CES lecture

Intergenerational Mobility

Session 2

Jan Stuhler

July 8, 2020

Introduction

Research frontier in economics is moving towards administrative data sources.

Applies likewise to intergenerational research (e.g., compare recent articles in "top" journals). Reasons:

- The scale, "completeness", and "linkability" of administrative data opens the door for new research designs (as we illustrate in today's and tomorrow's topics):

1. Descriptive studies that describe intergenerational mobility more thoroughly and/or from new perspectives
2. Causal research designs to identify determinants of (lack of) intergenerational mobility

Content

1. Intergenerational Mobility: Theory and Measurement
2. The Geography of Intergenerational Mobility
   - 2.1 Descriptive area design
   - 2.2 Causal event study design
   - 2.3 Causal "mover" design
3. Multigenerational Mobility

The geography of intergenerational mobility

With availability of large-scale administrative data sources, can study intergenerational mobility in much finer units/areas. Examples:

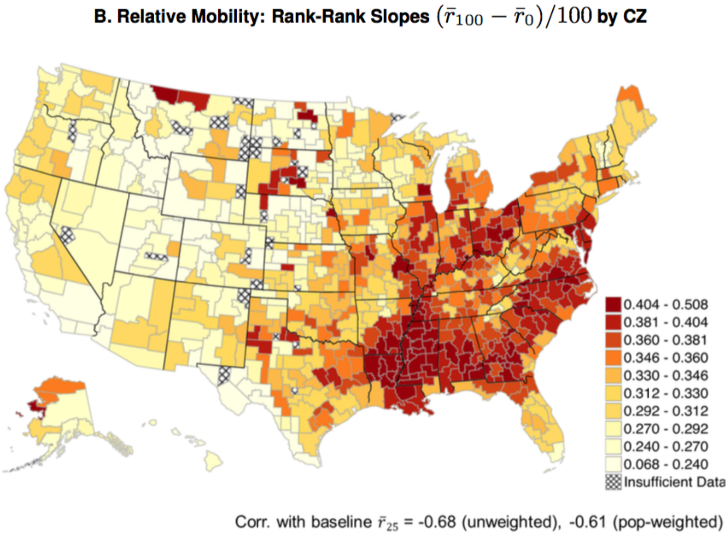
Descriptive

- Chetty, Hendren, Kline and Saez (2014),"Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." Quarterly Journal of Economics

Causal

- Chetty and Hendren (2018a) "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects." QJE
- Chetty and Hendren (2018b) "The Impact of Neighborhoods on Intergenerational Mobility II: County-Level Estimates." QJE

Figure: Chetty et al. (2014)



Intergenerational Mobility in the US: Chetty et al (2014)

Chetty, Hendren, Kline and Saez (2014):

- Use tax data from the US Internal Revenue Service (IRS), match records of parents and children to study intergenerational mobility in the U.S.
- Core sample of nearly 10 million children born between 1980 and 1982 (14- to 16-year-olds), tracked until age 30.

Income definitions:

- Parent's income: average total family income in 1996-2000. Children's income: measured over two years, 2011-2012

Measuring intergenerational dependence

- Intergenerational mobility is often summarized by the *intergenerational elasticity* of income (IGE), defined as the slope coefficient in the regression of log incomes of offspring y^* on log income of parents x^*,

$$y^* = \alpha + \beta x^* + \varepsilon$$

- More recent research often considers income ranks instead of log incomes ($\rightarrow$ rank-rank regression or rank correlation)

