The long term effects of early educational selection

a quasi-natural policy experiment from Hungary

Klára Gurzó* and Dániel Horn**

*CEU and **MTA KRTK KTI and ELTE
The project is supported by the Hungarian Scientific Research Fund (no. 109338)
Preliminary. Do not cite without permission!

Efficiency in education workshop, London 19-20. September 2014.

Literature

- There is a convincing set of empirical studies that show that early selection associates well with inequality both on the short- and the long-run (e.g. OECD 2005; Brunello and Checchi 2007; Schütz et al. 2008)
- There is evidence on the short run (e.g. Hanushek and Woesmann 2006; Allmendinger 2005; but see Waldinger 2007; or Kerckhoff 1986; Galindo-Rueda and Vignoles 2005; but see Manning and Pischke 2006), as well as the long run (e.g. Meghir and Palme 2005; Pekkarinen, Uusitalo, and Kerr 2009; but see Hall 2012 or Malamud and Pop-Eleches 2010) on (the lack of) causal effects
- However most of these studies looked at general reforms and not the partial effect of age-of-selection (but see van Elk et al. 2011; or Horn 2013; but both for short run effects)
- And most studies utilize a 'comprehensive' rather than a 'selective' reform

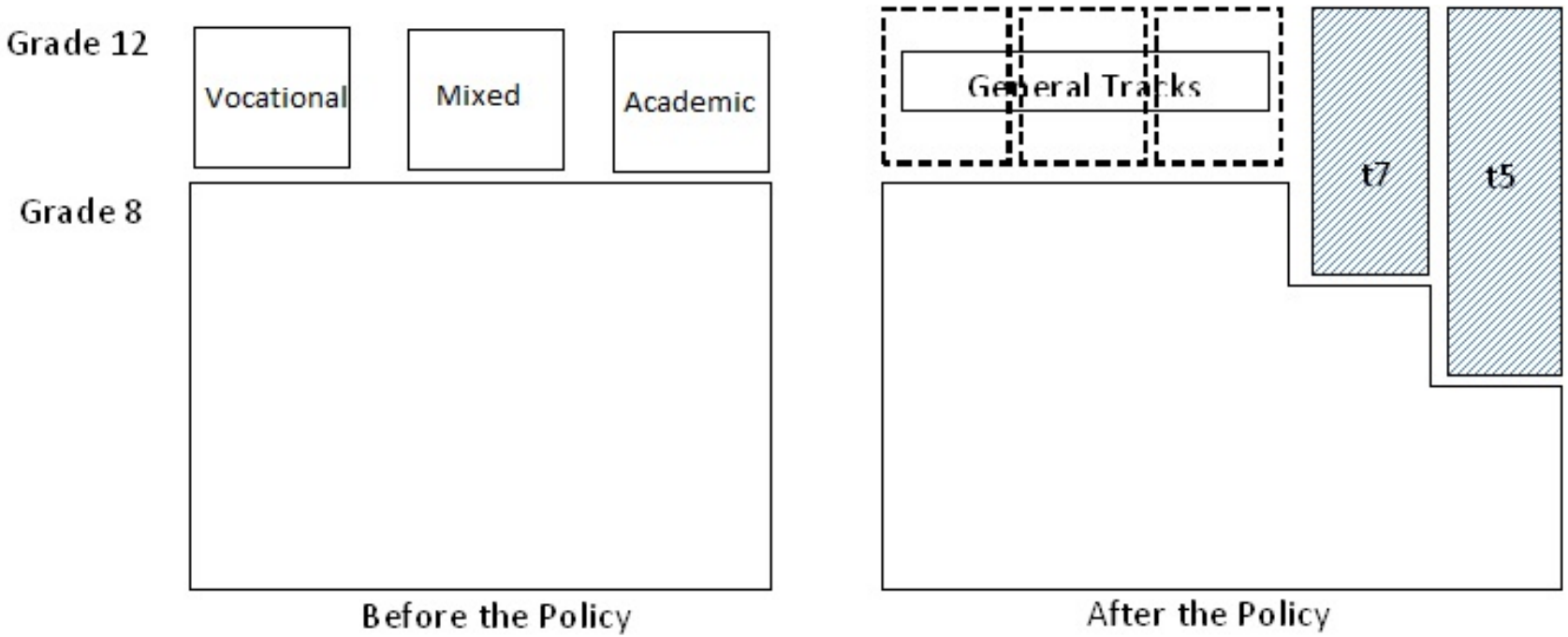
Hungary

- The pre-transition Hungarian education system resembled that of the Soviet system
  - 8 years of general training and 4 years of tracked training: academic, mixed, vocational
  - relatively centralized
- Post-transition the public administration was highly decentralized
  - over 3000 local governments
  - which had to provide public education (among others)
- Dramatic demographic decline (no. 1st grade students 1980: 175.000; 1990: 125.000)
- Increasing pressure on school providers to attract children from outside the catchment area
  - regulation implicitly (1985) and later explicitly (1993) allowed this

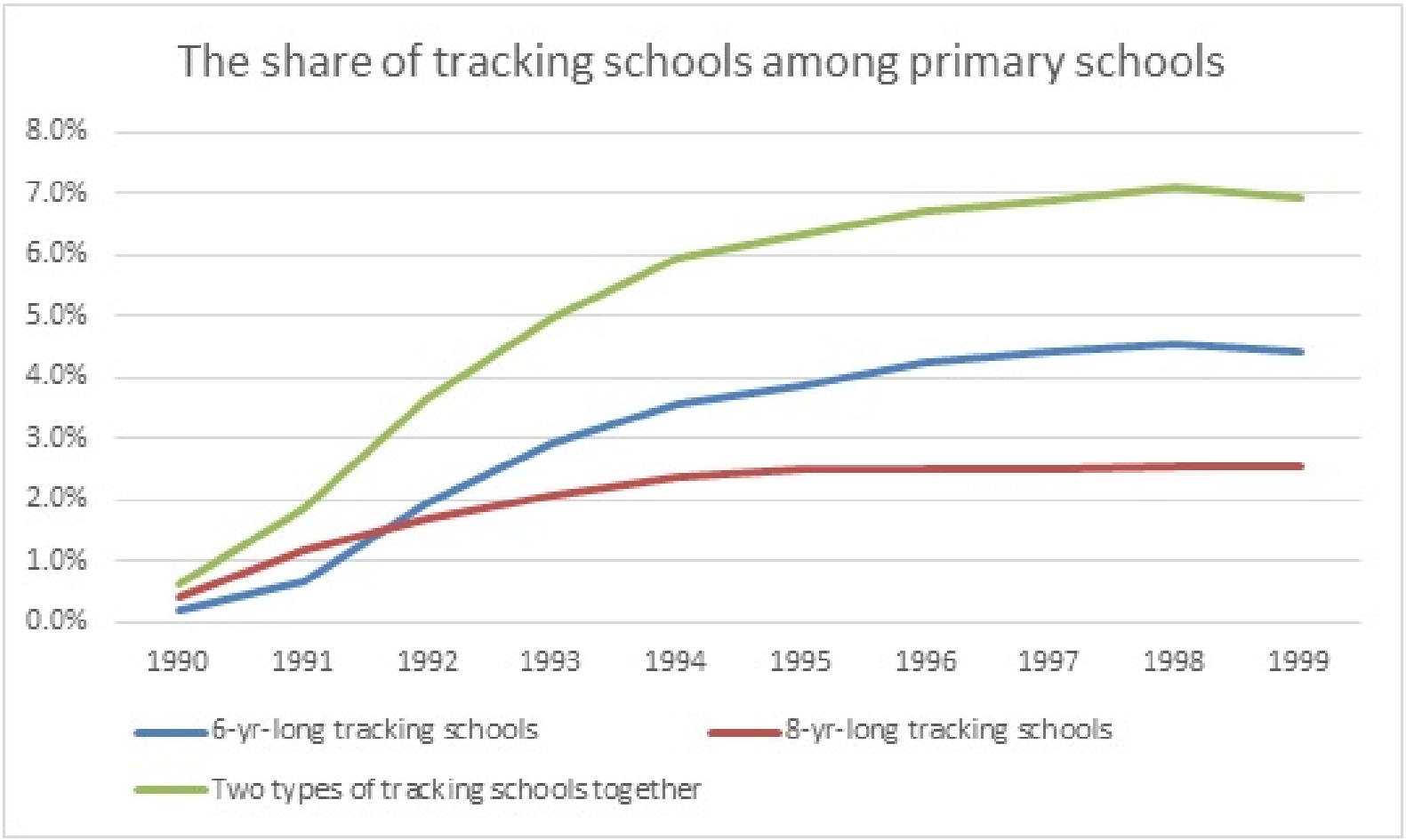
Hungary

- An obvious solution to this pressure was to introduce elite academic tracks as early as possible
- This coincided with the interest of the church(es)
- And was (implicitly) supported by two main parties in the parliament (Christian Democrats/Conservatives and Liberals)
- Two types of early-selective academic tracks were introduced
  - an 8-year-long, which selects at age 10 before grade 5 (t5)
  - a 6-year-long, which selects at age 12, before grade 7 (t7)

