Students' revealed preferences and ranking of school quality

Marcello Sartarelli
European University Institute

January 29, 2009

Motivation

How does a student choose a school?

What is school quality? It is observed by a student but only imperfectly by an econometrician

Governments publish school statistics but very few rankings

A Google search 'school ranking' returns 4 million hits

Newspaper rankings are easily manipulated by schools, Avery et al (2005)

I present a ranking of school quality that pins down between two schools the one that attracts high ability students at the lowest additional cost of quality

Intuition of the model and results

I characterise schooling as a differentiated product market (Hotelling model)

I obtain an indicator that captures the substitutability among average quality of schools and ability of students

I obtain the ranking by sorting the indicators by degree type, e.g. Economics, in ascending order

The ranking says that the best school is the one making the least quality effort to attract high ability students, i.e. the cheapest to finance ceteris paribus

The ranking that I compute using data on Italian college students differs from available rankings

Other countries whose colleges have low fees and non-selective admission: Austria, Belgium, France, Germany, Italy and The Netherlands

Literature

Revealed preferences of students admitted to top US colleges, Avery et al. (2005)

Strategic interaction college-student in a differentiated market, De Fraja and Iossa (2002)

Revealed preferences of students' parents over schooling, Hastings et al. (2006)

School quality and dropouts, Hanushek et al. (2006)

Applied econometrics ranking, Baltagi (1999)

Schooling as a differentiated product market

Students with heterogeneous ability and geographic location demand schooling

Schools are firms that sell a single good, schooling, with 2 dimensions of differentiation

- Horizontal: geographic location
- Vertical: desirability/quality

Equilibrium allocation of students across schools and quality

Model: demand of student i for college j

$$u_{ij} = \alpha_j + \beta_j a_i + \gamma d_{ij} + \delta_j a_i d_{ij} + \tau f_j + \epsilon_{ij} \quad (1)$$

$$x_{ij} = \underbrace{Pr(u_{ij} = \max_k u_{ik})}_{P_{ij}=Pr(y_i=j)} \quad (2)$$

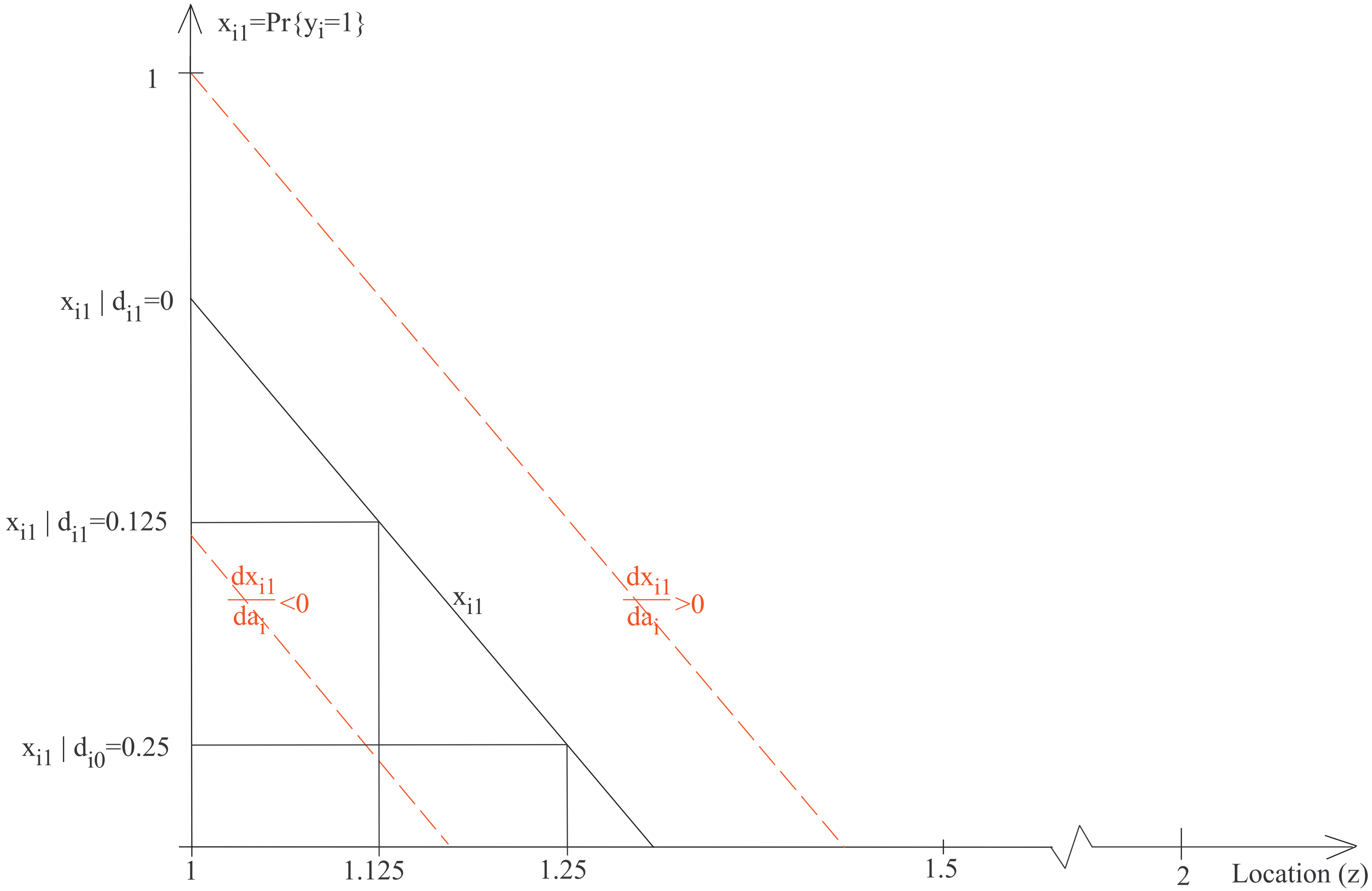
$$= \frac{e^{\alpha_j + \beta_j a_i + \gamma d_{ij} + \delta_j a_i d_{ij} + \tau f_j}}{\sum_k e^{\alpha_k + \beta_k a_i + \gamma d_{ik} + \delta_k a_i d_{ik} + \tau f_k}} \quad (3)$$

Utility u_{ij} depends on average quality α_j of school j, student i's ability a_i, the home-school distance d_{ij} and their interaction

The demand x_{ij} by student i is the probability that i chooses college j

Errors ϵ_{ij} with a logistic distribution yield a multinomial logit model of school choice with individual- and choice-invariant characteristics

Illustration of demand by student i for school 1



Demand x_{i1} for school 1 decreases with home-school distance (continuous line)

A change $\frac{\partial x_{i1}}{\partial a_i}$ in demand from a unit change in ability can be $\lesseqgtr 0$ (dashed line)

Model: supply of school j

$$\max_{\alpha_j} \pi_{ij} = \underbrace{(f_j - \alpha_j)}_{p_j - c_j} \underbrace{P_{ij}}_{x_{ij}} \tag{4}$$

$$\pi'_{ij} = \frac{\partial \pi_{ij}}{\partial \alpha_j} = -P_{ij} + (f_j - \alpha_j) P_{ij}(1 - P_{ij}) = 0 \iff \tag{5}$$

$$\alpha_j^* = f_j - \frac{1}{1 - P_{ij}(\alpha_j^*)} \tag{6}$$

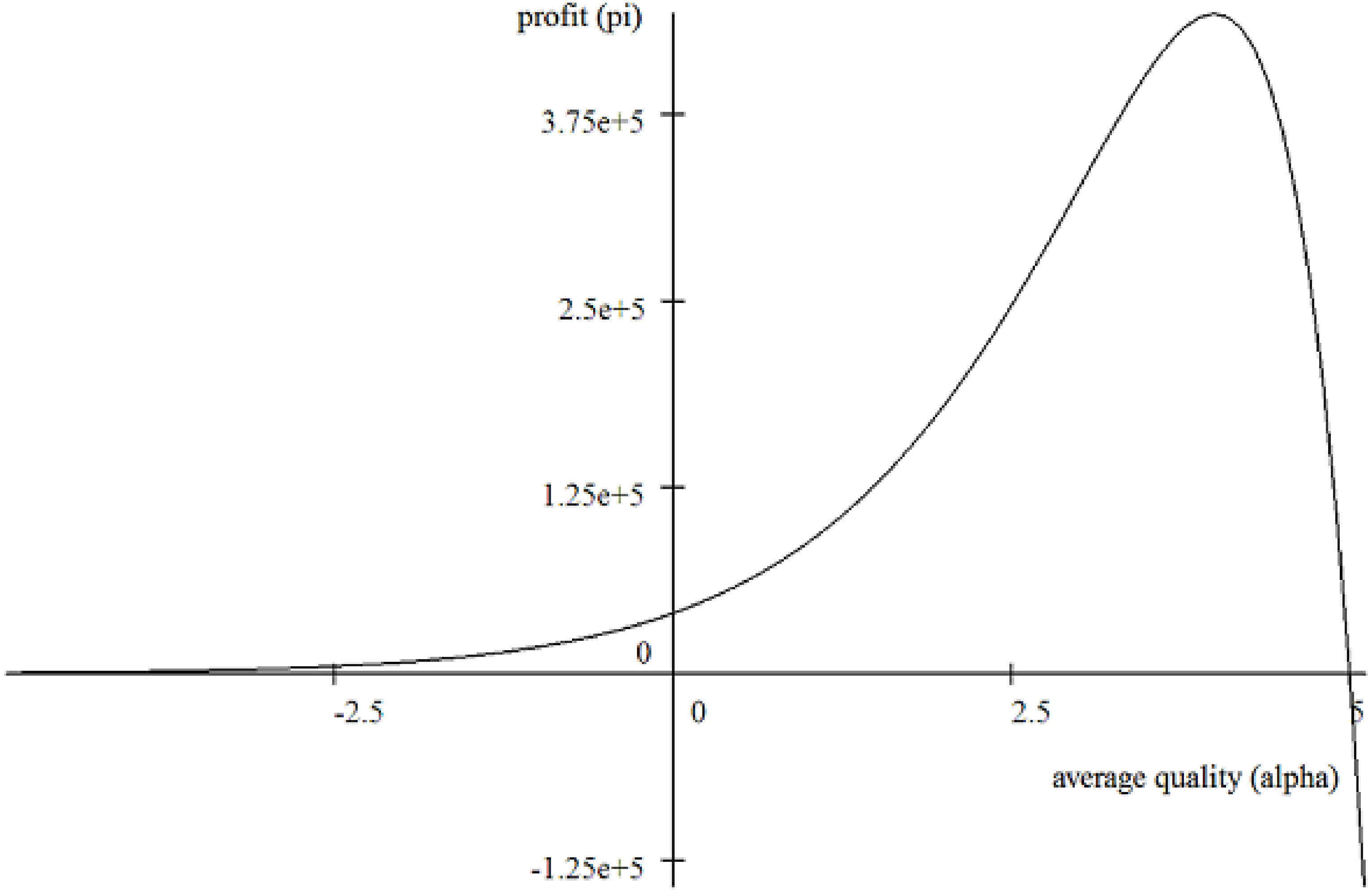
School j chooses average linear cost of quality $c(\alpha_j) = \alpha_j$ to maximise profit π_{ij} per student i

Fees f_j are regulated and admission is not selective

A very low α_j gives school j high profits but low probability that student i chooses it and viceversa