Figure: Mean Child Income Rank vs Parent Income Rank in the US

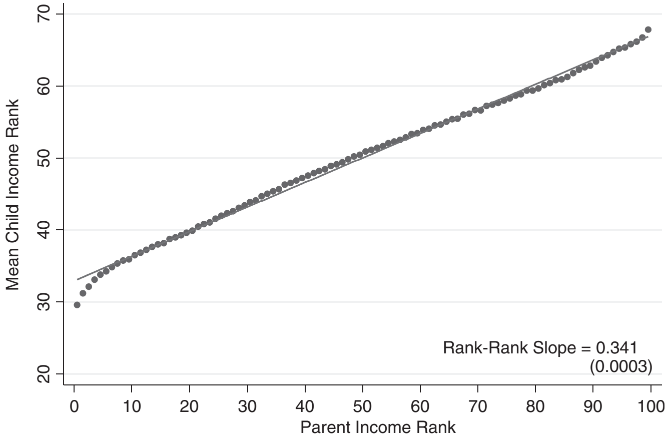
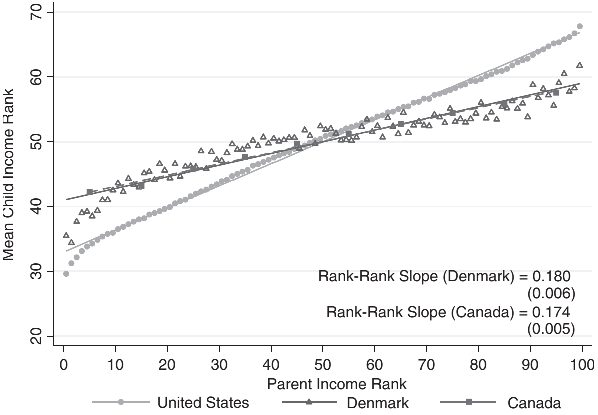
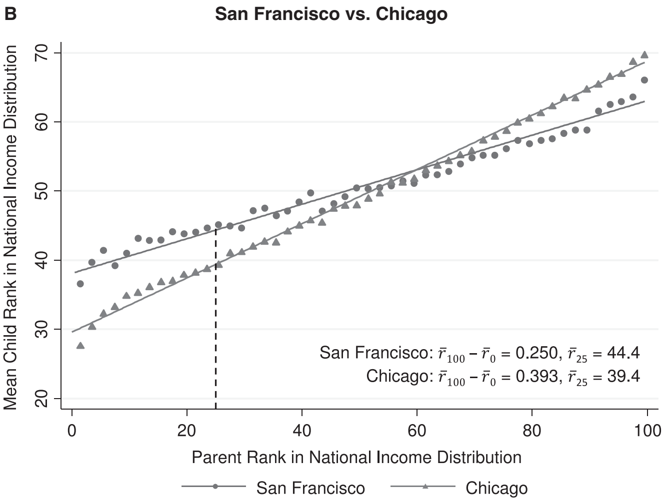


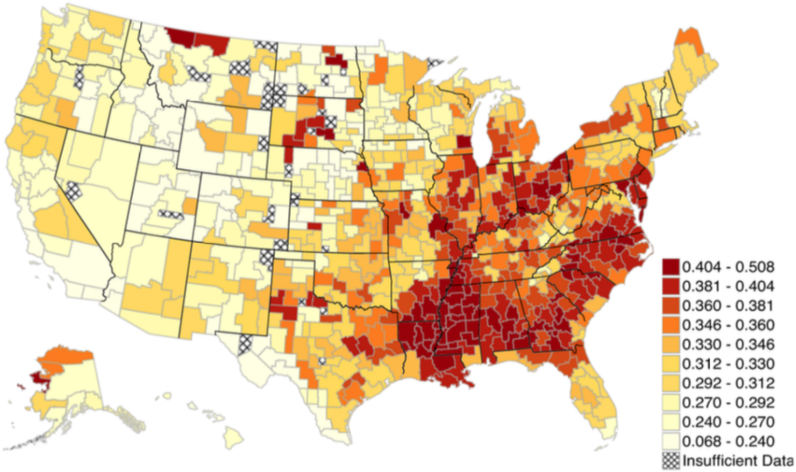
Figure: Rank-Rank Slope, Cross-Country Comparisons





B
San Francisco vs. Chicago
Mean Child Rank in National Income Distribution
20
30
40
50
60
70
0
20
40
60
80
100
Parent Rank in National Income Distribution
San Francisco: $\bar{r}_{100} - \bar{r}_0 = 0.250$, $\bar{r}_{25} = 44.4$
Chicago: $\bar{r}_{100} - \bar{r}_0 = 0.393$, $\bar{r}_{25} = 39.4$
San Francisco
Chicago

B. Relative Mobility: Rank-Rank Slopes $(\bar{r}_{100} - \bar{r}_0)/100$ **by CZ**



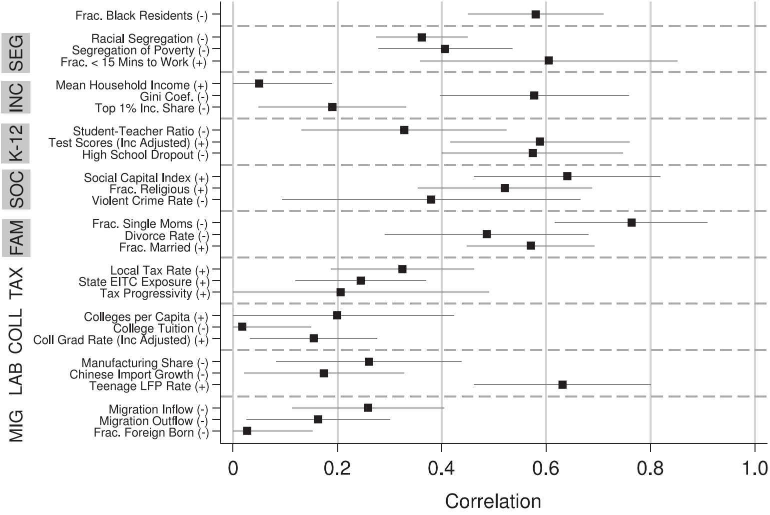
Corr. with baseline $\bar{r}_{25}$ = -0.68 (unweighted), -0.61 (pop-weighted)

Intergenerational Mobility in the 50 Largest Commuting Zones

(1) Upward mobility rank	(2) CZ name	(3) Population	(4) Absolute upward mobility	(5) P(child in Q5 | parent in Q1)	(6) Pct. above poverty line	(7) Relative mobility rank-rank slope
1	Salt Lake City, UT	1,426,729	46.2	10.8	77.3	0.264
2	Pittsburgh, PA	2,561,364	45.2	9.5	74.9	0.359
3	San Jose, CA	2,393,183	44.7	12.9	73.5	0.235
4	Boston, MA	4,974,945	44.6	10.5	73.7	0.322
5	San Francisco, CA	4,642,561	44.4	12.2	72.5	0.250
6	San Diego, CA	2,813,833	44.3	10.4	74.3	0.237
7	Manchester, NH	1,193,391	44.2	10.0	75.0	0.296
8	Minneapolis, MN	2,904,389	44.2	8.5	75.2	0.338
9	Newark, NJ	5,822,286	44.1	10.2	73.7	0.350
10	New York, NY	11,781,395	43.8	10.5	72.2	0.330
11	Los Angeles, CA	16,393,360	43.4	9.6	73.8	0.231
12	Providence, RI	1,582,997	43.4	8.2	73.6	0.333
13	Washington DC	4,632,415	43.2	11.0	72.2	0.330
14	Seattle, WA	3,775,744	43.2	10.9	72.0	0.273
15	Houston, TX	4,504,013	42.8	9.3	74.7	0.325
16	Sacramento, CA	2,570,609	42.7	9.7	71.3	0.257
17	Bridgeport, CT	3,405,565	42.4	7.9	72.4	0.359
18	Fort Worth, TX	1,804,370	42.3	9.1	73.6	0.320
19	Denver, CO	2,449,044	42.2	8.7	73.3	0.294
20	Buffalo, NY	2,369,699	42.0	6.7	73.1	0.368
21	Miami, FL	3,955,969	41.5	7.3	76.3	0.267
22	Fresno, CA	1,419,998	41.3	7.5	71.3	0.295

23	Portland, OR	1,842,889	41.3	9.3	70.5	0.277
24	San Antonio, TX	1,724,863	41.1	6.4	74.3	0.320
25	Philadelphia, PA	5,602,247	40.8	7.4	69.6	0.393
26	Austin, TX	1,298,076	40.4	6.9	71.9	0.323
27	Dallas, TX	3,405,666	40.4	7.1	72.6	0.347
28	Phoenix, AZ	3,303,211	40.3	7.5	70.6	0.294
29	Grand Rapids, Michigan	1,286,045	40.1	6.4	71.3	0.378
30	Kansas City, MI	1,762,873	40.1	7.0	70.4	0.365
31	Las Vegas, NV	1,568,418	40.0	8.0	71.1	0.259
32	Chicago, IL	8,183,799	39.4	6.5	70.8	0.393
33	Milwaukee, WI	1,660,659	39.3	4.5	70.3	0.424
34	Tampa, FL	2,395,997	39.1	6.0	71.3	0.335
35	Orlando, FL	1,697,906	39.1	5.8	71.5	0.326
36	Port St. Lucie, FL	1,533,306	39.0	6.2	71.2	0.303
37	Baltimore, MD	2,512,431	38.8	6.4	67.7	0.412
38	St. Louis, MO	2,325,609	38.4	5.1	69.0	0.413
39	Dayton, OH	1,179,009	38.3	4.9	68.2	0.397
40	Cleveland, OH	2,661,167	38.2	5.1	68.7	0.405
41	Nashville, TN	1,246,338	38.2	5.7	67.9	0.357
42	New Orleans, LA	1,381,652	38.2	5.1	69.5	0.397
43	Cincinnati, OH	1,954,800	37.9	5.1	66.4	0.429
44	Columbus, OH	1,663,807	37.7	4.9	67.1	0.406

Figure: Correlates of Upward Mobility



CHKS (2014) main findings

Findings:

1. Intergenerational (upward) mobility varies substantially across regions within the U.S.
2. Mobility correlates systematically with local demographic and economic factors
    - negatively with residential segregation, income inequality
    - positively with school quality, social capital
    - strong correlation with family structure
3. Intergenerational mobility is a "local problem" that could be tackled using place-based policies

Implications:

- Detailed descriptive evidence across regions within countries can be stepping stone for more causal research designs
- Scale, sample size and spatial resolution of data really matter

Recent spatial literature

Many similar papers from other countries:

1. Sweden: Heidrich (2017) and Branden (2019)
2. Norway: Risa (2019) and Bütikofer, Dalla-Zuanna and Salvanes (2018)
3. Denmark: Eriksen and Munk (2020)
4. Canada: Connolly, Corak and Haeck (2019), Connolly, Haeck and Lapierre (2019), Corak (2020)
5. Italy: Acciari, Polo and Violante (2016)
6. Australia: Deutscher and Mazumder (2019)
7. UK: Bell, Blundell and Machin (2018)
8. Netherlands: ...

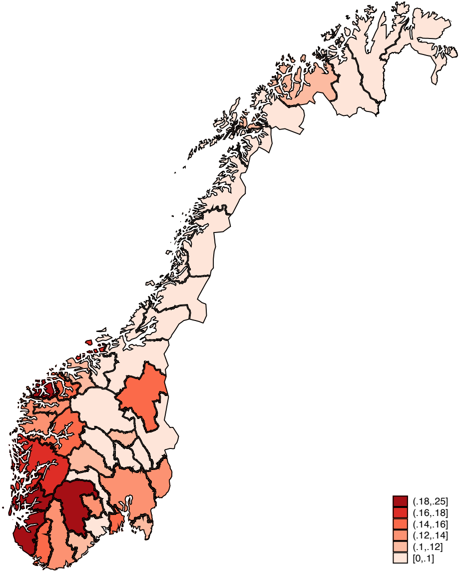


Figure: Probability of reaching the top income quintile when the father was in the lowest quintile (Bütikofer, Dalla-Zuanna and Salvanes, 2018)

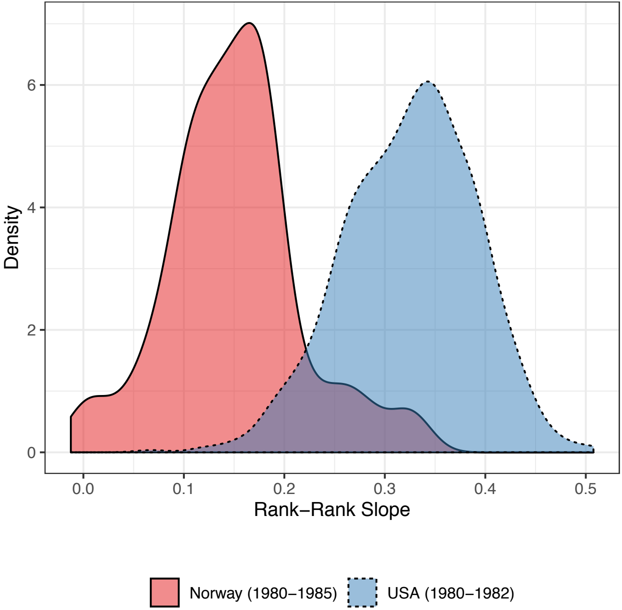


Figure: Rank-Rank slope distributions in Norway and the US (Risa, 2019)

Conceptual issue:

1. Region-level estimates of intergenerational mobility can be noisy
2. Few regions but many regional characteristics that might influence mobility

The literature has started to address these issues more systematically. In particular, in Risa (2019)

- "shrinkage techniques": the shrunk estimates are combination of raw estimates and estimated prior distribution
- machine learning techniques to select regional predictors of intergenerational persistence, using

$$\min_{(\beta_0,\beta)\in\mathbb{R}^{p+1}} \frac{1}{2N}\sum_{i=1}^{N}\left(y_i-\beta_0-x_i^T\beta\right)^2+\lambda\left[(1-\alpha)\|\beta\|_2^2/2+\alpha\|\beta\|_1\right]$$

where $\alpha\in[0,1]$ selects the penalty for non-zero coefficients ($\alpha=1$ lasso regression, $\alpha=0$ ridge regression).

Variable	Elastic Net Coef	Boot Coef	Lower Boot CI	Upper Boot CI	OLS Coef	OLS SE	Sample Mean	Sample SD
empshare_oil_gas_adult	-0.0024	-0.0022	-0.0046	0.0000	0.4186	2.1670	0.0085	0.0116
meanperc_mandatory_male_adult	-0.0017	-0.0013	-0.0045	0.0000	-0.0078	0.0313	0.4539	0.0638
empshare_construct_adult	-0.0011	-0.0010	-0.0039	0.0000	0.9136	4.7103	0.0797	0.0252
meanperc_college_female_child	0.0008	0.0008	0.0000	0.0037	0.0236	0.0175	0.4447	0.0514
lowearn_male_adult	0.0012	0.0008	0.0000	0.0037	0.0018	0.0145	0.0599	0.0261
meanperc_college_female_adult	0.0020	0.0016	0.0000	0.0049	-0.0300	0.0250	0.4571	0.0403
empshare_rtl_hotel_cens	0.0024	0.0013	0.0000	0.0042	0.1342	0.1689	0.1101	0.0417
lowearn_male_child	0.0036	0.0035	0.0000	0.0072	0.0133	0.0114	0.0484	0.0258
rank_longcollege_slope	0.0070	0.0063	0.0017	0.0106	0.0155	0.0032	0.0014	0.0007
rank_highschool_slope	0.0142	0.0133	0.0090	0.0178	0.0226	0.0041	0.0023	0.0011

Table: Most Predictive Variables - Rank-Rank slope (Risa, 2019)

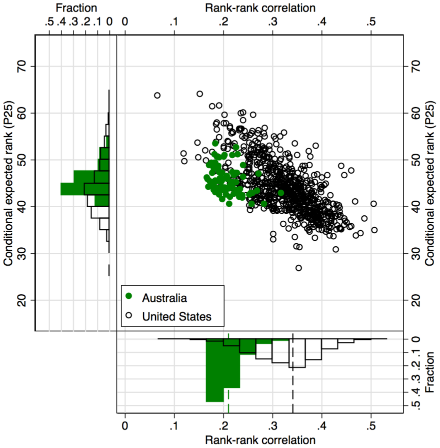


Figure: Distribution of intergenerational mobility measures across regions in Australia and the United States (Mazumder and Deutscher, 2019)

Figure: *From Güell et al. (2018)*

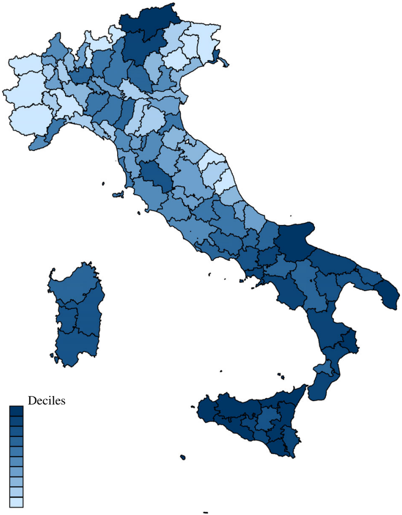


Fig. 2. *Social Mobility (ICS-30) across Italian Provinces*
Notes. Darker blue implies lower mobility. Colour figure can be viewed at wileyonlinelibrary.com.

Causal event study design

The correlation analysis may be suggestive of causal relationships, but ultimately we want to evaluate the effect of policies and institutions using tighter empirical designs.

A popular design is the "area" or "spatial correlation" approach:

1. Estimate mobility by region and cohort/period
2. Use these measures as dependent variable in a difference-in-differences or event study design

- Pekkarinen, Uusitalo and Kerr (2009) estimate the impact of an educational reform on intergenerational mobility in Finland [staggered policy adoption across regions]
- Bütikofer, Zuenna and Salvanes (2018) estimate the impact of the Norwegian oil boom on intergenerational mobility

Example: Pekkarinen et al (2009)

Pekkarinen, Uusitalo and Kerr (2009) is a classic example:

- First step: Estimate intergenerational mobility for each region and cohort, e.g.

$$\log y_s = a + b_{jt} \log y_f + e$$

- Second step: Use estimated slope coefficients from first step as dependent variable in DiD regression,

$$b_{jt} = b_0 + \delta R_{jt} + \Omega D_j + \Psi D_t + v_{jt}$$

  where j indexes regions, t birth cohorts, and R_{jt} equals 1 if the reform had taken place in the municipality by the time when the child was in the relevant age.

- How to weight second step? $\rightarrow$ Might be easier to estimate both steps at once.

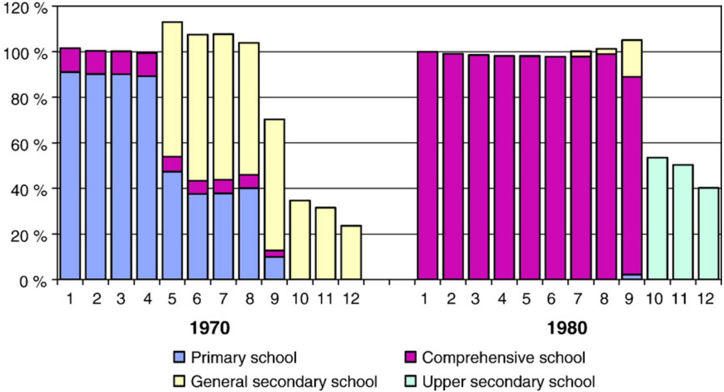


Figure: *The Finnish comprehensive school reform (Pekkarinen et al 2019)*

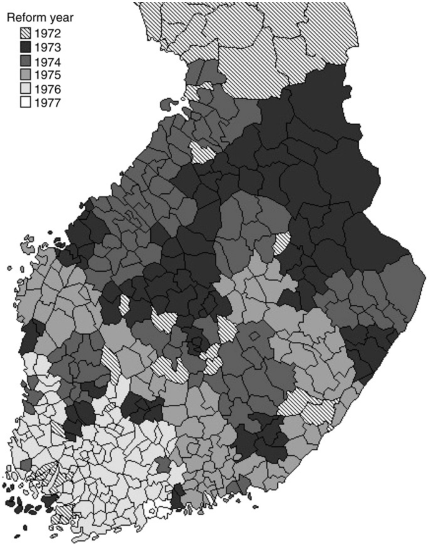


Figure: *The Finnish comprehensive school reform across regions (Pekkarinen et al, 2009)*

