Econ 2450B, Topic 4: Education

Nathaniel Hendren

Harvard

Spring, 2023

Education Policy

- Up until now: imagined taxable income the result of static effort
    - No role of "endogenous" wages
    - Human capital!
- Large literature on the impact of education on outcomes
    - Excellent administrative data on inputs and outputs and sharp micro-level variation

Education Policy

- Main Questions

1. When/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?
3. What are the fundamental constraints preventing efficient educational investment and what are the implications for optimal policy?

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 $\hat{\theta}$:

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

  - 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:

$$\int T\left(y\left(\theta\right)\right) = \int \left(y\left(\theta\right) - c\left(\theta\right)\right) d\theta \geq 0$$

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:

$$\theta l(\theta)\,\phi'(h(\theta)) = 1$$

- 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

- How do these results relate to last lecture on weak separability / Atkinson-Stiglitz?
- Exercise: Suppose individuals solve

$$\max_{c,y,h} u\left(c, y, h; \theta\right) \quad s.t.\ B\left(c, y, h\right) \leq 0$$

where B is the same for all people but utility varies with θ.

- Show that if

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

implies tax h as consumption

- Similarly if

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

implies tax y as pre-tax income (i.e. don't tax it!)

- Turns out we didn't need the Hamiltonian at all!

Private Optimization

- What if children are not the ones choosing h?
  - Becker: Can child-parent bargaining leads to efficient allocation?
    - Yes
  - 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?

Empirical Evidence on Education Policies

- Lecture 1 discussed Zimmerman (2014) finding large impacts of college admittance on earnings
  - Other literature documents impacts of scholarships/etc.
  - General result: policies that increase human capital tend to have large earnings effects (10% per year of education is the benchmark)
- Here, discuss two other post-k12 policies:
  - Michigan HAIL Aid – documents low-income students can be induced to apply
  - Year Up – positive news on job training programs (which have historically not been as successful)
- Then turn to discussion of K12 policies

Michigan HAIL Aid

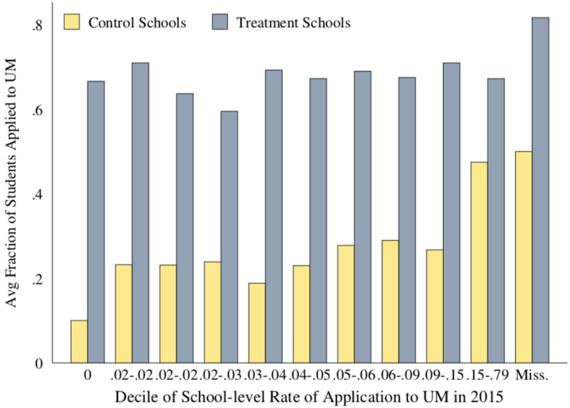
- Dynarski et al (2018) study the impact of an intervention aimed at increasing applications to the University of Michigan from Low-Income Students
  - Policy contacted "students (as well as their parents and principals) with an encouragement to apply along with a promise of four years of free tuition and fees upon admission."
- Dynarski et al (2018) look at impacts on application and enrollment at UM
  - Treated students are 2X likely to apply and 2X likely to enroll


HAIL SCHOLARSHIP
M
UNIVERSITY OF
MICHIGAN
#VICTORS2020
Four Years of Free Tuition
If Admitted to the University of Michigan
Sample,
Go Blue!

Dynarski et al. (2018)

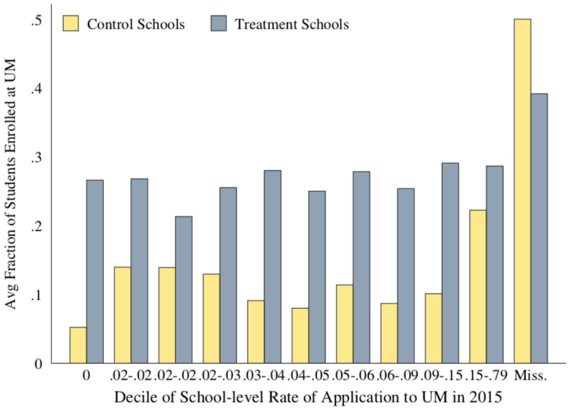
Figure 6
Estimated Effect of HAIL Scholarship on Application and Enrollment to UM by Prior School-Level Application Rate to UM (First and Second HAIL Cohorts)

(a) Applied



Dynarski et al. (2018)