Institutional change



Institutional change



Early-selective tracks today

Datasets

- Hungarian Census 2011
  - full population (cca. 10m)
  - place of birth and place of resindece (current and previous)
  - education path (restricted)
  - labor market status
  - NO wage
- Hungarian Wage Survey 2011
  - over 200.000 observations
  - full public and sample of private sphere (over 5 employed)
  - ISCO, place of work and gender
  - imputed wage $\approx$ occupational prestige
- Own School Survey
  - In which year has the settlement opened (or closed) an early-selective track

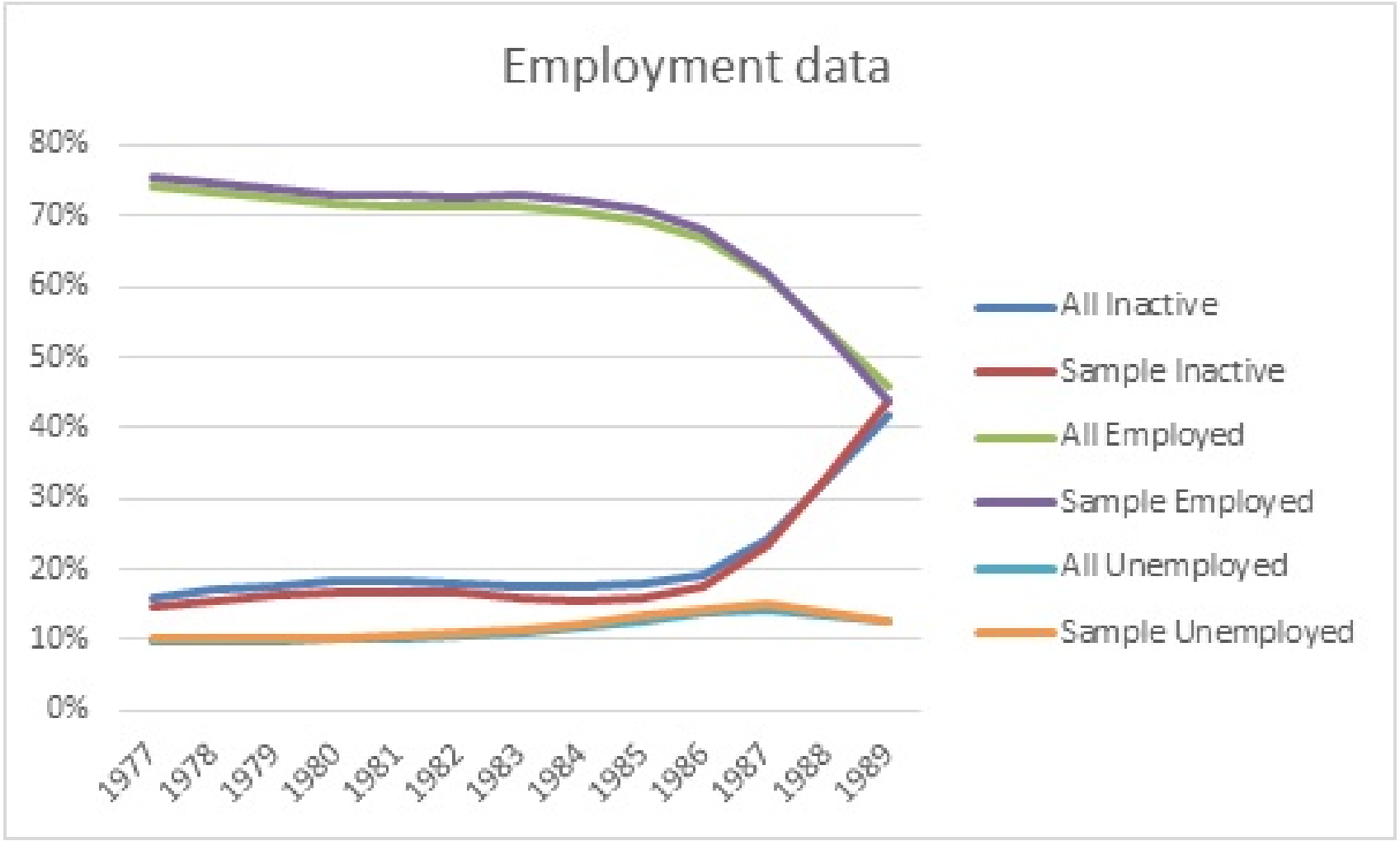
Sample

- born between May 31, 1976. and May 31, 1989.
- for whom the place of residence between ages 10 and 18 could be identified
- who lived in a settlement, which ever established an early selective track

Years	Non-Treated	Tracking From Grade 5	Tracking From Grade 7	Both	Total
1977	48018	0	289	0	48307
1978	42270	0	3529	0	45799
1979	31080	0	12325	0	43405
1980	16735	3071	20931	1311	42048
1981	8886	5458	19837	5763	39944
1982	4445	7152	18132	7621	37350
1983	2355	7435	16877	8756	35423
1984	1314	7297	14402	11319	34332
1985	844	6870	14931	12371	35016
1986	1175	6794	14726	13791	36486
1987	2808	7182	12675	13860	36525
1988	15432	22097	0	0	37529
1989	15346	21652	0	0	36998
Total	190708	95008	148654	74792	509162
Fraction	0.37	0.19	0.29	0.15	1

MTA KRTK KTI

Sample


Employment data
80%
70%
60%
50%
40%
30%
20%
10%
0%
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
All Inactive
Sample Inactive
All Employed
Sample Employed
All Unemployed
Sample Unemployed

Difference-in-difference

- Reduced-form estimates:

$$y_{ics} = \alpha + \beta t5_{cs} + \eta_s + \gamma_c + \kappa X_{ics} + \epsilon_{ics} \quad (1)$$

$$y_{ics} = \alpha + \beta t7_{cs} + \eta_s + \gamma_c + \kappa X_{ics} + \epsilon_{ics} \quad (2)$$

$$y_{ics} = \alpha + \beta t5_{cs} + \gamma t7_{cm} + \eta_s + \gamma_c + \kappa X_{ics} + \epsilon_{ics} \quad (3)$$

  - where y is the outcome (employment, unemployment, estimated wage and tertiary enrollment)
  - i - individual, c - cohort, s - settlement;
  - t5 and t7 are the two early-selective track dummies
  - X is a vector of individual characteristics (gender, year of birth, "R is a student", "R is on maternity leave")
  - η is the settlement, γ is the cohort fixed effect
  - α β and κ are parameters and ϵ individual error
  - Note: this model is a reduced form of an IV model, where individual participation in an early-selective track is instrumented by t5 or t7. (IV models tested, substantial results do not change)

Event study

- Event study estimates:

$$y_{ics} = \alpha + \beta_1 t5_{cs}^{<-2} + \beta_2 t5_{cs}^{-2} + \beta_3 t5_{cs}^{-1} + \beta_4 t5_{cs}^{0} + \\ + \beta_5 t5_{cs}^{1} + \beta_6 t5_{cs}^{2} + \beta_7 t5_{cs}^{>2} \dots \quad (4)$$

$$y_{ics} = \alpha + \beta_1 t7_{cs}^{<-2} + \beta_2 t7_{cs}^{-2} + \beta_3 t7_{cs}^{-1} + \beta_4 t7_{cs}^{0} + \\ + \beta_5 t7_{cs}^{1} + \beta_6 t7_{cs}^{2} + \beta_7 t7_{cs}^{>2} + \dots \quad (5)$$

$$y_{ics} = \alpha + \beta_1 t5_{cs}^{<-2} + \beta_2 t5_{cs}^{-2} + \beta_3 t5_{cs}^{-1} + \beta_4 t5_{cs}^{0} + \\ + \beta_5 t5_{cs}^{1} + \beta_6 t5_{cs}^{2} + \beta_7 t7_{cs}^{>2} + \\ + \beta_8 t7_{cs}^{<-2} + \beta_9 t7_{cs}^{-2} + \beta_{10} t7_{cs}^{-1} + \beta_{11} t7_{cs}^{0} + \\ + \beta_{12} t7_{cs}^{1} + \beta_{13} t7_{cs}^{2} + \beta_{14} t7_{cs}^{<2} \dots \quad (6)$$

- where superscripts indicate years to/from the establishment of the early-selective track on a given settlement for a given cohort

Full sample

	unemployed			employed		
VARIABLES	both	t7 only	t5 only	both	t7 only	t5 only
t5	0.00234		0.00292	-0.00243		-0.00244
	(0.00219)		(0.00263)	(0.00181)		(0.00206)
t7	-0.00245	0.000560		0.000896	0.000391	
	(0.00239)	(0.00271)		(0.00183)	(0.00223)	
gender	-0.00985***	-0.0118***	-0.00908***	0.00969***	0.0112***	0.00919***
	(0.00223)	(0.00266)	(0.00254)	(0.00186)	(0.00221)	(0.00213)
Constant	0.140***	0.144***	0.134***	0.858***	0.858***	0.863***
	(0.0161)	(0.0201)	(0.0177)	(0.0124)	(0.0152)	(0.0138)
Observations	408,405	244,365	330,815	509,274	303,797	413,522
R-squared	0.009	0.008	0.009	0.534	0.543	0.534
No. of settlements	139	61	106	139	61	106

Robust settlement*cohort level clustered standard errors in parentheses;

*** p<0.01, ** p<0.05, * p<0.1

settlement and cohort FE, year of birth, "is in school", "is on maternity leave" are controlled