α_j^* is the implicit equilibrium condition of average school quality for school j

Figure 1: **Illustration: profit function and average quality of school j**



The profit function π_j is concave in average quality and α_j^* is a global maximum for reasonable parameter values

Interpretation of equilibrium average quality α_j^*

$$\alpha_j^* = f_j - \frac{1}{1 - P(\alpha_j^*, a_i, d_{ij}, f_j)}$$

The greater the equilibrium average quality α_j^*, the lower ability a_i or home-school distance d_{ij} and viceversa

An increase in α_j^* (similar to inframarginal reasoning under monopoly) can make high ability students "switch" from school j into a different school

Ranking schools by α_j or P_{ij} would only capture the overall effect of average and marginal quality

Change in equilibrium quality as an indicator of school quality

$$\alpha^{*'}_{j_a} = \frac{\partial \alpha^*_j}{\partial a_i} = - \underbrace{\frac{\frac{\partial \hat{P_{ij}}}{\partial a_i}}{\frac{\partial \hat{P_{ij}}}{\partial \alpha_j}}}_{MQ_a} \underbrace{\frac{\left[1 + \alpha^*_j (1 - 2\hat{P_{ij}})\right]}{\left[1 + \frac{1}{\alpha^*_j} + \alpha^*_j (1 - 2\hat{P_{ij}})\right]}}_{AQ}$$

$\alpha^{*'}_{j_a}$ says by how much school j has to change α^*_j to change the average ability of applicants by a unit

Sorting $\alpha^{*'}_{j_a}$ in ascending order ranks as top the school that spends the least in quality to get high ability students, i.e. the cheapest to finance

The ranking can offer a back-of-the-envelope calculation of quality provision in other services with regulated prices such as hospitals

Estimation method

I consider for example students in Economics in all colleges in Italy

I estimate

- the probability $Pr(y_i = j)$ that a student i chooses a degree in Economics in college j
- a conditional logit model with individual- and choice-invariant characteristics

To identify α_js I use as a baseline category a synthetic college that I obtain by drawing

- randomly from students in Economics
- a number of students equal to the mean size of a college in Economics

Dataset of Italian college graduates (Source: Almalaurea)

Variable	Mean	Standard deviation	Min	Max
High school final grade	48.93	7.23	36	60
Male	0.47		0	1
Year of birth	1974	4	1923	1982
Graduation calendar year	2002	2	2000	2003
Home-college distance	159.33	245.58	0	1118.42
Pre-college geographic area of residence				
North	0.57		0	1
Center	0.18		0	1
South	0.25		0	1
High school type				
Gymnasium	0.565		0	1
Teaching	0.036		0	1
Languages	0.044		0	1
Technical	0.302		0	1
Vocational	0.029		0	1
Other	0.024		0	1
Mother's education				
No schooling	0.01		0	1
Primary school	0.209		0	1
Junior high school	0.271		0	1
High school	0.335		0	1
College degree	0.173		0	1

Summary of results: Rank order correlation of rankings

(1)	(2) La Repubblica ranking	(3) Equilibrium ranking $\hat{P}_j$	(4) Distance ranking $\hat{\alpha}_d^{*\prime}$	(5) Ability ranking $\hat{\alpha}_a^{*\prime}$
La Repubblica ranking	1			
Equilibrium ranking $\hat{P}_j$	0.300 *0.069*	1		
Distance ranking $\hat{\alpha}_j^*$	0.279 *0.091*	0.358 *0.030*	1	
Ability ranking $\hat{\alpha}_j^*$	-0.200 *0.229*	-0.142 *0.398*	0.100 *0.559*	1

The correlation between the

- the ranking that I propose in the bottom row and the one by the newspaper La Repubblica are not statistically significant as the *p-values in Italic font* in the table show
- equilibrium ranking and the ability or distance ones are not statistically significant

Results: Economics

(1) College	(2) Equilibrium ranking $\hat{P}$	(3) Equilibrium ranking LR	(4) Equilibrium predicted value $\hat{P}$	(5) Distance (d) ranking $\hat{\alpha_d^{*\prime}}$	(6) Distance (d) ranking LR	(7) Distance (d) MFX	(8) Ability (a) ranking $\hat{\alpha_a^{*\prime}}$	(9) Ability (a) ranking LR	(10) Ability (a) MFX	(11) Observations
Bologna	1	1	0.165	6	1	-0.090*	9	1	0.004*	5462
						(0.050)			(0.003)	
Torino	2	7	0.097	2	7	-0.052*	17	7	0.040	2876
						(0.050)			(0.051)	
Messina	3	16	0.086	17	16	-0.048	16	16	-0.006	2855
						(0.050)			(0.008)	
Parma	4	9	0.069	4	9	-0.045*	15	9	0.011*	2191
						(0.027)			(0.007)	
Firenze	5	12	0.063	16	12	-0.039*	11	12	-0.003*	2040
						(0.036)			(0.003)	
Bari	6	11	0.061	10	11	-0.036***	8	11	-0.003*	2046
						(0.004)			(0.001)	
Siena	7	5	0.058	8	5	-0.038*	7	5	-0.005*	1868
						(0.023)			(0.003)	
Trento	8	2	0.054	1	2	-0.033*	19	2	0.015	1808
						(0.033)			(0.021)	
Modena-Reggio Emilia	9	10	0.050	13	10	-0.033*	3	10	0.012*	1640
						(0.024)			(0.010)	
Catania	10	14	0.047	12	14	-0.028	10	14	0.001*	1563
						(0.040)			(0.001)	
Genova	11	3	0.043	11	3	-0.028*	12	3	-0.023	1310
						(0.028)			(0.028)	
Chieti	12	15	0.042	3	15	-0.024	3	15	-0.002	1172
						(0.039)			(0.005)	
Udine	13	13	0.028	14	13	-0.018	2	13	0.011	897
						(0.026)			(0.019)	