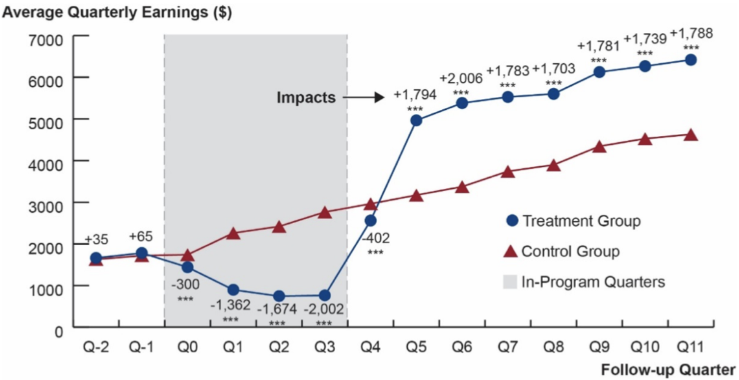
(b) Enrolled



Job training

- Large literature on job training programs, often tested with an RCT
  - National Supported Work Experiment
  - Job Training Partnership Act
  - Range of welfare reform policies in the 80s and 90s
- More promising recent evidence from sectoral training programs that provide training for particular sectors of the economy
- PACE/Abt Conducted an RCT of Year Up
- Year Up program provides a year of training to prep low-income adults (18-24) for jobs:
  - 6 months skills training (e.g. computer repair, communication skills, etc)
  - 6 month internship with major corporate partners (e.g. CVS, State Street, etc.)
  - Job search assistance

Exhibit ES-3: Impact on Average Earnings in Successive Follow-up Quarters



SOURCE: Match to wage records in the National Directory of New Hires for 1,638 treatment and 858 control group members.
NOTES: Impacts appear as numbers giving differences in average earnings for treatment and control group members in each follow-up quarter. *** Impact in a two-tailed test is statistically significant at the 1-percent level, ** at the 5-percent level, * at the 10-percent level.

K12-Education

- Large debate about whether K12 school funding increases outcomes
- Coleman (1966) documents minimal correlation between spending and outcomes
  - But concerns about selection effects
- Jackson, Johnson, and Persico (2016 QJE) exploit variation in school finance reform
  - Find large effects of school funding on performance
  - Identify from variation in timing of court-mandated reforms in the 1970s-80

School Finance Reform

Figure 4: *Hypothesized Patterns with a Causal Effect of an Exogenous Increase in Spending*

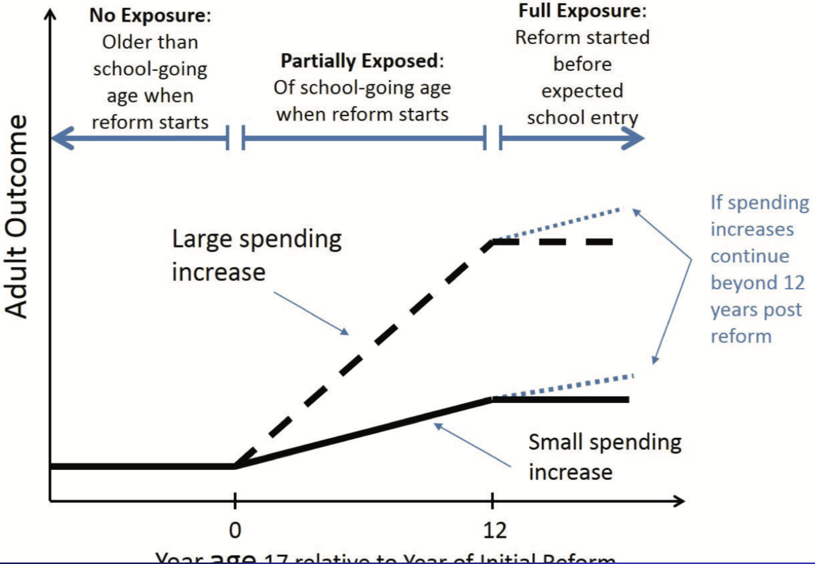
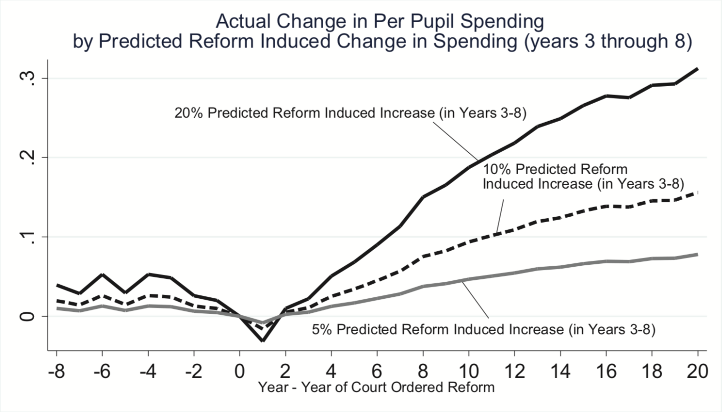


