Demand for secondary school characteristics

Evidence from school choice in Hungary

Carla Haelermans[1], Zoltán Hermann[2], Thomas Wouters[3]

[1]Maastricht University, [2]Insitute of Economics(Budapest) [3]KU Leuven

Maastricht, February 27, 2018

Research questions

- Does school quality and school composition affect students' choices?
- Do parents pay attention to the level of the school, or rather to school quality (learning gains)? Are these preferences heterogeneous along (parental) social/educational lines?
- What are the preferences for school (socio-economic) composition?

Motivation

- Many countries have a system of school choice. School choice gains popularity (e.g. controlled school choice in the US)
- School choice is assumed to increase welfare and even quality (through competition)
- Detrimental effects on equality of opportunity (if disadvantaged students end up in lower quality schools)

Contribution

- Application data (ordinal preference lists instead of realised choices)
  - Not directly affected by schools' decisions
  - Hungarian case: high-stakes decision
  - Rank-ordered logistic regression
- School choice in upper secondary education
- Distinction between school quality and level

Overview

Previous literature

Evidence on determinants of school choice (Hastings et al, Burgess et al) indicates the following determinants:

- Distance to home/work
- School quality
- School denomination (religion/church schools) & pedagogical philosophy
- School composition (ethnic/SES)
- Teachers, school management
- Odds of admission
- Availability of information

Education in Hungary

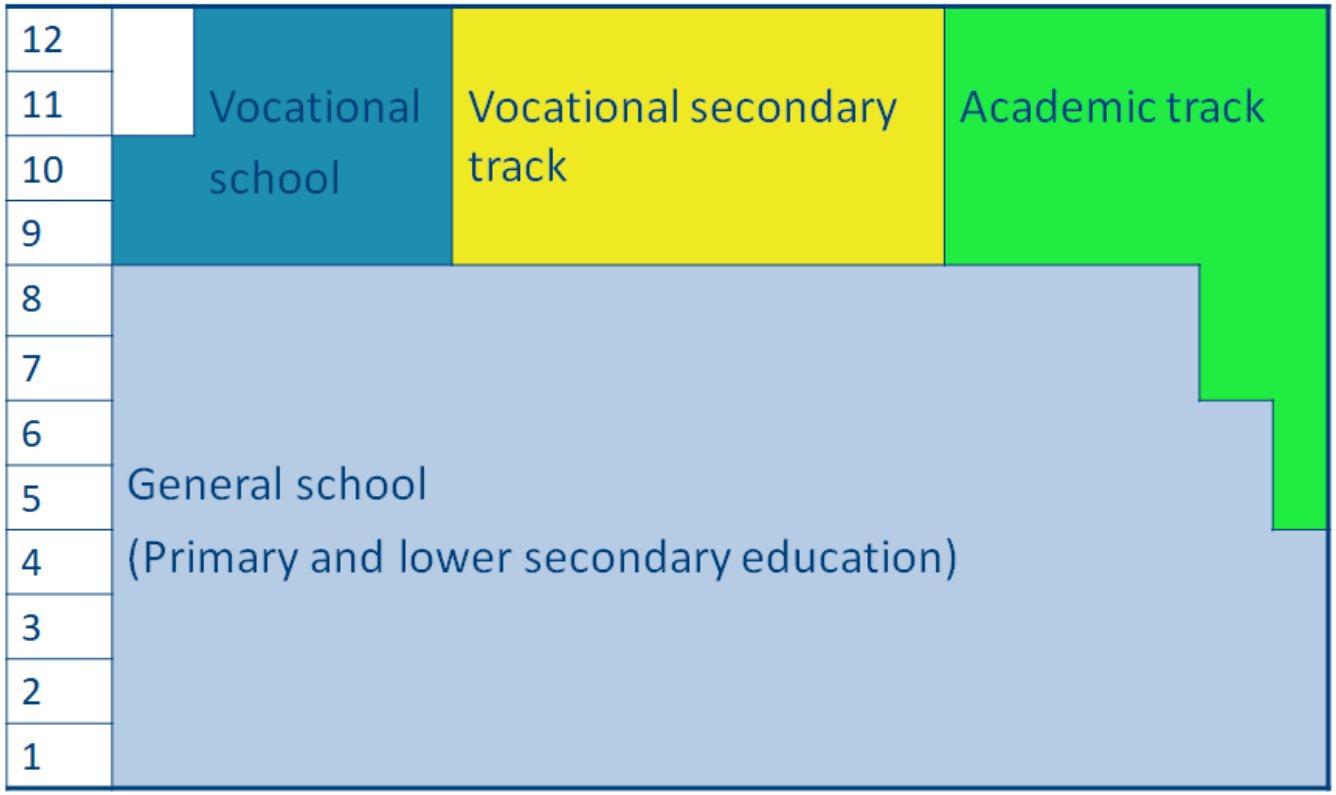

Figure: The Hungarian school system

School admission and enrolment

1. Students apply to programs (we will observe tracks)
   - Choice among educational programs within tracks within schools
   - Students submit a rank-ordered list of preferences
2. Schools rank students
   - Decision on a cutoff entry score
   - Priority for higher test scores
3. Assignment in a centralized manner
   - No incentive problems for students while ranking schools
   - A second round of assignment for those not matched (not observed)

