Econ 2450B, Topic 4: Education[1]

Nathaniel Hendren

Harvard

Fall, 2018

[1] I would like to thank Raj Chetty for sharing his slides on education, which comprise much of this lecture.

Education Policy

- Education is one of the largest public goods provided by government
  - Approximately 5.5% of GDP or 1/6 of government expenditure
  - More than 90% at the state and local level
- From a research perspective:
  - Excellent admin. data on inputs and outputs, sharp micro-level variation, and direct policy relevance

Education Policy

- Main Questions

  1. Why should the government intervene? What do we need to estimate for the welfare impact of intervention?
  2. How can we estimate the impact of education policies?

1. Theory and Motivation for Gov't Intervention

2. Estimating the Impact of Education Policies

Motives for Government Intervention

- Motives for government intervention
  - Socially inefficient choices:
    - Fiscal externalities: higher incomes increase future tax revenue
    - Externalities on others: more education may reduce crime, make for more enjoyable conversations, other externalities?
  - Privately inefficient choices
    - Divergence between parent and child preferences
    - Borrowing constraints: Children cannot efficiently invest
    - Optimization failures: individuals misperceive returns to education

Fiscal Externalities

- Part of there return to education falls on the government budget
- Model setup
  - l is labor effort (unobserved)
  - y is an individual's production (observed)
  - θ is an individual's type (unobserved)
  - h is human capital investment (observed)
- Arbitrary production:

$$y = f\left(h, l, \theta\right)$$

  - Condition for maximizing production for each θ:

$$\frac{\partial f}{\partial h} = 1$$

- Utility

$$u\left(c, l, h, \theta\right)$$

Bovenberg and Jacobs (2005)

- Follow Bovenberg and Jacobs (2005, JPubEc)
  - Assume h only affects production of y
    - QUESTION: What if h only entered the utility function and not the production function?

- Production function

$$y = \theta l \phi \left(h\right)$$

  - Production maximized for each θ iff

$$\theta l \phi' \left(h\right) = 1$$

- Utility

$$u = c - \frac{l^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}}$$

Hamiltonian

- Common method for solving uni-dimensional screening problems: Use a Hamiltonian
- Government chooses menu of observable variables, $\{c(\theta), y(\theta), h(\theta)\}_\theta$ to maximize social welfare:

$$\int u(\theta)\,\psi(\theta)\,d\theta$$

where $u(\theta) = u\left(c(\theta), \frac{y(\theta)}{\theta\phi(h(\theta))}\right)$ and $\psi(\theta)$ is a social welfare weight

- Subject to IC constraints and aggregate resource constraints (defined below)

Switch to utility space

- Often helpful to solve these problems in utility space, instead of consumption space
- Define consumption required to obtain utility level u for individual with income y and human capital h

$$c\left(u,l\right)=u+\frac{l^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}}$$

  - Helpful to have quasilinear utility...why?

IC and Resource Constraints

- Start with IC constraint:
  - Define utility a type θ obtains if they say they are type $\hat{\theta}$:

$$\hat{v}\left(\theta,\hat{\theta}\right)=u\left(c\left(\hat{\theta}\right),y\left(\hat{\theta}\right),h\left(\hat{\theta}\right);\theta\right)=c\left(\hat{\theta}\right)-\frac{\left[\frac{y\left(\hat{\theta}\right)}{\phi\left(h\left(\hat{\theta}\right)\right)\theta}\right]^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}}$$

  - IC constraint is (with abuse of notation):

$$u\left(\theta\right)=max_{\hat{\theta}}\hat{v}\left(\theta,\hat{\theta}\right)\quad\forall\theta$$

  - Each type prefers truth-telling

- Resource Constraint:

$$\int T\left(y\left(\theta\right)\right)=\int\left(y\left(\theta\right)-c\left(\theta\right)\right)d\theta\geq0$$

First Order Approach

- Under a single crossing assumption, the global incentive constraints can be replaced with local incentive constraints
- Local IC constraints described by envelope theorem:

$$
\begin{aligned}
u'(\theta) &= \frac{\partial u}{\partial \theta} = -\left(\frac{y(\theta)}{\phi(h(\theta))}\right)^{1+\frac{1}{\epsilon}} \frac{\frac{d}{d\theta}\theta^{-\left(1+\frac{1}{\epsilon}\right)}}{1+\frac{1}{\epsilon}} \\
&= -\left(\frac{y(\theta)}{\phi(h(\theta))}\right)^{1+\frac{1}{\epsilon}} \frac{\frac{d}{d\theta}\theta^{-\left(1+\frac{1}{\epsilon}\right)}}{1+\frac{1}{\epsilon}} \\
&= -\left(\frac{y(\theta)}{\phi(h(\theta))}\right)^{1+\frac{1}{\epsilon}} \theta^{\frac{1}{\epsilon}} \\
&= \frac{1}{\theta}\left(\frac{y(\theta)}{\phi(h(\theta))}\right)^{1+\frac{1}{\epsilon}}
\end{aligned}
$$

- Note that $u'(\theta) > 0$. Implies more productive types must get higher utility...

Hamiltonian

- Hamiltonian:
  - Think of θ as "time"
  - $u(\theta)$ is the state variable (we have a constraint for $u'(\theta)$)
  - Control variables (aka co-state variables): $h(\theta)$, $y(\theta)$, and $c(\theta)$

Hamiltonian

$$H = u(\theta)\psi(\theta) - \gamma_{IC}(\theta)\left[\frac{1}{\theta}\left(\frac{y(\theta)}{\phi(h(\theta))}\right)^{1+\frac{1}{\epsilon}}\right] + \gamma_{RC}\left[y(\theta) - h(\theta) - c\left(u(\theta), \frac{y(\theta)}{\theta\phi(h(\theta))}\right)\right]$$

- Key insight: at the optimum, $\frac{\partial H}{\partial f(X)} = 0$ for all ctsly diff functions of control variables $f(X)$.
  - Trick: substitute back $l(\theta)$ instead of $y(\theta)$.

$$H = u(\theta)\psi(\theta) - \gamma_{IC}(\theta)\frac{1}{\theta}l(\theta)^{1+\frac{1}{\epsilon}} + \gamma_{RC}\left[\theta l(\theta)\phi(h(\theta)) - h(\theta) - c(u(\theta), l(\theta))\right]$$

  - Now, take derivative wrt h holding l and v constant:

$$\frac{\partial H}{\partial h} = \gamma_{RC}\left[\theta l(\theta)\phi'(h(\theta)) - 1 - \frac{dc}{dh}|_{l,v}\right] = 0$$

Full Deductibility

- Note that $\frac{dc}{dh}\big|_{l,v} = 0$, so that:

$$\theta l\left(\theta\right)\phi'\left(h\left(\theta\right)\right) = 1$$

- All education expenses, h, should not be taxed!
- If all income is taxed, then h should be deductible.
- How general is this?
  - Depends on shape of incentive constraints (Stantcheva 2013).

Stantcheva 2016

- Consider general case: $y(\theta) = \phi(h, \theta)$:

$$\hat{v}\left(\theta, \hat{\theta}\right) = c\left(\hat{\theta}\right) - \frac{\left[\frac{y\left(\hat{\theta}\right)}{\phi\left(h\left(\hat{\theta}\right), \theta\right)}\right]^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}}$$

  - Then, IC constraints imply:

$$v'(\theta) = \left[\frac{y\left(\hat{\theta}\right)}{\phi\left(h\left(\hat{\theta}\right), \theta\right)}\right]^{\frac{1}{\epsilon}} \frac{y(\theta)\frac{\partial\phi}{\partial\theta}}{\phi^2} = \left[\frac{y\left(\hat{\theta}\right)}{\phi\left(h\left(\hat{\theta}\right), \theta\right)}\right]^{1+\frac{1}{\epsilon}} \frac{\frac{\partial\phi}{\partial\theta}}{\phi}$$

  - Note $\frac{\frac{\partial\phi}{\partial\theta}}{\phi}$ does NOT depend on h when $\phi = h\theta$.
  - The general IC constraint now enters derivative of H wrt h
  - Hicksian coefficient of complementarity:

$$\rho = \frac{\frac{\partial^2\phi}{\partial\theta\partial h}\phi}{\frac{\partial\phi}{\partial\theta}\frac{\partial\phi}{\partial h}}$$

  Subsidize human capital more (less) than taxes if $\rho < 1$ ($\rho > 1$)

Weak Separability (Hendren + Kaplow ongoing...)