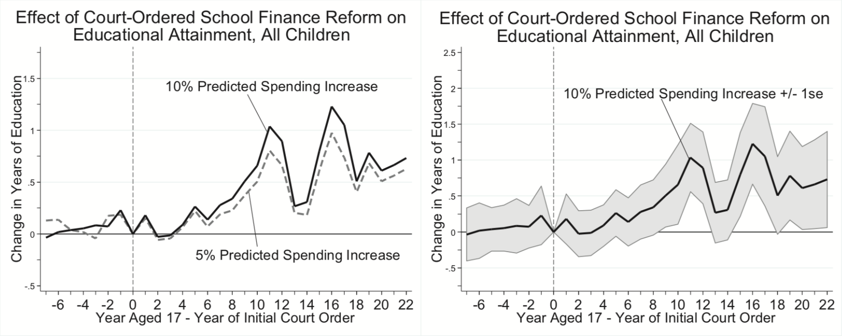
Figure 2: *Effect of Predicted Reform Induced Spending Changes Interacted with Time Relative to the First Court Ordered Reform on Actual Spending Over Time: CCD Population*



Data: The sample includes all school districts in the United States between the years of 1967 and 2010. The sample is made up of 483,047 district-year observations. Each district is weighted by average enrollment for the full sample.
Model: These plots present the estimated event time coefficients of a regression on per-pupil spending at the district level on year fixed effects, district fixed effects, and the percentile group of the district in the state distribution of median income interacted with a full set of event-time indicator variables from 10 years prior to 19 years after the first equalization plan. The event study plots are shown for the top and bottom 10 percent of districts in the state distribution of median district income. The event time plot has been re-centered at zero for the 10 pre-reform years so that the estimated coefficients represent the change in spending relative to

School Finance Reform

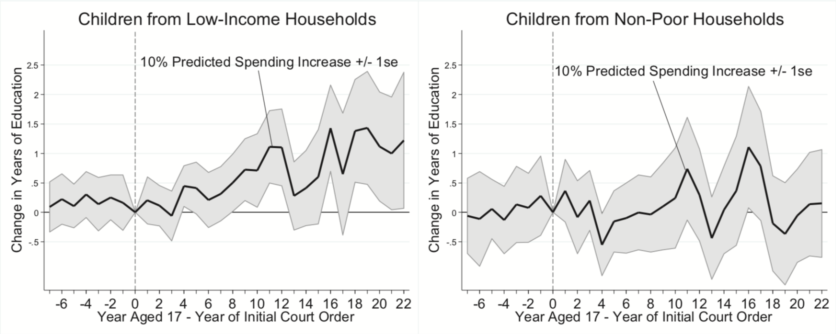
Figure 5: *Effect of Predicted Spending Changes Interacted with Time Relative to the First Court Ordered Reform on Years of Completed Schooling*



Data: PSID geocode Data (1968-2011), matched with childhood school and district characteristics. Analysis sample includes all PSID individuals born 1955-1985, followed into adulthood through 2011, (N=15,353 individuals from 1,409 school districts (1,031 child counties, 50 states).
Models: Results are based on event-study models that include: school district fixed effects, race-specific year of birth fixed effects, race*census division-specific linear cohort trends, controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, roll-out of "War on Poverty" & related safety-net programs (community health centers, county expenditures on Head Start (at age 4), food stamps, medicaid, AFDC, UI, Title-I (average during childhood years), timing of state-funded Kindergarten), controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends, and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). Standard errors are clustered at the childhood county level. Main school finance reform variables allowed to affect outcomes through both the amount of induced school spending changes and the duration of school-age years of exposure to reform-induced spending changes (i.e., models include intercept and slope terms of intensity of treatment (district

School Finance Reform

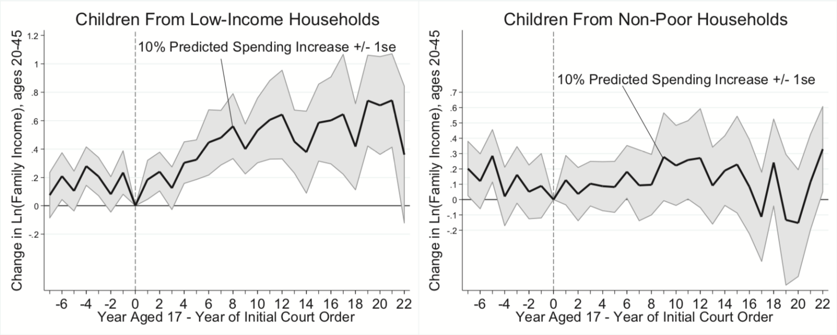
Figure 6: *Effect of Predicted Spending Changes Interacted with Time Relative to the First Court Ordered Reform on Years of Completed Schooling: By Childhood Poverty Status*



Data: PSID geocode Data (1968-2011), matched with childhood school and district characteristics. Analysis sample includes all PSID individuals born 1955-1985, followed into adulthood through 2011, (N=15,353 individuals from 1,409 school districts (1,031 child counties, 50 states).
Models: Results are based on event-study models that include: school district fixed effects, race-specific year of birth fixed effects, race*census division-specific linear cohort trends, controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, roll-out of "War on Poverty" & related safety-net programs (community health centers, county expenditures on Head Start (at age 4), food stamps, medicaid, AFDC, UI, Title-I (average during childhood years), timing of state-funded Kindergarten), controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends, and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). Standard errors are clustered at the childhood county level. