only in statistical inference, but also in estimation (shrinkage)

Shrinkage

- In most applications of value-added models it is common to shrink value-added estimates toward a common Bayeasian prior.
- The following formula illustrates empirical shrinkage:

$$\hat{A}^{EB}_{j,t} = \delta_j \cdot \hat{A}_{j,t} + (1 - \delta_j)\overline{A}_t,$$

  where $\hat{A}_{j,t}$ is the unshrunken (or original) estimate, $\overline{A}_t$, is the average school value-added and δ_j and $1 - \delta_j$ are the weights based on uncertainty of $\hat{A}_{j,t}$.
- If school performance is evaluated without taking the uncertainty resulting from small cohorts into account, the best and worst schools each year tend to be the smallest ones.
  - Many of Finnish USS are small, with only one cohort of perhaps 20-30 students or even less graduating each year.
- We apply **the value-added estimator with the shrinkage property**.
  - The estimates of smaller schools are shrunk more than larger ones.

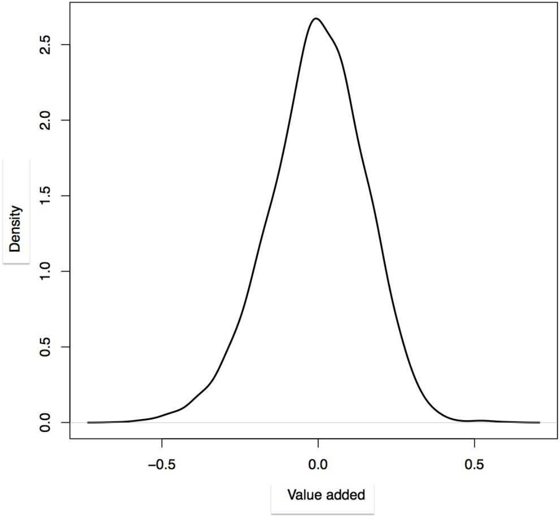
Selection bias

- The value-added calculated can give valid estimates of the “true” school effects $\mu_{j,t}$ only if the errors are not correlated with the school effects.
- This assumption is violated for example if students select into different quality schools based on time-varying unobservable variables.
- Using randomized experiments and quasi-experiments, some previous studies have not found signicant selection bias for typical value-added models.
- In our application, we have used some methods proposed in the literature to evaluate potential bias.
  - The idea is to include more control variables and check whether the results stay robust.

Data

- The data used is mostly combined from two sources:
  - The first source is the database of the Finnish Matriculation Examination Board, consisting of all matriculation exam results from 1990 to 2013.
  - The second source is the National Board of Education database, which has data on (nearly) all upper secondary school applications from 1998 to 2012.
- The final data contains all exams taken in the years 2002-2013, about 386 000 students in total.
- There are around 400 upper secondary schools in Finland.

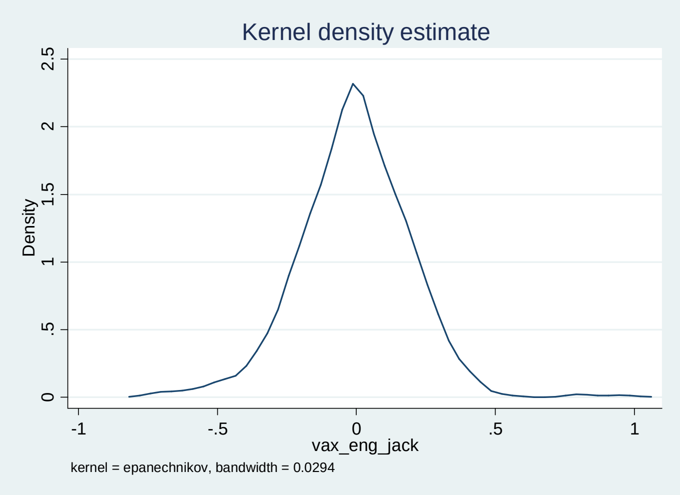
Distribution of value-added (y = average scores of exams)



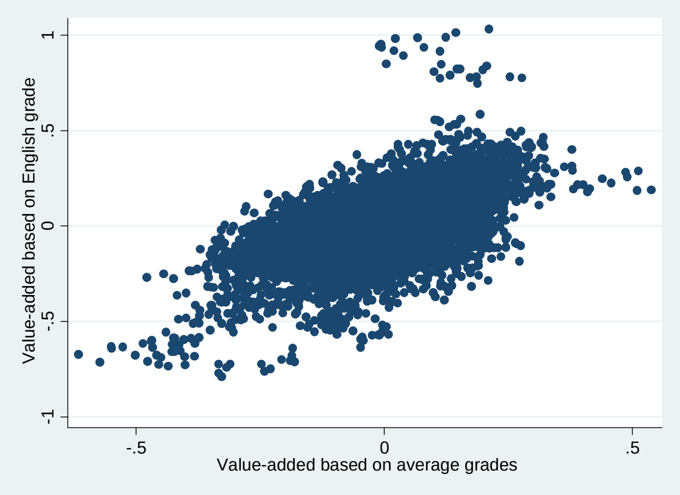
Distribution of value-added (cont.)

- There are considerable differences between top and bottom schools.
  - The difference between these value-added estimates is about 1 grade point.
- A difference of this magnitude would translate into a large advantage in university entry.
  - For example, a person having an average that is one grade point higher would have to score 10-15 % less in entry examinations for the faculties of law or medicine in the University of Helsinki.
- But differences are this big only between the very top and bottom institutions.
  - Middle 80 % of schools are within at most 0.4 grade points of each other.
  - Similar figure for the middle 50 % is only 0.2 grade points.

Distribution of value-added (y = English grade)

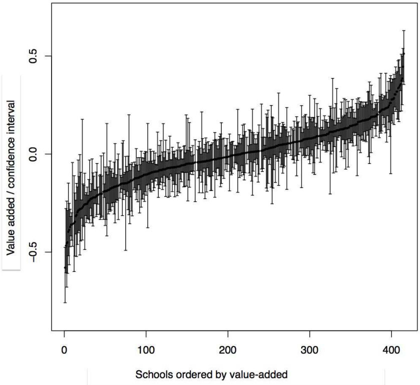


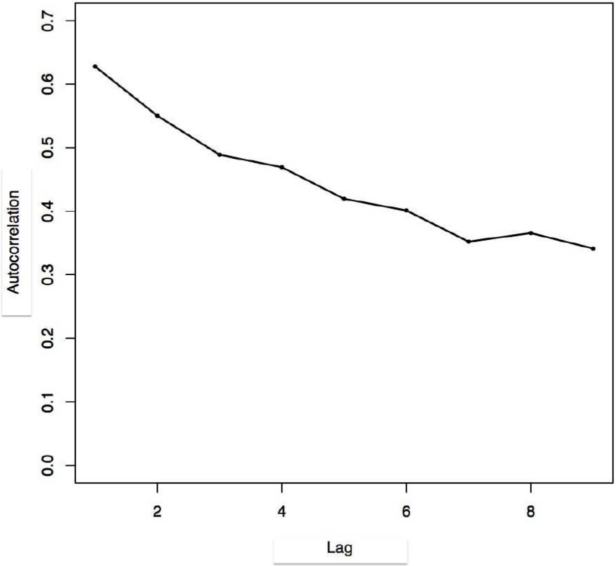
Value-added estimates based on different outcomes



Are differences statistically significant?

- We used block bootstrap to estimate confidence intervals for schools' value-added estimates.
- Most schools in the middle are indistinguishable (in one year)
- For about 50% of schools, we cannot reject the null hypothesis: $H_0: \mu_{j,t} = 0$.

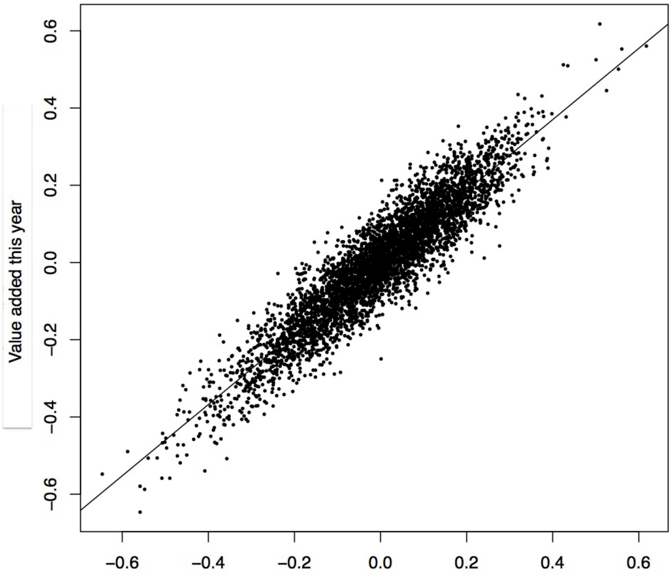



Autocorrelation
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
2
4
6
8
Lag

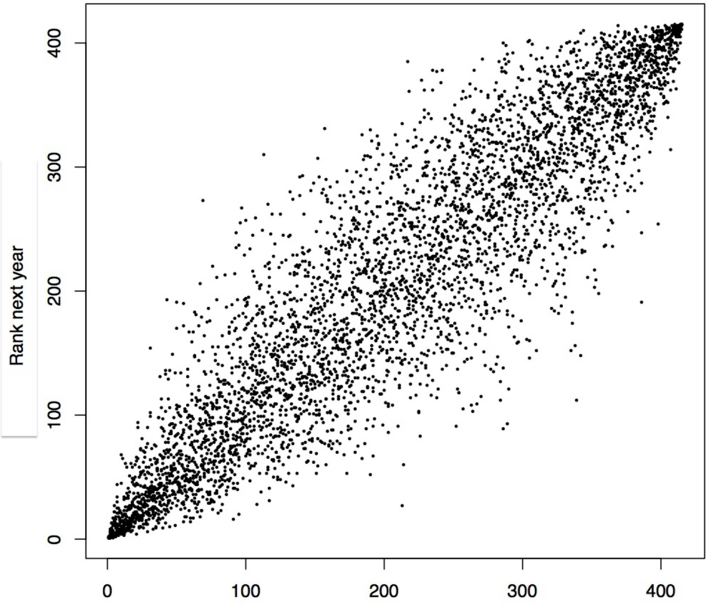
Persistence

- Besides cross-sectional differences, we studied the persistence of value-added estimates using transition matrices and rank correlations.
- The results (not presented here) show that persistence is quite strong for schools in top and bottom 10%, but there is a lot of turnover for other deciles.
  - almost 80 % of the bottom 10 % of schools stay in the bottom in the next year
- Overall, the results show that rankings based on value-added estimates are unstable over time.

Adjacent year value addeds



Adjacent rankings



Rankings are not very stable (Kendall rc matrix)

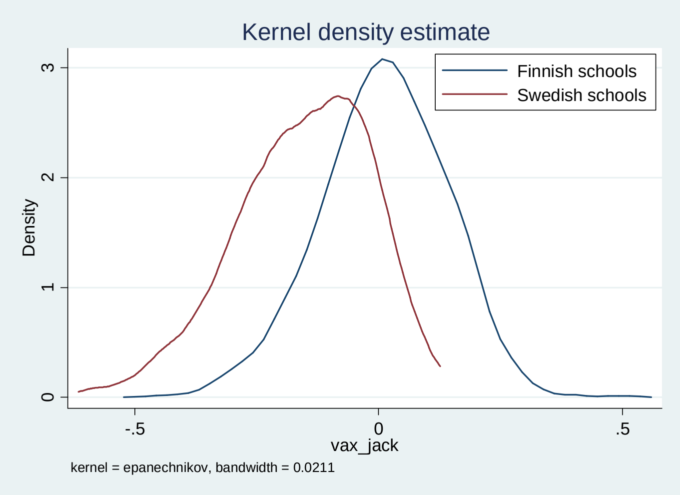
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
2002	1											
2003	0.77	1										
2004	0.68	0.74	1									
2005	0.6	0.66	0.76	1								
2006	0.55	0.58	0.65	0.73	1							
2007	0.5	0.54	0.57	0.62	0.72	1						
2008	0.45	0.47	0.52	0.54	0.62	0.72	1					
2009	0.4	0.43	0.46	0.5	0.53	0.62	0.74	1				
2010	0.4	0.41	0.44	0.45	0.49	0.55	0.64	0.75	1			
2011	0.38	0.4	0.39	0.42	0.44	0.5	0.56	0.68	0.77	1		
2012	0.38	0.38	0.4	0.4	0.43	0.47	0.54	0.6	0.68	0.77	1	
2013	0.41	0.4	0.4	0.41	0.42	0.47	0.51	0.57	0.63	0.7	0.79	1