- Can frame this as weak separability / Atkinson-Stiglitz:
- Note that

$$u(c, y, h; \theta) = \tilde{u}(v(c, h), y; \theta) \implies h, c \text{ same tax rate}$$

  implies tax h as consumption
- Similarly

$$u(c, y, h; \theta) = \tilde{u}(c, v(y, h); \theta) \implies h, y \text{ same tax rate}$$

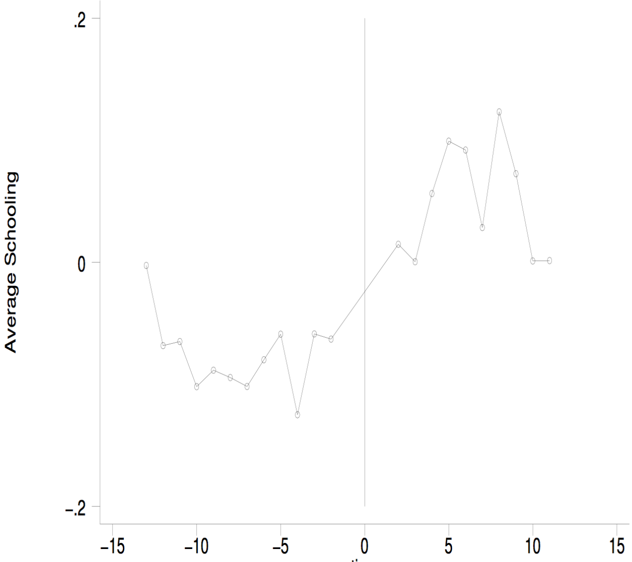
  implies tax y as pre-tax income (i.e. don't tax it!)
- Sigh...turns out we didn't need the Hamiltonian at all!

Externalities

- Education provides fiscal externalities
- What about other externalities?
  - Ex: crime, voting behavior, others' wage rates
  - Classic Pigouvian externality (recall $\frac{\partial u}{\partial E}\frac{dE}{d\theta}$)
- Focus here on Lochner and Moretti (2003), who study effects of schooling on imprisonment using 1960-80 Census data
  - Research design: changes in state compulsory schooling laws that affect cohorts differentially

Impact of Schooling Laws on Educational Attainment

Figure 3: The Effect of Increases in Compulsory Attendance Laws on Average Years of Schooling



Effect of Years of Schooling on Imprisonment

	IV Estimates			Control Function
	(1)	(2)	(3)	(4)
WHITES				
Second-Stage				
Years of Schooling	-0.11	-0.09	-0.14	-0.09
	(0.02)	(0.05)	(0.06)	(0.05)
First Stage				
Compulsory Attendance = 9	0.278	0.222	0.202	
	(0.026)	(0.024)	(0.024)	
Compulsory Attendance = 10	0.213	0.199	0.176	
	(0.035)	(0.034)	(0.033)	
Compulsory Attendance $\geq$ 11	0.422	0.340	0.329	
	(0.037)	(0.033)	(0.033)	
First Stage F-test (d.o.f. = 3)	52.5	38.6	36.2	

Source: Lochner and Moretti 2003

Externalities: Lochner and Moretti (2003)

- Lochner and Moretti show an extra year of schooling reduces incarceration rates significantly
  - 0.1 pct decline for white males relative to a mean of 1%
  - 0.3 pct decline for black males relative to mean of 3%
- Gap in schooling between whites and blacks accounts for more than one-fourth of difference in crime rates
- Externality from crime reduction is about 20% of private return

Private Optimization

- Socially inefficient choices provide rationale for government intervention
  - Suggests need to measure externalities
  - But only externalities?
- But what about the impact of education on children? Can this matter for welfare?
  - Not if investment is privately efficient
  - Do children privately optimize their choice of eduction?

Private Optimization

- Becker: Yes
  - Child-parent bargaining leads to efficient allocation
    - Even if parent and child preferences differ
  - Why?
  - Parents invest and children repay in future (or take less bequest)
  - Implies optimal investment in human capital as long as bequests are positive (Becker and Tomes)

Becker

- Child utility
$$u_k\left(c_k, l_k\right)$$
- Earnings given by
$$y_k = f\left(l_k, h_k; \theta_k\right)$$
- Budget constraint
$$c_k \leq y_k + t$$
where t is transfers from parents

Becker

- Parents altruistic utility

$$u_p\left(c_p, l_p, u_k\right)$$

- Budget constraint

$$c_p + t + h \leq f_p\left(l_p; \theta_p\right)$$

  - Note: t and h do not affect u_p other than through u_k.
  - Therefore, choose t versus h to maximize u_k.
  - Should be indifferent to \$1 more of h and \$1 less of t:

$$\frac{\partial f}{\partial h} = 1$$

Private Optimization

- Optimal private investment requires no constraints on t and h
  - Optimal allocation may involve $t < 0$...feasible?
- Key questions:
  - Are there borrowing constraints?
  - Do individuals / parents know the returns to education?

Borrowing Constraints

- U.S. govt. disbursed $47 billion in grant aid and loans in 2000
  - Does this have a significant causal effect on college attendance?
- Dynarski (2003) studies elimination of SSA program to provide aid to students with deceased or disabled SSA beneficiaries in 1982
  - Average annual payment to children attending college with deceased parent pre-1982 was $6,700
- DD estimates of impacts on college attendance using NLSY data
  - Treatment group: children with deceased father

Effect of SSA college aid on probability of attending college

TABLE 2—OLS, EFFECT OF ELIGIBILITY FOR STUDENT BENEFITS ON PROBABILITY OF ATTENDING COLLEGE BY AGE 23

	(1) Difference-in-differences	(2) Add covariates
Deceased father × before	0.182	0.219
	(0.096)	(0.102)
Deceased father	−0.123	Y
	(0.083)	
Before	0.026	Y
	(0.021)	

Source: Dynarski 2003

Optimal Level of Investment in Education

- Reason to think that human capital is privately and socially under-provided
  - Setting riddled with externalities
  - Can't just rely on envelope theorem and private optimization for welfare analysis
    - Exercise: If human capital is mis-allocated, then causal effect of policies on private choices affects private welfare
- Key empirical objects of interest:
  - Impact of investment in education on outcomes
    - Tax revenue (FE)
    - Externalities ($\frac{dE}{d\theta}$)
    - Earnings? Under what conditions does this measure $\frac{\frac{\partial U}{\partial G}}{\lambda}$?