Results: Economics

(1) College	(2) Equilibrium ranking $\hat{P}$	(3) Equilibrium ranking LR	(4) Equilibrium predicted value $\hat{P}$	(5) Distance (d) ranking $\hat{\alpha}_d^{*\prime}$	(6) Distance (d) ranking LR	(7) Distance (d) MFX	(8) Ability (a) ranking $\hat{\alpha}_a^{*\prime}$	(9) Ability (a) ranking LR	(10) Ability (a) MFX	(11) Observations
Bologna	1	1	0.165	6	1	-0.090*	9	1	0.004*	5462
						(0.050)			(0.003)	
Torino	2	7	0.097	2	7	-0.052*	17	7	0.040	2876
						(0.050)			(0.051)	
Messina	3	16	0.086	17	16	-0.048	16	16	-0.006	2855
						(0.050)			(0.008)	
...										
						(0.033)			(0.021)	
Catania	10	14	0.047	12	14	-0.028	10	14	0.001*	1563
						(0.040)			(0.001)	
...										

Messina scores higher than Catania in the equilibrium ranking and viceversa in the ability or distance ones

At the margin high ability students move out of Messina and to Catania

Results: Engineering

(1) College	(2) Equilibrium ranking $\hat{P}$	(3) Equilibrium ranking LR	(4) Equilibrium predicted value $\hat{P}$	(5) Distance (d) ranking $\hat{\alpha_d^{*'}}$	(6) Distance (d) ranking LR	(7) Distance (d) MFX	(8) Ability (a) ranking $\hat{\alpha_a^{*'}}$	(9) Ability (a) ranking LR	(10) Ability (a) MFX	(11) Observations
Torino	1	1	0.224	4	1	-0.091*	14	1	-0.014*	5090
						(0.054)			(0.014)	
Bologna	2	6	0.200	8	6	-0.107**	11	6	-0.001*	4811
						(0.050)			(0.001)	
Padova	3	4	0.155	13	4	-0.074*	13	4	-0.002*	3405
						(0.059)			(0.002)	
...										
Vercelli	13	8	0.012	3	8	-0.009*	8	8	0.0003*	277
						(0.009)			(0.0003)	
...										

Torino scores higher than Vercelli in the equilibrium ranking and viceversa in the ability or distance ones

In Engineering there are fewer departments than in other fields and the three top departments have higher $\hat{P}$ values and lower MFX ones

Results: Law

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
College	Equilibrium			Distance (d)			Ability (a)			Observations
	ranking		predicted	ranking		MFX	ranking		MFX	
	$\hat{P}$	LR	value $\hat{P}$	$\widehat{\alpha_d^{*\prime}}$	LR		$\widehat{\alpha_a^{*\prime}}$	LR		
Bologna	1	2	0.168	10	2	-0.135*	14	2	-0.024*	6062
						(0.086)			(0.015)	
Parma	2	10	0.093	18	10	-0.082*	12	10	-0.002*	3360
						(0.070)			(0.002)	
Bari	3	3	0.089	5	3	-0.048	3	3	0.003	3201
						(0.085)			(0.007)	
...										
Genova	9	1	0.047	16	1	-0.025	18	1	-0.006	1686
						(0.063)			(0.016)	
...										

In Liguria Genova is the only college and has mid-ranking values of the equilibrium ranking and bottom values of distance and ability ones

In Emilia Romagna there are 4 colleges and Bologna scores above in the equilibrium ranking but below in the ability ranking

Conclusions

The ranking offers a methodology to pin down best and worst schools by setting apart average and marginal information contained in the students' revealed preferences

Two preliminary policy recommendations:

- Use in the financing of schools indicators of quality capturing the
  - preferences of students
  - optimising behaviour of schools over financial resources

- Jointly use in ranking schools the
  - equilibrium ranking as a measure of average quality
  - ability ranking as marginal quality to differentially award different schools

Extensions

Model

- Characterise the joint distribution of students' ability and geographic location
- Use observed and unobserved information on school quality
- Obtain optimal weights of marginal and average college quality to inform policy makers' decisions

Estimation

- Mixed models free of IIA assumption
- I/O type of IV exploiting characteristics of neighbouring schools
- Panel dimension of college changes by students and survey data on quality