	1	2	3	4
Father's earnings	0.277	0.297	0.298	0.296
	(0.014)	(0.011)	(0.010)	(0.014)
Reform		−0.063	−0.019	...
		(0.012)	(0.021)	
Father's earnings*reform		−0.055	−0.069	−0.066
		(0.009)	(0.022)	(0.031)
Cohort dummies			√	√
Father's earnings*cohort dummies			√	√
Region dummies			√	√
Father's earnings*region dummies			√	√
Cohort*region dummies				√
Region-specific trends				√
Observations	20824	20824	20824	20824
R-squared	0.05	0.05	0.05	0.06

Figure: *Regression DiD results (Pekkarinen et al, 2009)*

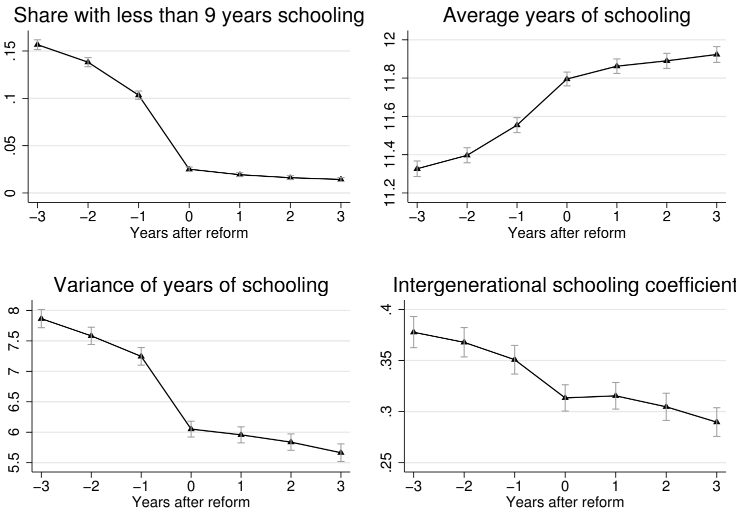


Figure: *Swedish compulsory school reform (Nybom and Stuhler, 2014)*

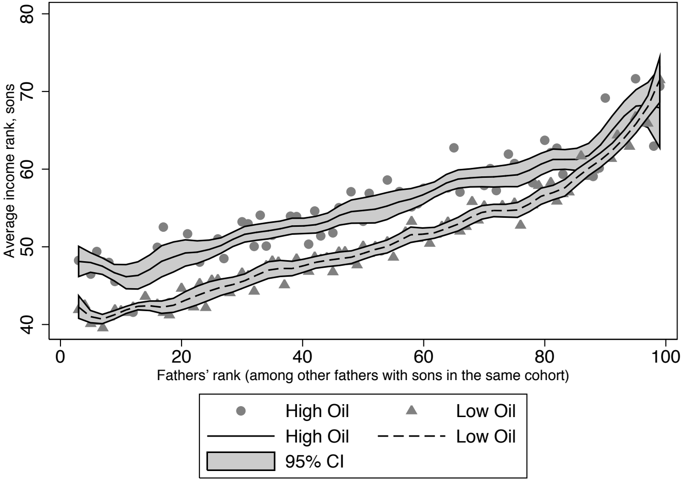


Figure: *Association between sons' and fathers' earnings ranks by region (Bütikofer, Dalla-Zuanna and Salvanes, 2018)*

Causal "mover" design

An alternative design is the "mover" design as used in Chetty et al (2018a, 2018b):

- Ideal experiment: Randomly assign children to new neighborhoods d starting at age m (for the rest of childhood)
- In observational data: We have to control for selection, as choice of neighborhood is likely to be correlated with children's potential outcomes.

$\rightarrow$ See

- Slides by Chetty et al.
- Paper by Mogstad et al. (Figures 6 and Figures 7)

Intergenerational mobility across regions: Causal estimates

Figure: Chetty et al. (2018)