Large Literature Looking at Impact of Education Policies

- Common outcome: test scores
  - More recent estimates of long-run outcomes
- Focus here: Head Start
- Two empirical approaches:
  - Quasi-experimental approach: e.g. Deming (2009)
  - Experimental Approach: HSIS

Siblings Design: Deming (2009)

is uncorrelated with the unobservable determinants of outcomes. The estimating equation is

$$Y_{ij} = \alpha + \beta_1 HS_{ij} + \beta_2 PRE_{ij} + \delta \mathbf{X}_{ij} + \gamma_j + \varepsilon_i, \tag{1}$$

where i indexes individuals and j indexes the family, $\mathbf{X}$ is a vector of family-varying controls, and γ_j is the family fixed effect. HS_{ij} and PRE_{ij} are the estimated effect of Head Start and other preschools, respectively, on the outcomes $Y_{ij.}$ Threats

Siblings Design: Deming (2009)

TABLE 3—THE EFFECT OF HEAD START ON COGNITIVE TEST SCORES

	(1)	(2)	(3)	(4)	(5)
Head Start					
Ages 5–6	−0.025	0.081	0.093	0.131	0.145*
	(0.091)	(0.083)	(0.079)	(0.087)	(0.085)
Ages 7–10	−0.116	0.040	0.067	0.116*	0.133**
	(0.072)	(0.065)	(0.061)	(0.060)	(0.060)
Ages 11–14	−0.201***	−0.053	−0.017	0.029	0.055
	(0.070)	(0.065)	(0.061)	(0.061)	(0.062)
Other preschools					
Ages 5–6	0.167**	0.022	−0.019	−0.102	−0.079
	(0.083)	(0.082)	(0.078)	(0.084)	(0.085)
Ages 7–10	0.230***	0.111*	0.087	0.031	0.048
	(0.070)	(0.064)	(0.061)	(0.061)	(0.065)
Ages 11–14	0.182**	0.076	0.037	−0.040	−0.022
	(0.072)	(0.068)	(0.065)	(0.066)	(0.069)
Permanent income (standardized) mean (0), SD (1)			0.112*		
			(0.064)		
Maternal AFQT (standardized) mean (0), SD (1)			0.353***		
			(0.057)		
Mom high school			0.141**		
			(0.071)		
Mom some college			0.280***		
			(0.080)		
p (all age effects equal—Head Start)	0.074	0.096	0.161	0.092	0.151
Pre-treatment covariates	N	Y	Y	N	Y
Sibling fixed effects	N	N	N	Y	Y
Total number of tests	4,687	4,687	4,687	4,687	4,687
R^2	0.028	0.194	0.268	0.608	0.619
Sample size	1,251	1,251	1,251	1,251	1,251

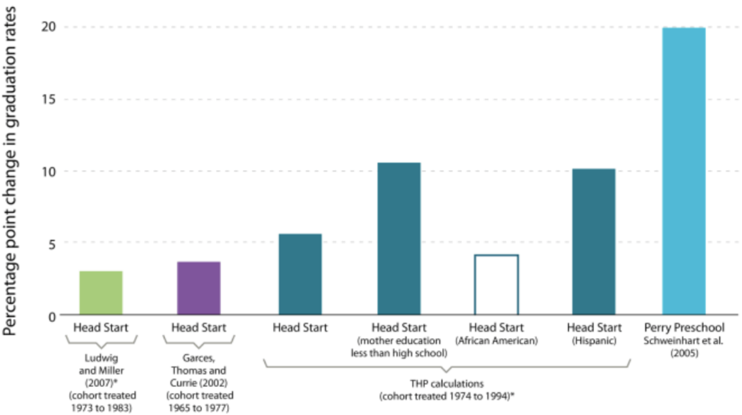
Notes: The outcome variable is a summary index of test scores that includes the child's standardized PPVT and PIAT math and reading scores at each age. Head Start and other preschool indicators are interacted with the three age groups (5–6, 7–10, and 11–14) listed above. Each column includes controls for gender, first born status, and age-at-test and year fixed effects, plus the covariates indicated in the bottom rows. The unit of observation is child-by-age. Standard errors are clustered at the family level.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Hamilton Project Summary

FIGURE 1.

Effect of Early Education on High School Graduation Rates

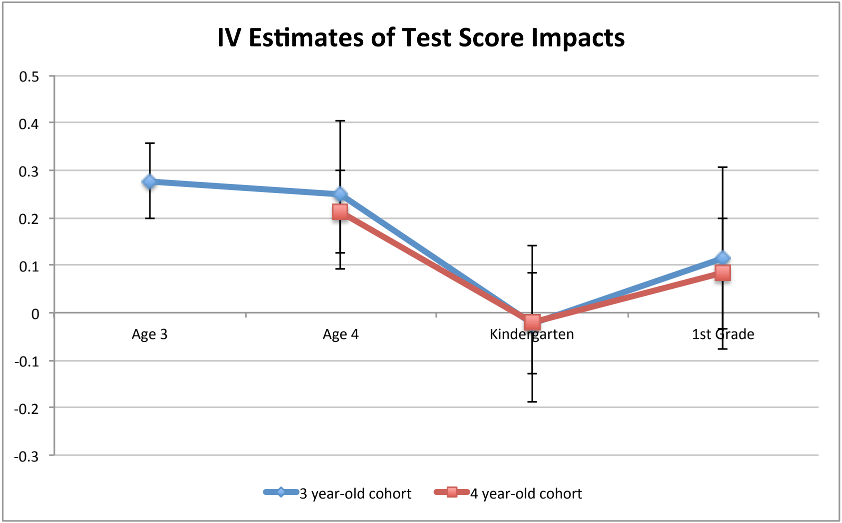


Note: Hollowed bars are not statistically significant at the 10 percent level. *Estimate for Ludwig and Miller (2007) may include and THP calculations do include some GED completers, but authors estimate their contribution to the total is small. There is no differential effect for white students. See the technical appendix.

THE HAMILTON PROJECT BROOKINGS

BUT! HSIS Finds minimal effect

- Head Start Impact Study
- Randomized controlled trial
- Finds minimal effect of being offered head start!
    - Quasi-experimental design flawed?


IV Estimates of Test Score Impacts
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
Age 3
Age 4
Kindergarten
1st Grade
3 year-old cohort
4 year-old cohort

Evaluating Public Programs with Close Substitutes: The Case of Head Start

Patrick Kline Christopher Walters

UC Berkeley and NBER

November 2016

This Paper

Revisit HSIS results in view of wide availability of substitute preschools

Key facts:

- 1/3 of HSIS control group attended other preschools
  - Fraction increased after first year of experiment
- Most of these preschools were publicly funded

Cost-Benefit Analysis

- Basic cost-benefit analysis: would it pay off to admit an extra person into Head Start?

- Toy model: focus on a single benefit (earnings) and compare to impact on government budget
  - When market for preschool substitutes clears:
    - IV-LATE is policy-relevant benefit
    - But costs need to be adjusted for "fiscal externalities"
  - When substitutes are rationed: LATE is not enough

- Empirical analysis:
  - PDV projected earnings impacts $\sim$ HS enrollment costs
  - But accounting for public savings $\Rightarrow$ Benefits $>$ Costs
  - With rationing: Benefits $\gg$ Costs

Technology vs Market Structure

- Develop selection model parameterizing heterogeneity in effects of Head Start vs home care / other preschools
  - Identify using interactions of experimental status with household and site characteristics
  - Decompose LATE into "subLATE's" with respect to particular alternatives
  - Predict effects of changing selection into the program

- Findings:
  - Head Start and other preschools have roughly equivalent average impacts on test scores relative to home care
  - "Reverse-Roy" selection: those with lowest gains most likely to participate
  - Rate of return can be raised further by drawing in new populations

Background on Head Start

- Enrolls one million 3- and 4- year-olds at a cost of $8 billion per year
- Grants awarded to public, private non-profit, and for-profit organizations
- Eligibility: 100% of FPL, with some exceptions
- Competing center-based care programs are ubiquitous:
    - State preschool programs
    - TANF
    - Child Care Development Fund (CCDF)

The Head Start Impact Study

- 1998 Head Start reauthorization included a mandate to determine program's effects: resulted in the HSIS, a large-scale randomized trial
- Stratified random sample of Head Start centers
  - Baseline randomization in Fall 2002
  - Two age cohorts: 55% age 3, 45% age 4
- We focus on summary index of cognitive outcomes based upon average of PPVT and WJ III test scores
  - Normed to have mean zero, std dev. one in control group each year

Table 3: Funding Sources

Largest funding source	Head Start (1)	Other centers (2)	Other centers attended by c -> h compliers (3)
Head Start	0.842	0.027	0.038
Parent fees	0.004	0.153	0.191
Child and adult care food program	0.011	0.026	0.019
State pre-K program	0.004	0.182	0.155
Child care subsidies	0.013	0.097	0.107
Other funding or support	0.022	0.118	0.113
No funding or support	0.000	0.003	0.001
Missing	0.105	0.394	0.375

Table 3: Funding Sources

Largest funding source	Head Start (1)	Other centers (2)	Other centers attended by c -> h compliers (3)
Head Start	0.842	0.027	0.038
Parent fees	0.004	0.153	0.191
Child and adult care food program	0.011	0.026	0.019
State pre-K program	0.004	0.182	0.155
Child care subsidies	0.013	0.097	0.107
Other funding or support	0.022	0.118	0.113
No funding or support	0.000	0.003	0.001
Missing	0.105	0.394	0.375

Table 3: Funding Sources

Largest funding source	Head Start (1)	Other centers (2)	Other centers attended by c -> h compliers (3)
Head Start	0.842	0.027	0.038
Parent fees	0.004	0.153	0.191
Child and adult care food program	0.011	0.026	0.019
State pre-K program	0.004	0.182	0.155
Child care subsidies	0.013	0.097	0.107
Other funding or support	0.022	0.118	0.113
No funding or support	0.000	0.003	0.001
Missing	0.105	0.394	0.375

Table 3: Funding Sources

Largest funding source	Head Start (1)	Other centers (2)	Other centers attended by c -> h compliers (3)
Head Start	0.842	0.027	0.038
Parent fees	0.004	0.153	0.191
Child and adult care food program	0.011	0.026	0.019
State pre-K program	0.004	0.182	0.155
Child care subsidies	0.013	0.097	0.107
Other funding or support	0.022	0.118	0.113
No funding or support	0.000	0.003	0.001
Missing	0.105	0.394	0.375

Inputs

Benefits and Costs of Head Start

Benefits and Costs of Head Start

Benefits

Increased earnings

Benefits and Costs of Head Start

Benefits

Increased earnings

Tuition / time savings for parents

Reductions in crime

Health improvements

Benefits and Costs of Head Start

Benefits

Increased earnings

Tuition / time savings for parents

Reductions in crime

Health improvements

Net Costs

Administrative costs

Benefits and Costs of Head Start

Benefits	**Net Costs**
Increased earnings	Administrative costs
Tuition / time savings for parents	Reduced funding of competing preschool programs
Reductions in crime	Extra tax revenue from more productive children
Health improvements	

Benefits and Costs of Head Start

Benefits

Increased earnings

Tuition / time savings for parents

Reductions in crime

Health improvements

Net Costs

Administrative costs

Reduced funding of competing preschool programs

Extra tax revenue from more productive children

Extra tax revenue from parents Table

Reduced participation in transfer programs

Savings from reduced grade repetition / Special Ed

Standard approach (CEA, 2015)

Table 1: Summary of Cost-Benefit Studies

	Tulsa Full-Day Preschool	Tulsa Half-Day Preschool	Oklahoma & Georgia Preschool	Head Start	Perry Preschool
Year children entered program	2005	2005	1995/98	2002	1962
Value of earnings gains per child	\$27,897	\$16,683	\$24,094	\$14,459	\$92,020
Value of total benefits per child					\$180,257[b]
Cost of program per child	\$9,118	\$4,559	\$4,086	\$9,173	\$20,948
Net benefit per child	\$18,779	\$12,124	\$20,008	\$5,286	\$159,309[b]
Benefit to cost ratio (earnings only)	3.06	3.66	5.90	1.58[a]	4.39
Benefit to cost ratio (all benefits)	–	–	–	–	8.60[b]
Study	Bartik et al. (2012)	Bartik et al. (2012)	Cascio et al. (2013)	Duncan et al. (2010)	Heckman et al. (2010b)

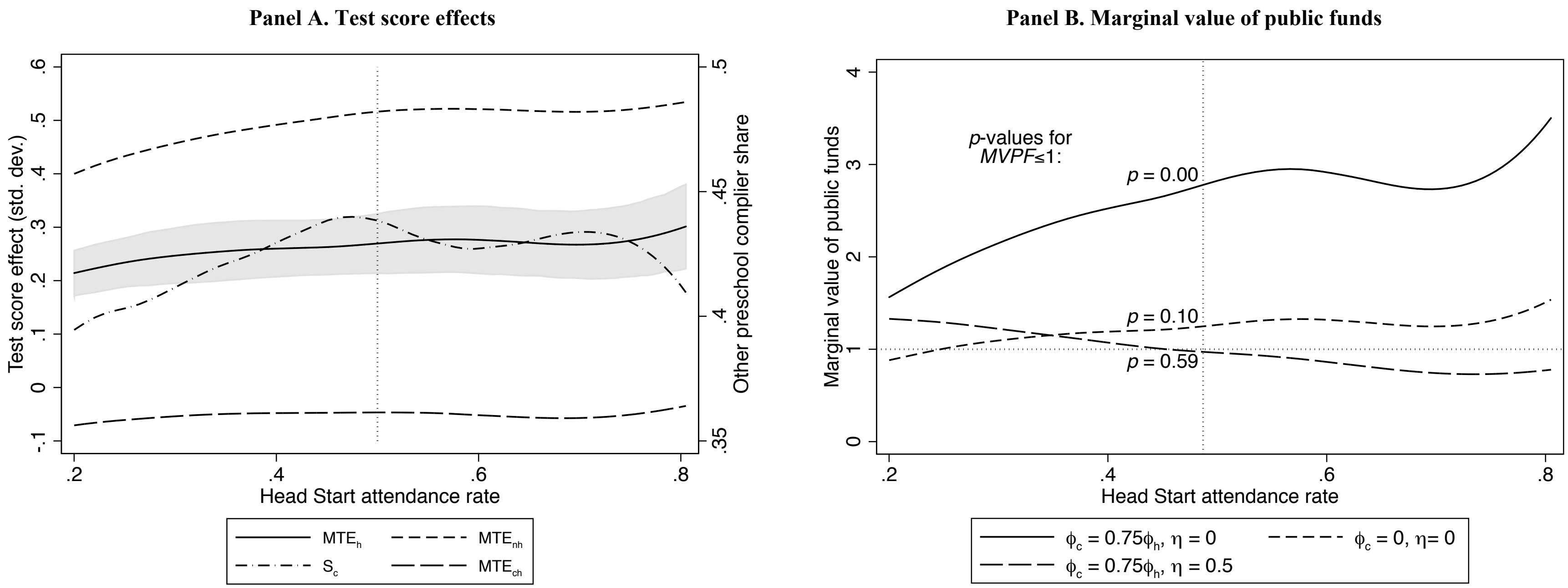


Figure I. Effects of Structural Reforms

Notes: This figure plots predicted test score effects and marginal values of public funds for various values of the program feature f, which shifts the utility of Head Start attendance. Horizontal axes shows the Head Start attendance rate at each f, and a vertical line indicates the HSIS attendance rate ($f = 0$). Panel A shows marginal treatment effects and competing preschool compliance shares. The left axis measures test score effects. MTE_h is the average effect for marginal students, while MTE_{nh} and MTE_{ch} are effects for subgroups of marginal students drawn from home care and other preschools. The right axis measures the share of marginal students drawn from other preschools. The shaded region shows a 90-percent symmetric bootstrap confidence interval for MTE_h. Panel B shows predicted marginal values of public funds for structural reforms, using the same parameter calibrations as Table IV. P-values come from bootstrap tests of the hypothesis that the marginal value of public funds is less than or equal to one at $f = 0$.

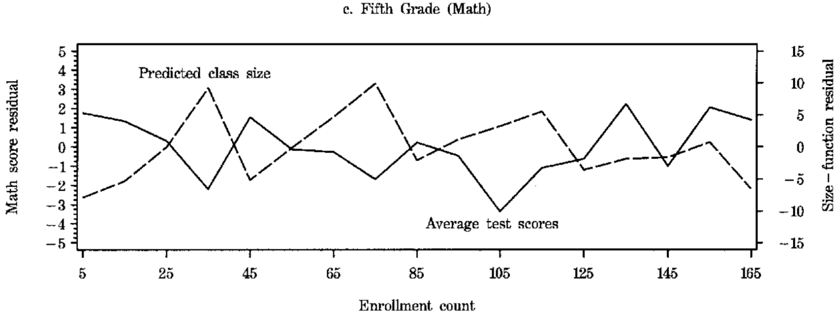
Large Literature...

- There is a large literature estimating the impact of education policies on children's outcomes
- Common outcome: test scores
  - More recent estimates of long-run outcomes
- Wide range of policies/effects that have been studied
- Focus on three here:
  - Classes (and class size)
  - Teachers
  - Schools

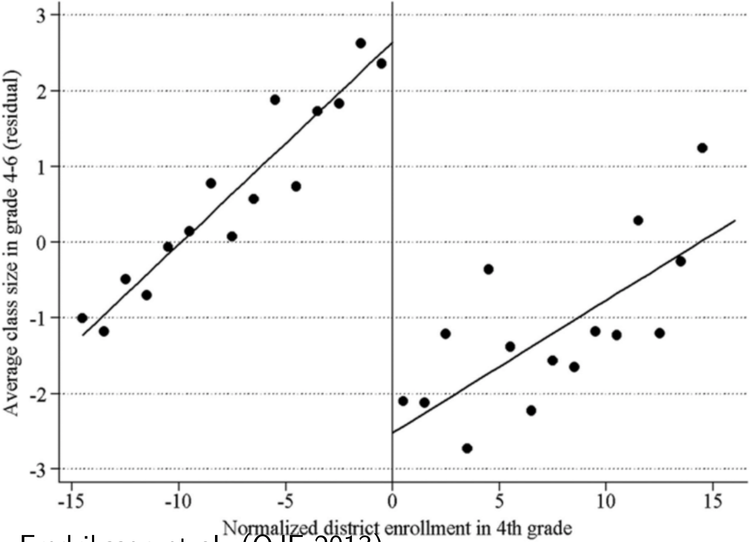
Class Size

- Robust evidence that smaller class sizes improve outcomes
- Quasi-experimental evidence: RD estimates using maximum class size limits (Angrist and Lavy 1997, Fredriksson et al. 2013)
  - Angrist and Lavy: test score impacts in Israel
  - Fredriksson et al.: long-term impacts in Sweden

Angrist and Lavy, 1999 QJE

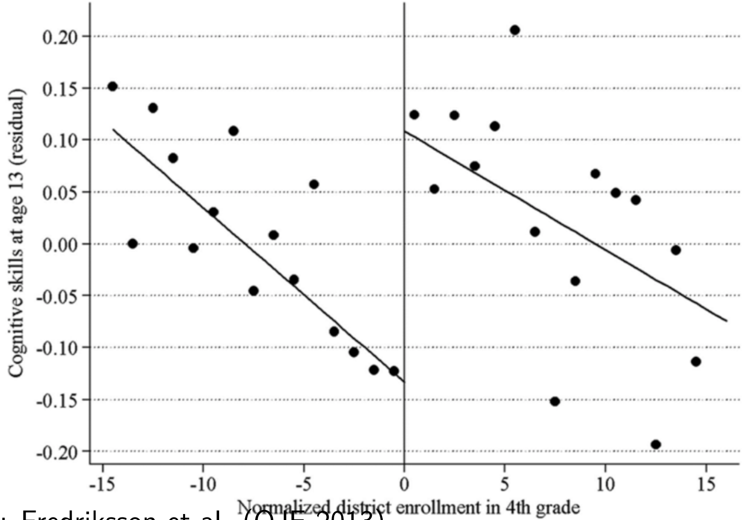


RD Evidence: Class Size vs. Enrollment in Grade 4



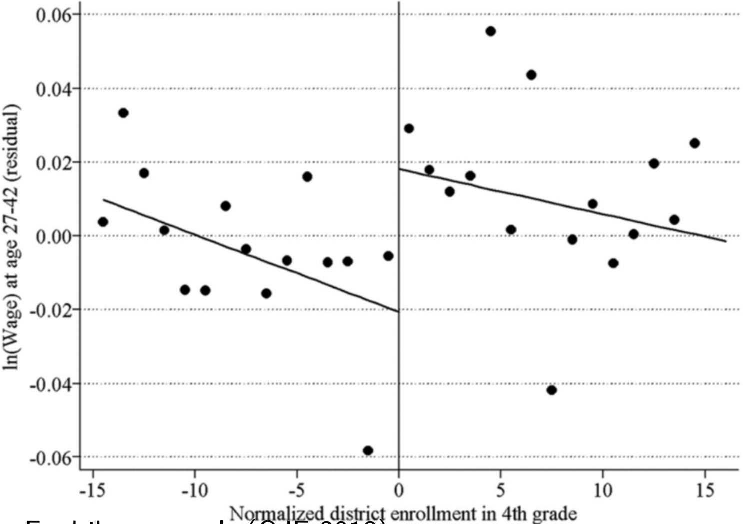
Source: Fredriksson et al. (QJE 2013)

Test Scores at Age 13 vs. Enrollment in Grade 4



Source: Fredriksson et al. (QJE 2013)

Earnings vs. Enrollment in Grade 4



Source: Fredriksson et al. (QJE 2013)

Class Size

- Experimental evidence: Project STAR (Krueger 1999, Chetty et al. 2011)
    - Random assignment of 12,000 kids in Tennessee to classrooms in grades K-3 in mid 1980's
    - Small classes: 15 students, large classes: 23 students

STAR Experiment: Impacts of Class Size

Dep Var:	Test Score	College in 2000	College Quality	Wage Earnings	Summary Index
	(1)	(2)	(3)	(4)	(5)
Small Class	4.81	2.02%	$119	-$4	5.06%
	(1.05)	(1.10%)	($97)	($327)	(2.16%)
Observations	9,939	10,992	10,992	10,992	10,992
Mean of Dep. Var.	48.67	26.4%	$27,115	$15,912	0.00

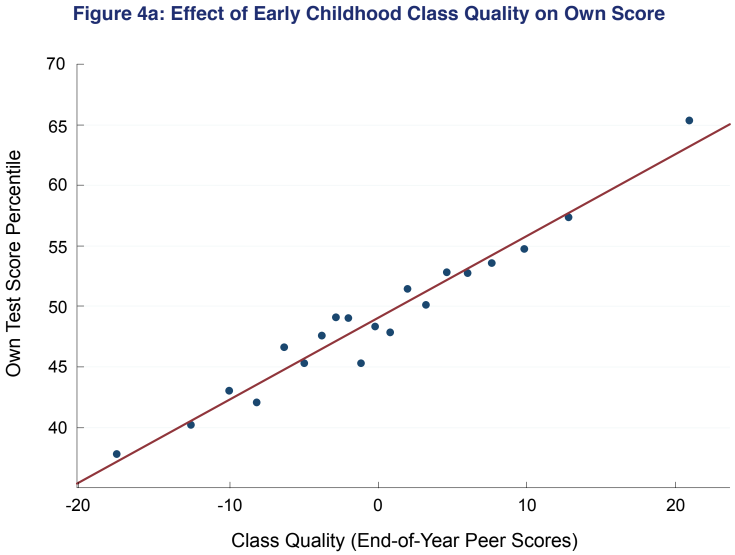
Source: Chetty et al. (QJE 2011)

Project STAR

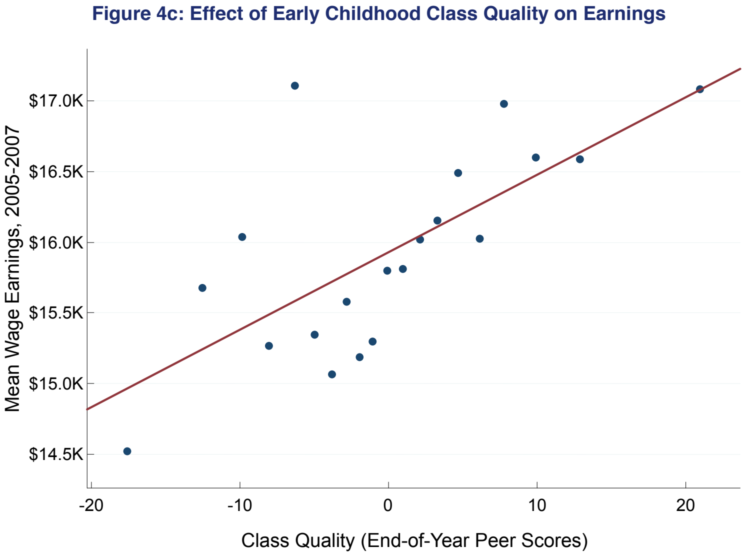
- Project STAR also provides random allocation of children to classrooms
- Maybe more than class size matters?
- Idea: Look for impact of unobservables of classroom (teacher, students, etc.)
- Notation: child i randomly assigned to classroom c
- Define $s_{-i,c}$ to be the test scores of other children in the classroom
  - Measured at the end of the school year

$$y_i = \alpha + \beta s_{-i,c} + \epsilon_i$$

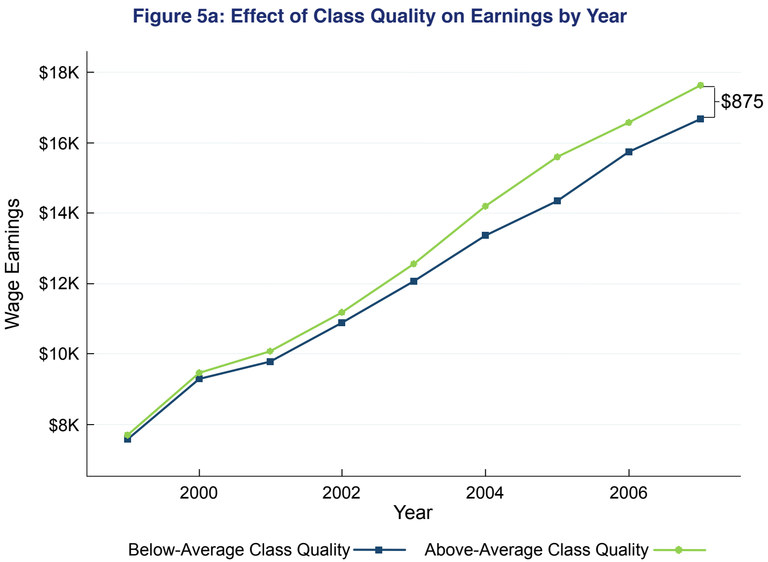
STAR: Impact of Class Size (Chetty et al 2011)



STAR: Impacts of Class Size (Chetty et al 2011)



STAR: Impacts of Class Size (Chetty et al 2011)



Teacher Value-Added Metrics

- Classes matter. Is this:
  - Teachers?
  - Peers?
  - Quality of the blackboard?
  - The air?
- How do we isolate the impact of teachers?
  - Common method: value added modeling (Hanushek (1971), Murnane (1975), Kane and Staiger (2008), Rothstein (2010)

Debate About Teacher Value-Added

- Basic idea: measure teacher's impact on child test scores by conditioning on lagged test scores

1. Potential for bias [Kane and Staiger 2008, Rothstein 2010]
    - Do differences in test-score gains across teachers capture causal impacts of teachers or are they driven by student sorting?
2. Lack of evidence on teachers's long-term impacts
    - Do teachers who raise test scores improve students' long-term outcomes or are they simply better at teaching to the test?

Chetty, Friedman, Rockoff 2014

- Chetty, Friedman, Rockoff (2014a,b) study 2.5 million children from childhood to early adulthood

  1. Develop new quasi-experimental tests for bias in VA estimates
  2. Test if children who get high VA teachers have better outcomes in adulthood

Constructing Value-Added Estimates

- Model the estimation of VA as a forecasting problem
- Simplest case: teachers teach one class per year with N students
- All teachers have test score data available for t previous years
- Objective: predict test scores for students taught by teacher j in year $t+1$ using test score data from previous t years
  - Define $\hat{\mu}_{j,t+1}$ as forecasted impact of teacher j in year $t+1$
  - Use test scores from teacher's past classes from 0 to time t

Constructing Value-Added Estimates

- Three steps to estimate VA ($\hat{\mu}_{j,t+1}$) for teacher j in year $t+1$

1. Form residual test scores A_{is}, controlling for observables X_{is}
   - Regress raw test scores A^*_{is} on observable student characteristics X_{is}, including prior test scores $A^*_{i,s-1}$

2. Regress mean class-level test score residuals in year t on class-level test score residuals in years 0 to $t-1$:

$$\bar{A}_{jt} = a + \psi_{t-1}\bar{A}_{j,t-1} + ... + \psi_0 \bar{A}_{j0} + \varepsilon_{jt}$$

3. Use estimated coefficients $\psi_1, \ldots, \psi_t$ to predict VA in year $t+1$ based on mean test score residuals in years 1 to t for each teacher j:

$$\hat{\mu}_{j,t+1} = \sum_{s=1}^{t} \psi_s \bar{A}_{js}$$

Constructing Value-Added Estimates

- Two special cases:

1. Forecast VA in year t using data from only year $t-s$:

$$\hat{\mu}_{jt} = r_s \bar{A}_{j,t-s}$$

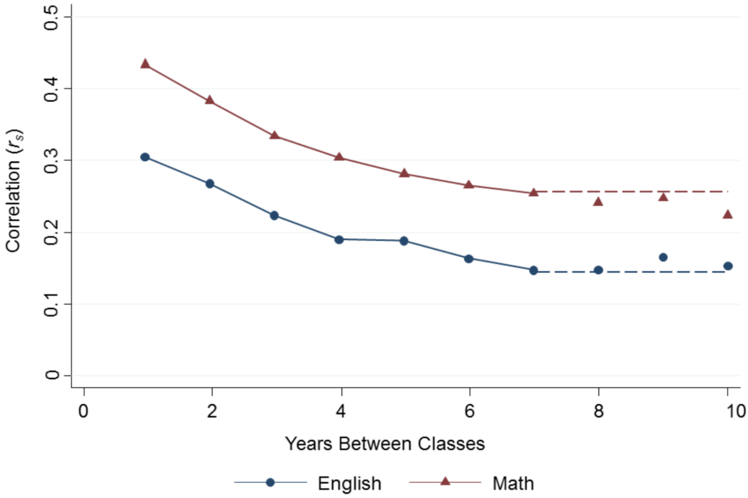
where $r_s = Corr(\bar{A}_t, \bar{A}_{t-s})$ is autocorrelation at lag s

2. Without drift, put equal weight on all prior scores:

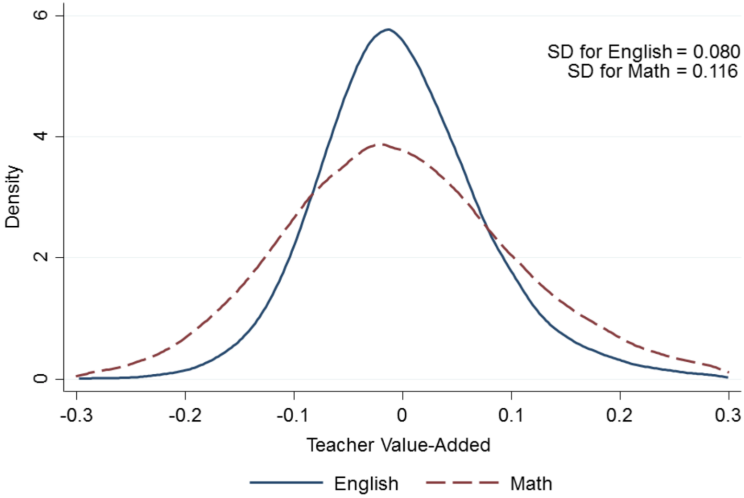
$$\hat{\mu}_{jt} = \bar{A}_j^{-t} \frac{\sigma_\mu^2}{\sigma_\mu^2 + (\sigma_\theta^2 + \sigma_{\bar{\epsilon}}^2 / n) / T}$$

- Bayesian interpretation: shrinkage based on signal-noise ratio (Kane and Staiger 2008)
- Why does this deal with measurement error in $\bar{A}_{j,t}$?

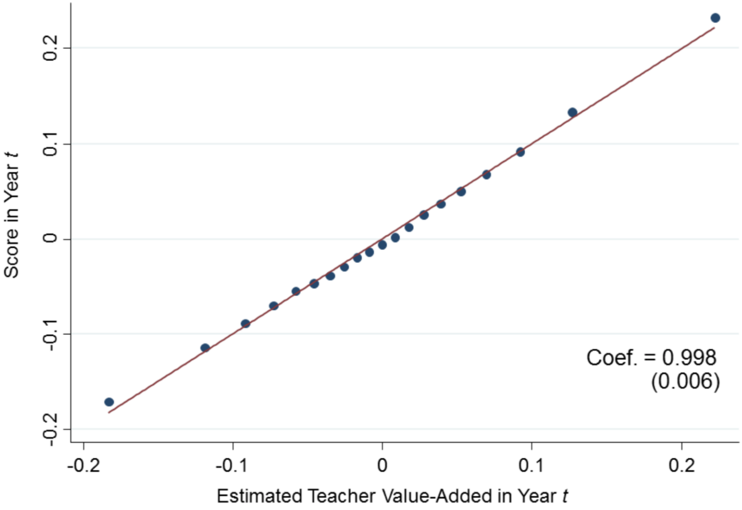
Autocorrelation Vector in Elementary School



Distribution of VA Estimates



Test Score Residuals vs. VA in Cross-Section



Are VA Estimates Biased?

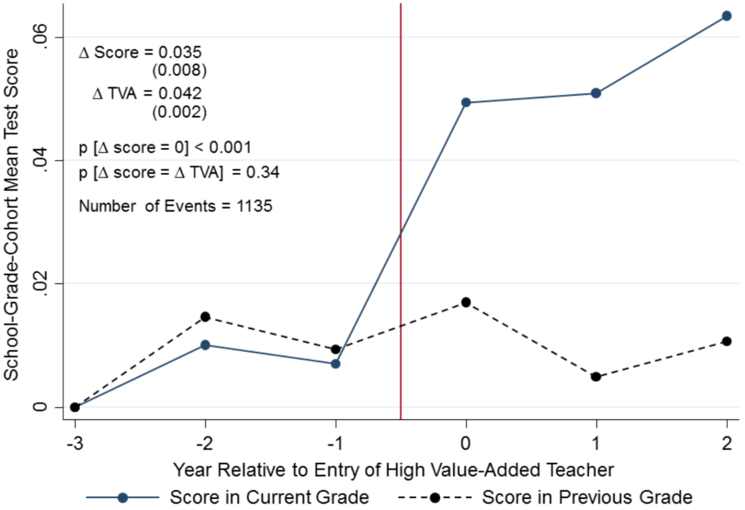
- Let γ denote causal impact of 1 unit increase in teacher's estimated VA on student's test score
  - Define forecast bias as $B = 1 - \gamma$
- Ideal experiment to estimate forecast bias (Kane and Staiger 2008): randomly assign students to teachers with different VA estimates
  - Does a student who is randomly assigned to a teacher previously estimated to be high VA have higher test score gains?
- Use teacher switching as a quasi-experimental analog

Teacher Switchers in School-Grade-Subject-Year Data

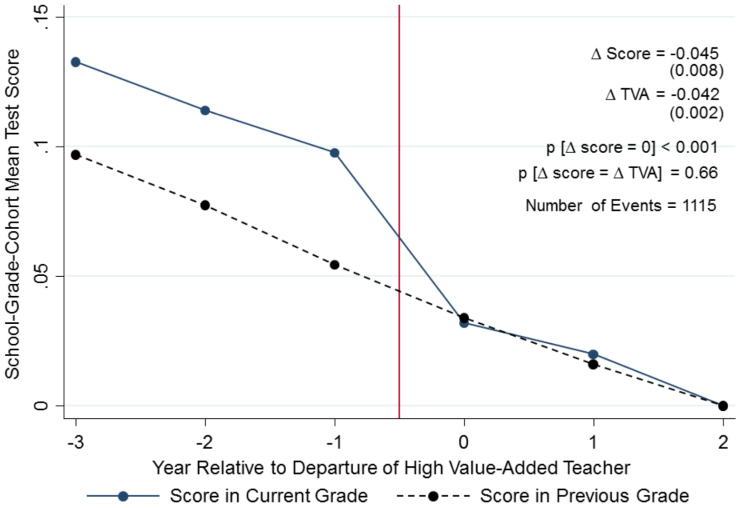
School	Grade	Subject	Year	Teachers	Mean Score	Mean Age 28 Earnings
1	5	math	1992	Jones, Heckman, ...	-.09	$15K
1	5	math	1993	Jones, Heckman, ...	-.04	$17K
1	5	math	1994	Jones, Heckman, ...	-.05	$16K
1	5	math	1995	Katz, Heckman, ...	0.01	$18K
1	5	math	1996	Katz, Heckman, ...	0.04	$17K
1	5	math	1997	Katz, Heckman, ...	0.02	$18K

- Jones switches to a different school in 1995; Katz replaces him

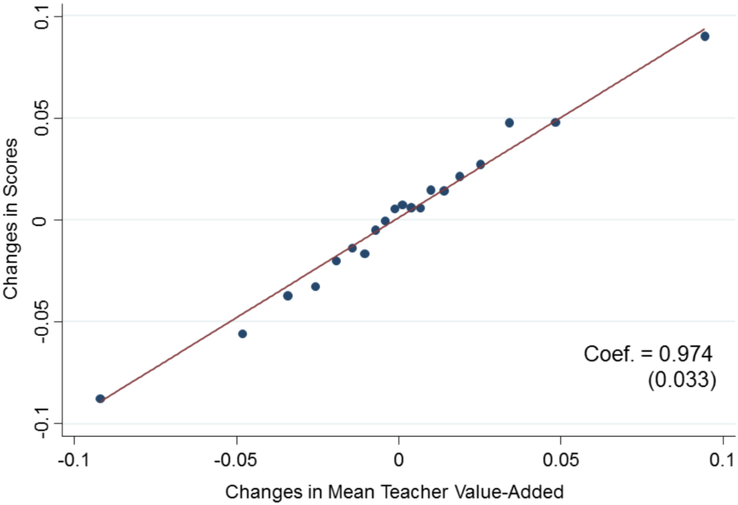
Impact of High VA Teacher Entry on Cohort Test Scores