- To evaluate whether value-added are sensitive to bias, we have also investigated how adding or removing control variables affect the results.

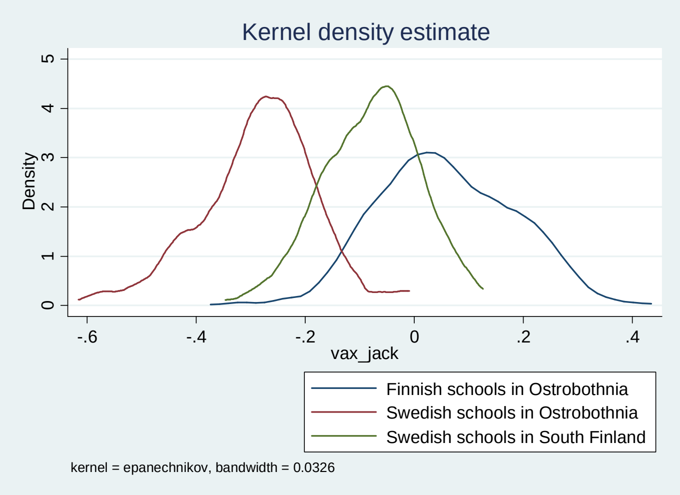
Table 6: Calculations of bias

	Left-out / new variables	Estimate of bias
1	Comprehensive school effects	60.3 %
2	Math / Fin / Swe grades	0.1 %
3	Socioeconomic	0.7 %

Finnish-speaking schools vs. Swedish-speaking schools



Value-added distribution for different groups



Associations of value-added with different variables

- There is no statistically significant association between value-added estimates and various school and municipal variables:

Explanatory variable	Coefficient, simple regression model	s.e. (cluster)	t
Number of students in school	0,000	0,000	1,240
Average expenditure per student	0,000	0,000	-1,210
Teacher-student ratio (2013)	-0,142	0,139	-1,020
Share of qualified teachers (2013)	0,036	0,083	0,440
Number of schools in municipality	0,001	0,001	0,520
Municipal income per capita	0,000	0,000	-1,030
Municipal income tax rate	-0,002	0,006	0,270
State subsidy per capita	0,000	0,000	1,330

Conclusions

- We found significant differences between the schools with highest and lowest value-added estimates.
- However, most schools are in the middle and are statistically indistinguishable.
- Interestingly, we find statistically significant differences in value-added between Finnish- and Swedish-speaking schools.
  - Finnish schools outperform Swedish schools
  - Especially, Swedish schools in Ostrobothnia seem to perform poorly.
- Consistently with previous literature, value-added estimates do not correlate with many observed variables.
  - e.g. per capita expenditure, number of students in school, share of qualified teachers.

Ongoing research: Elite Schools

- In a related project (Kanninen and Kortelainen 2016), we study the causal effects of elite upper secondary schools in Finland on various educational and labor market outcomes.
- We will exploit the entrance thresholds of the Finnish USS application system.
  - Students submit a list of 5 ranked choices usually at the end of their 9th grade of comprehensive schooling.
  - The selection itself is based on the announced preferences and students' grade point averages (GPA).
- Identification based on regression discontinuity design (RDD)
  - compare the students who were just accepted to specific school (the treatment group) to students who applied the same schools but were just rejected

Ongoing research: Elite Schools (cont.)

- We study both short- and long-term effects:
  - ST: matriculation examination scores
  - LT: income, higher education degree, cognitive and non-cognitive test scores
- Preliminary estimates based on RDD are consistent with value-added estimates for elite schools: short-term effects are close to 0.
  - Benefits of attending elite schools seems to be low.