Full sample

	ln(wage)			tertiary		
VARIABLES	both	t7 only	t5 only	both	t7 only	t5 only
t5	0.00125		0.00238	0.00167		0.00269
	(0.00276)		(0.00268)	(0.00379)		(0.00382)
t7	-0.00174	-0.000872		-0.00518	-0.00402	
	(0.00338)	(0.00336)		(0.00441)	(0.00483)	
gender	0.0449***	0.0544***	0.0447***	-0.149***	-0.147***	-0.149***
	(0.00539)	(0.00780)	(0.00603)	(0.00251)	(0.00343)	(0.00286)
Constant	11.84***	11.86***	11.84***	0.336***	0.350***	0.337***
	(0.0149)	(0.0188)	(0.0157)	(0.0165)	(0.0182)	(0.0182)
Observations	450,927	269,582	365,414	509,274	303,797	413,522
R-squared	0.024	0.025	0.025	0.084	0.081	0.084
No. of settlements	139	61	106	139	61	106

Robust settlement*cohort level clustered standard errors in parentheses;

*** p<0.01, ** p<0.05, * p<0.1

settlement and cohort FE, year of birth, "is in school", "is on maternity leave" are controlled

MTA KRTK KTI

Problems

One thing is clear: Pareto-efficiency is unlikely! But:

- Take-up is too small! Why do we expect any effect from 5-8% of the population?
- Reduced form shows average effects. What we see is that winners win and losers lose.

Academic students only

- higher take-up
- but selection on RHS

	unemployed			employed		
VARIABLES	both	t7 only	t5 only	both	t7 only	t5 only
t5	0.000954		-0.000940	-0.00127		0.000333
	(0.00340)		(0.00389)	(0.00281)		(0.00314)
t7	-0.00239	-0.000137		0.00185	0.00138	
	(0.00348)	(0.00373)		(0.00264)	(0.00271)	
gender	-0.0135***	-0.0145***	-0.0137***	0.0116***	0.0123***	0.0119***
	(0.00229)	(0.00252)	(0.00264)	(0.00191)	(0.00214)	(0.00219)
Constant	0.102***	0.118***	0.101***	0.889***	0.881***	0.891***
	(0.0237)	(0.0293)	(0.0259)	(0.0172)	(0.0205)	(0.0187)
Observations	140,080	83,677	114,951	181,681	108,633	149,428
R-squared	0.015	0.015	0.016	0.623	0.627	0.623
No. of settlements	139	61	106	139	61	106

Robust settlement*cohort level clustered standard errors in parentheses;

*** p<0.01, ** p<0.05, * p<0.1

settlement and cohort FE, year of birth, "is in school", "is on maternity leave" are controlled

Academic students only

- higher take-up
- but selection on RHS

	ln(wage)			tertiary		
VARIABLES	both	t7 only	t5 only	both	t7 only	t5 only
t5	0.00444		0.00490	0.000378		-0.000272
	(0.00419)		(0.00460)	(0.00521)		(0.00621)
t7	0.00662	0.00344		-0.00270	8.62e-05	
	(0.00637)	(0.00542)		(0.00767)	(0.00776)	
gender	0.0989***	0.107***	0.100***	-0.0841***	-0.0835***	-0.0802***
	(0.00519)	(0.00728)	(0.00589)	(0.00440)	(0.00574)	(0.00466)
Constant	11.95***	11.96***	11.95***	0.553***	0.586***	0.556***
	(0.0277)	(0.0355)	(0.0288)	(0.0273)	(0.0305)	(0.0309)
Observations	154,066	92,050	126,437	181,681	108,633	149,428
R-squared	0.042	0.044	0.043	0.043	0.041	0.042
No. of settlements	139	61	106	139	61	106

Robust settlement*cohort level clustered standard errors in parentheses;

*** p<0.01, ** p<0.05, * p<0.1

settlement and cohort FE, year of birth, "is in school", "is on maternity leave" are controlled

Full cohort - wage above 95th percentile

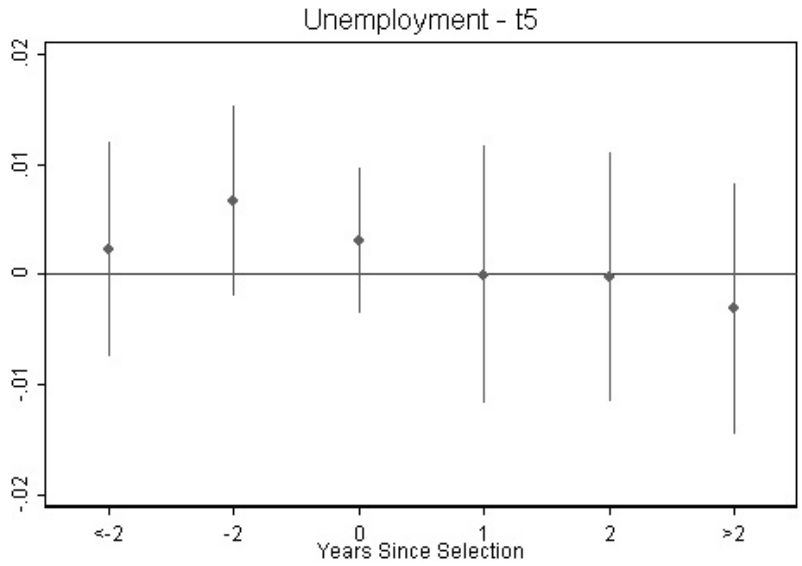
- selection on LHS
- wage ≈ occupational prestige ⇒ wage95 ≈ highest prestige occupations
- where early-selection should really matter