Impact of High VA Teacher Exit on Cohort Test Scores



Changes in Mean Scores vs. Changes in Mean Teacher VA



Estimates of Forecast Bias with Alternative Control Vectors

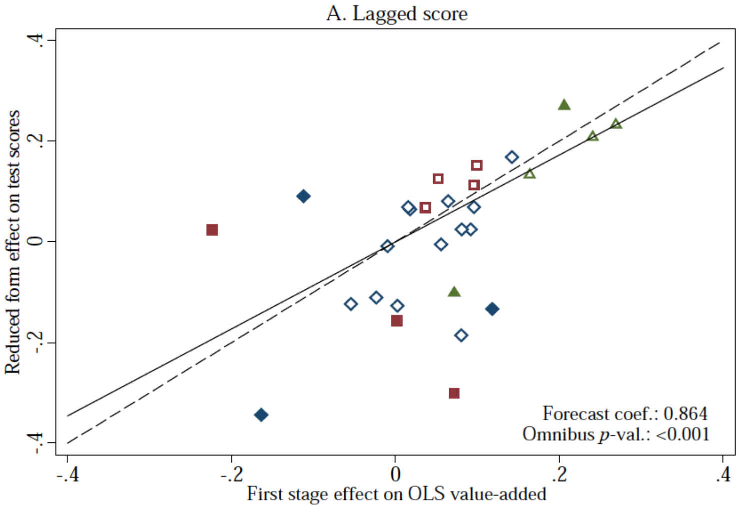
Control Vector	Quasi-Experimental Estimate of Bias (%)
Baseline	2.58 (3.34)
Student-level lagged scores	4.83 (3.29)
Non-score controls only	45.39 (2.26)
No controls	65.58 (3.73)

Impacts of Schools: Combining Lotteries and Value Added

- Large literature exploiting lotteries for over-subscribed schools
- Use lottery to generate exogenous variation in child assignment to schools
  - Note: Schools (as opposed to Classes and Teachers)
- Estimates of more bias at school level (e.g. $B = 0.1 - 0.2$)
- Angrist et al. (2016): Combine Value Added and Lotteries
  - Lotteries available for some schools (but noisy)
  - Value added available for all schools

Angrist et al. (2016)

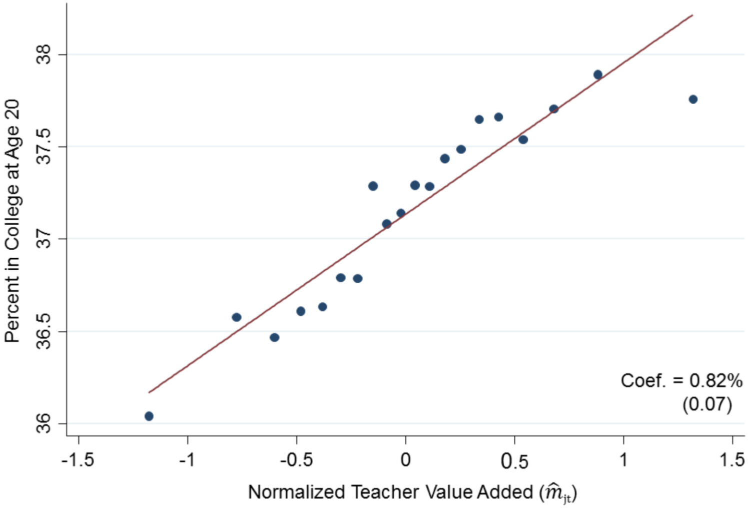
Figure 2: Visual instrumental variables tests for bias



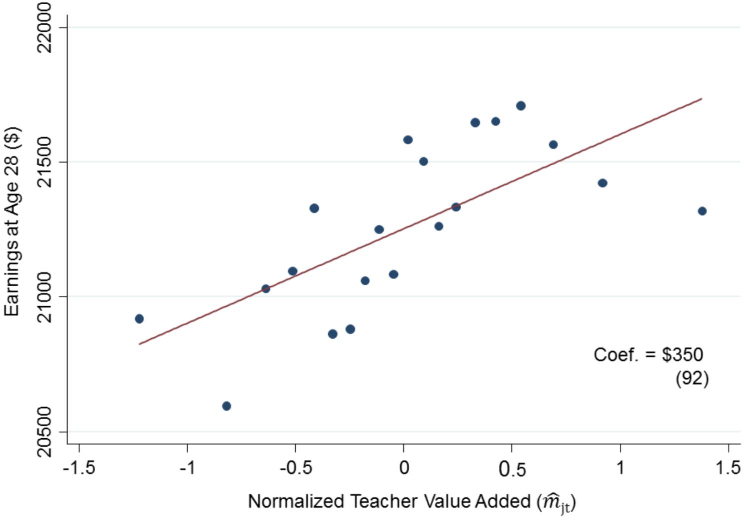
Impacts on Outcomes in Adulthood

- Do teachers who raise test scores also improve students' long-run outcomes?
- CFR, paper #2: Regress long-term outcomes on teacher-level VA estimates
  - Then validate using cross-cohort switchers design
- Interpretation of these reduced-form coefficients (Todd and Wolpin 2003):
  - Impact of having better teacher, as measured by VA, for single year during grades 4-8 on earnings
  - Includes benefit of better teachers, peers, etc. in later grades via tracking

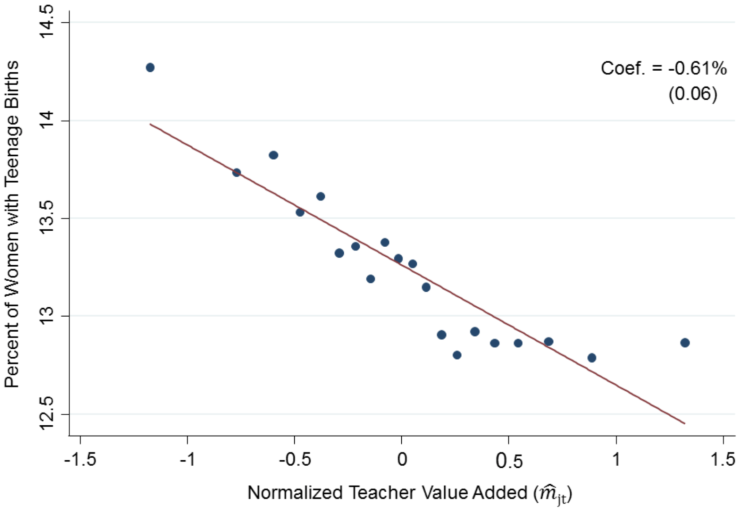
College Attendance at Age 20 vs. Teacher Value-added



Earnings at Age 28 vs. Teacher Value-Added



Women with Teenage Births vs. Teacher Value-Added



Unanswered Questions

- Many important questions remain
  - Distributional incidence: Are some teachers better at teaching some types of students?
  - Which grades are most important?
  - Optimal allocation of students to classes/teachers/peers
    - Redistribution vs. efficiency tradeoff?
- GE Effects:
  - Evidence of Peer Effects (Hoxby, 2000) using cohort variation
  - Benefits capitalized into housing prices
    - Even for school choice policies de-linked from neighborhood choice (Avery and Pathak (2015, NBER WP #21525))

Welfare Implications

- More broadly, value-added methodology can be used in other contexts:
- Tax preparer effects, manager effects, doctor effects, neighborhood effects (next topic...)
- Quasi-experimental designs growing in applied work
    - Can validate using experimental evidence