Data

- Administrative data for a single cohort (2006)
- Matched data from three datasets (75% of students)
  - Secondary school application register: rank-ordered lists by students
  - National Assessment of Basic Competences
    - Math and reading test scores
    - Individual characteristics
  - Travel time data (public transport)

Method: utility function

Individuals maximize a simple (linear) utility function of the form:

$$U_{is} = \alpha_i + \beta_1 TravelTime_{is} + \beta_2 SES_i TravelTime_{is} + \gamma_1 Quality_s + \gamma_2 SES_i Quality_s + \delta_1 Level_s + \delta_2 SES_i Level_s + \zeta_1 SchoolSES_s + \zeta_2 SES_i SchoolSES_s + \epsilon_{is} \quad (1)$$

ϵ_{is} is IID according to extreme value type 1 distribution.

Method: rank-ordered logit

- Individuals have strict preference orderings over schools
- e.g. a choice set with 3 schools A, B, C
  - $A \succ_i B \succ_i C$
  - $U_{iA} > U_{iB} > U_{ic}$
- Probability we observe this ordering is $\pi_{iA}^{A,B,C} \pi_{iB}^{B,C}$, with $\pi_{is}^{S} = \frac{e^{V_{is}}}{\sum_{j \in S} e^{V_{is}}}$ This is the probability that school s is chosen from set S.

Measuring school quality

We estimate a value-added model.

$$T_{is}^{10} = \beta_1 T_{is}^{8} + \beta_2 (T_{is}^{8})^2 + \beta_3 (T_{is}^{8})^3 + \delta X_{is} + \theta_s + \epsilon_{is} \tag{2}$$

T: test score in grade 8 or 10
X: gender, SES, parental education, number of books
θ_s: school quality - random effects, shrinkage estimator. We standardise this measure.

Other variables

- School level: school mean of grade 8 scores (mean of math and reading) (standardised)
- School SES composition: share of parents with (at least 1) secondary education degree (mixed or academic track) (0-1)
- Distance: travel time between towns (public transport) (hours)
- High SES: parents with (at least 1) secondary education degree (mixed or academic track) (0/1)
- Test score: grade 8 test score (standardised)

Feasible choice sets

- Students do not rank all schools
- We only consider schools within 90 minutes of travel time
- We also add schools to the choice set
  - For each student, we determine a radius (based on the most distant school)
  - All schools within this radius (which are always within 90 minutes of travel time) are included in the feasible choice set
  - We only consider tracks the student has ranked explicitly

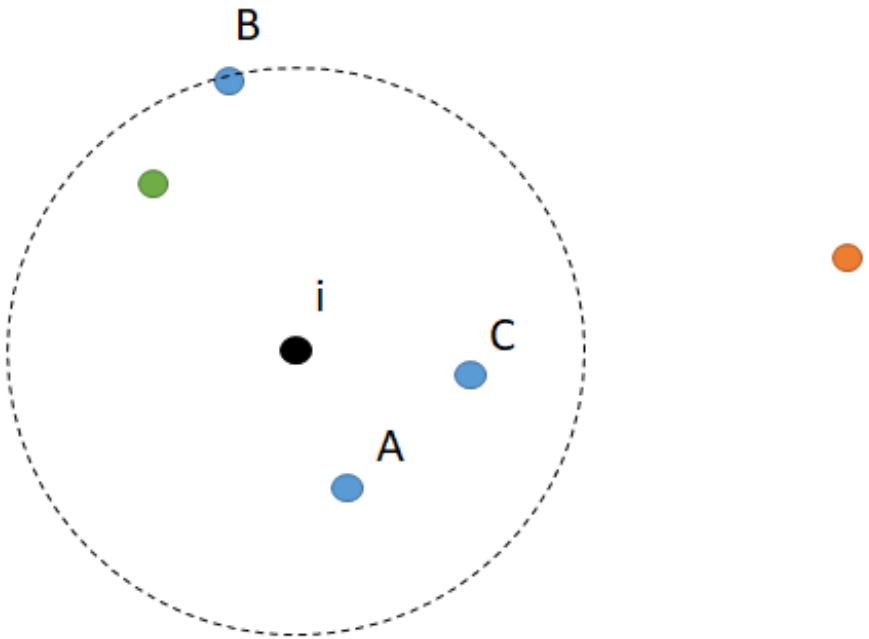

Figure: Adding schools that were not ranked to the choice set

Baseline model

First analysis: including the 4 main variables: travel time, SES composition, school level and school quality.