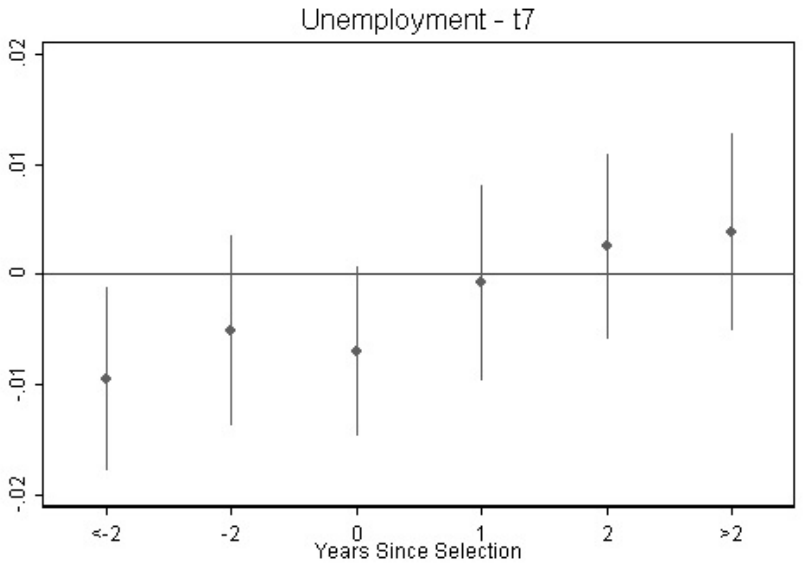
	wage95		
VARIABLES	both	t7 only	t5 only
t5	-0.00193		-0.00477**
	(0.00139)		(0.00238)
t7	-0.00265*	-0.00029	
	(0.000155)	(0.00131)	
gender	0.0281**	0.0268**	0.0337***
	(0.00318)	(0.00367)	(0.00462)
Constant	0.0201**	0.0220	0.0172
	(0.00811)	(0.0075)	(0.0111)
Observations	450,927	269,582	365,414
R-squared			
No. of settlements	139	61	106

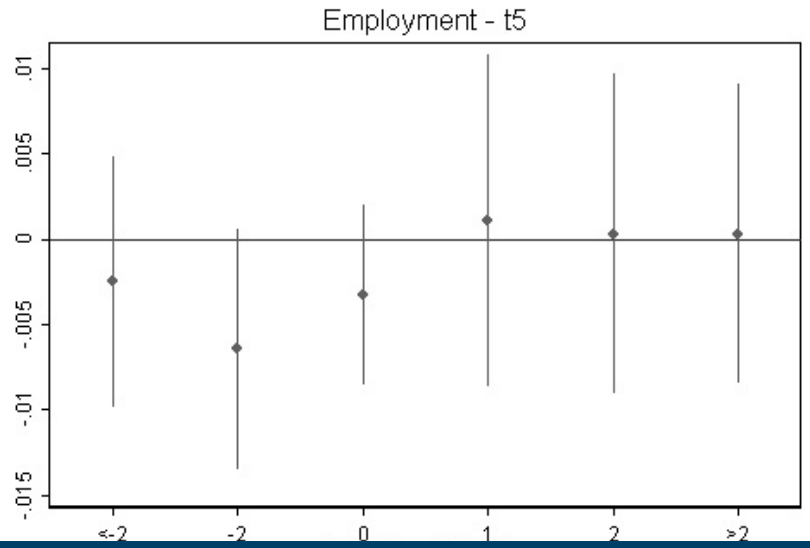
Robust settlement*cohort level clustered standard errors in parentheses;
*** p<0.01, ** p<0.05, * p<0.1

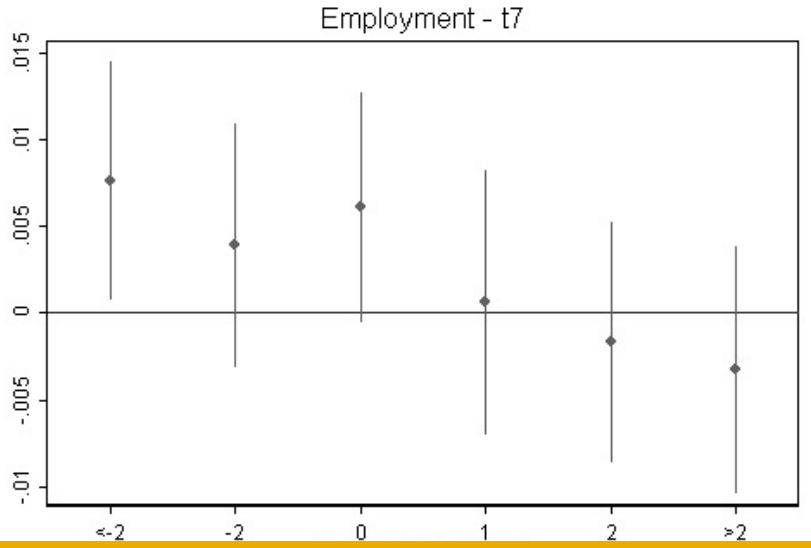
settlement and cohort FE, year of birth, "is in school", "is on maternity leave" are controlled

Event study

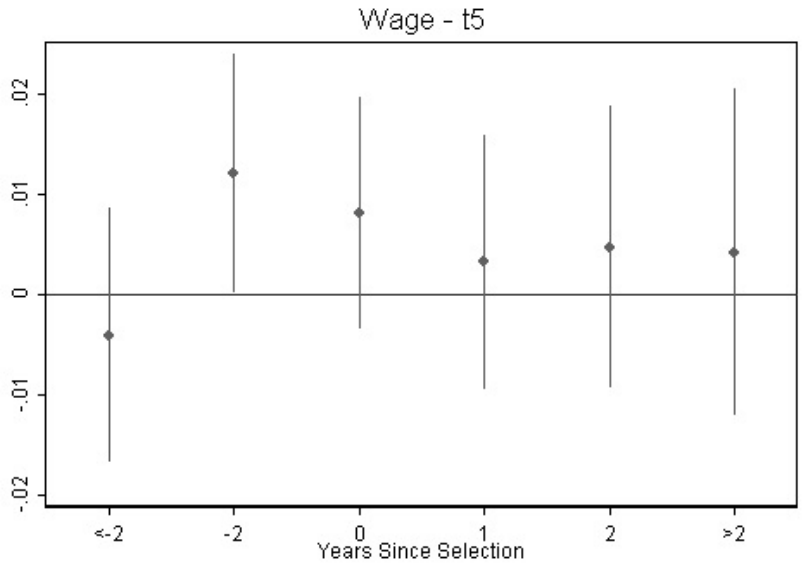

Unemployment - t5
.02
.01
0
-.01
-.02
<-2
-2
0
1
2
>2
Years Since Selection

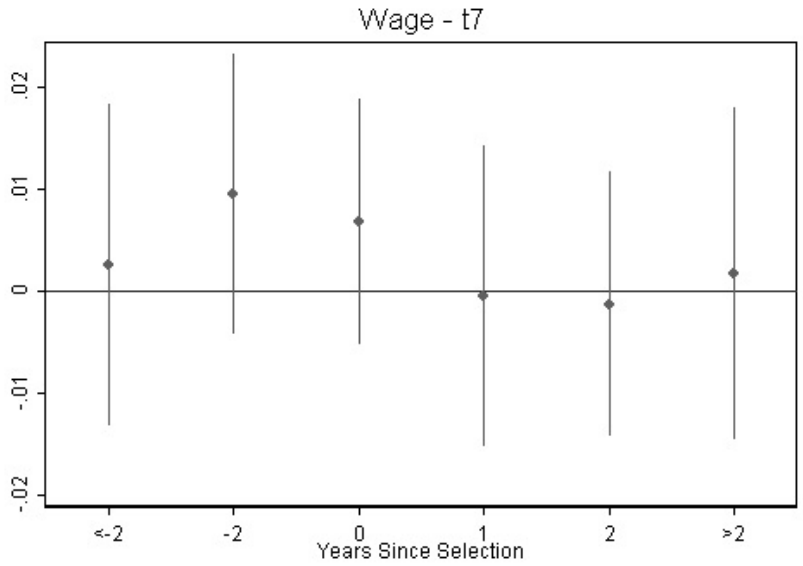

Unemployment - t7
.02
.01
0
-.01
-.02
<-2
-2
0
1
2
>2
Years Since Selection

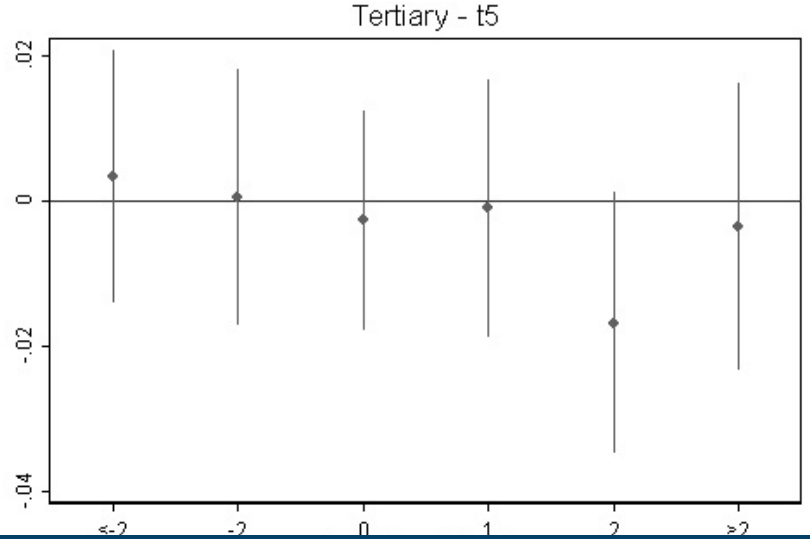

Employment - t5
.01
.005
0
-.005
-.01
-.015
<-2
-2
0
1
2
>2

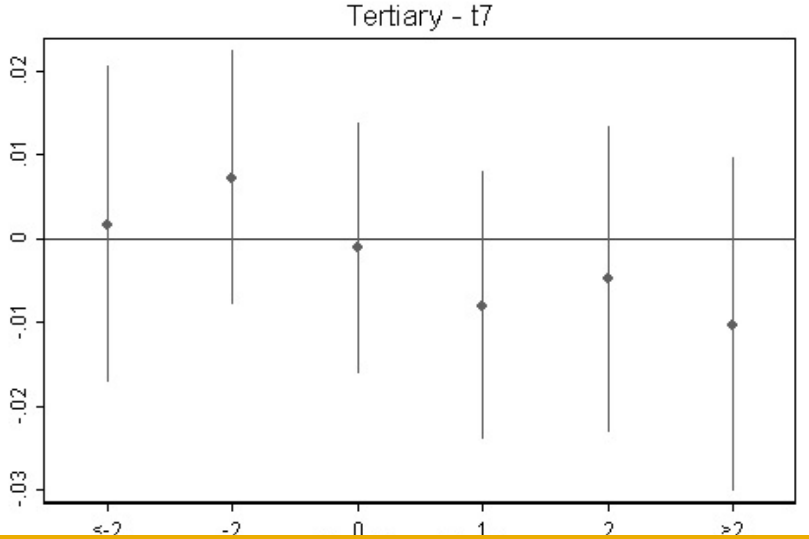

Employment - t7
.015
.01
.005
0
-.005
-.01
<-2
-2
0
1
2
>2

Event study


Wage - t5
.02
.01
0
-.01
-.02
<-2
-2
0
1
2
>2
Years Since Selection


Wage - t7
.02
.01
0
-.01
-.02
<-2
-2
0
1
2
>2
Years Since Selection


Tertiary - t5
.02
0
-.02
-.04


Tertiary - t7
.02
.01
0
-.01
-.02
-.03

Summary, thoughts and further tests

- In sum: we do not see any effect of early selection on labor market and tertiary enrollment
- Answers to the potential critique:
  1. No effects in the academic subsample. (Take-up there is much larger. But potential composition problems.)
  2. No difference (or negative difference) at the upper end of the wage/occupational prestige distribution.
  3. And most importantly: no difference (or negative difference) between t7 and t5? Two years of additional tracking should at least turn the coefficients to the right direction.
- To do:
  1. Try to incorporate take-up rates. Include average take-up by cohort and settlement instead of dummies.
  2. Try other LHS cutoffs: e.g. those, who finished tertiary

Summary, thoughts and further tests

- But what if results are really robust? How do we explain this in view of the numerous counter-evidence?
- Answer: Although, it might be that these early-selective tracks increase test scores (cf. Horn 2013), but (un)employment, wage or even tertiary enrollment depends on various other factors.
  - Tracking might not have an effect (or may have a negative effect) on some of these.
  - Short-run test scores might capture a little effect, but 2 or 4 years of separated education has just no real effect a decade later.
- Also, most other studies looked at general reforms that included other systemic change (curricular, teacher training, more resources), this reform was purely redistributive. It captures the pure effect of segregation on labor market outcomes. And there is none!

Thank you for any comments!

horn.daniel@krtk.mta.hu