	All students
	Coeff (SE)
Travel time	-3.136***
	(0.01)
School SES composition	-1.748***
	(0.02)
School level	0.860***
	(0.01)
School quality	-0.136***
	(0.02)
R^2	
N	4240322

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

Table: Basic model including only the 4 main determinants

Baseline model

First analysis: including the 4 main variables: travel time, SES composition, school level and school quality.

	All students	Academic track	Mixed track	Vocational track
	Coeff (SE)			
Travel time	-3.136***	-3.896***	-3.158***	-2.932***
	(0.01)	(0.02)	(0.01)	(0.02)
School SES composition	-1.748***	-1.783***	-1.792***	-1.790***
	(0.02)	(0.05)	(0.05)	(0.06)
School level	0.533***	0.651***	0.689***	0.981***
	(0.01)	(0.01)	(0.01)	(0.02)
School quality	-0.022***	-0.154***	0.087***	-0.050***
	(0.00)	(0.01)	(0.00)	(0.01)
R^2				
N	4240322	1685460	1621676	933186

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

Table: Basic model including only the 4 main determinants

Baseline model with interaction terms

Now including interaction terms: weak homophily
With quadratic terms: higher SES profile is preferred up to some point

	Rank-ordered logit
	Coeff (SE)
Travel time	-2.916***
	(0.01)
High SES × Travel time	-0.506***
	(0.02)
School SES composition	-2.206***
	(0.03)
High SES × School SES composition	2.028***
	(0.02)
School level	0.354***
	(0.01)
School level × Test score	0.663***
	(0.00)
School quality	-0.007**
	(0.00)
School quality × Test score	0.040***
	(0.00)
R^2	
N	4206191

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

Table: Basic model with interaction terms

First vs subsequent choices

The rank-ordered model weighs all choices equally, irrespective of their rank. We now consider the top-n choice for each student. We consider the probability that a school is chosen from the set, excluding higher-ranked alternatives. The rank-ordered logit the reduces to a conditional logit model.
In general, preferences become less outspoken, the lower the rank of the choices we explore.

	First choice	Second choice	Third choice	Fourth choice	Fifth choice
	Coeff (SE)	*Coeff (SE)*	*Coeff (SE)*	*Coeff (SE)*	*Coeff (SE)*
Travel time	-2.983***	-2.857***	-2.914***	-2.926***	-2.911***
	(0.02)	(0.02)	(0.03)	(0.05)	(0.07)
High SES × Travel time	-0.572***	-0.539***	-0.526***	-0.443***	-0.350***
	(0.03)	(0.03)	(0.04)	(0.06)	(0.09)
School SES composition	-1.855***	-2.175***	-2.321***	-2.577***	-2.885***
	(0.05)	(0.05)	(0.06)	(0.09)	(0.14)
High SES × School SES composition	2.359***	2.152***	1.943***	1.787***	1.598***
	(0.04)	(0.05)	(0.05)	(0.07)	(0.12)
School level	0.416***	0.343***	0.316***	0.294***	0.332***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
Test score × School level	0.755***	0.696***	0.609***	0.541***	0.471***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
School quality	0.008	-0.033***	0.001	0.012	-0.011
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Test score × School quality	0.043***	0.031***	0.053***	0.051***	0.037**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
R^2					
N	4206191	4152613	4104868	4071042	4055260

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

Table: Conditional logit on the top-n choice

Endogeneity issues

Omitted variable bias: school composition variables (SES composition and school level), and school quality, may correlate with other (unobserved) school characteristics.

- School level and SES composition: effect may be overestimated
- School quality: less likely to suffer from bias
- Proposed instruments: level and SES composition of students living in the school's town. No suitable instrument for school quality (yet).

Potential solution: control function approach

- Does not require the problem to be linearized
- Other solutions (Berry inversion, BLP) imply working with aggregated data

Control function approach

Two stages

- First stage: regress endogeneous variables on exogeneous ones (including instruments). Save the error terms.

$$Level_s = \alpha Level_s^{instr} + \beta SchoolSES_s^{instr} + \gamma Quality_s + \epsilon_{is} \tag{3}$$

- Second stage: original regression, but including the control function term (i.e. error term from first stage)

Control function approach: results

	Original	CF (level)	CF (SES composition)
	Coeff (SE)		
Travel time	-3.136***	-3.222***	-3.222***
	(0.01)	(0.01)	(0.01)
School SES composition	-1.748***		2.580***
	(0.02)		(0.08)
School level	0.533***	-0.253***	
	(0.01)	(0.01)	
School quality	-0.022***	-0.008***	0.029***
	(0.00)	(0.00)	(0.00)
Town SES composition		1.274***	
		(0.05)	
School level (CF)		0.617***	
		(0.01)	
School SES composition (CF)			-2.342***
			(0.08)
Town level			-0.492***
			(0.02)
R^2			
N	4240322	4246565	4240322

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

Table: Control function approach: results

Conclusion

Conclusions:

- Preferences for school level overshadow preferences for school quality (learning gains)
- School quality is harder to observe than school level and school composition
- Indications for homophily in terms of socio-economic status

Challenges:

- Implementation of control function approach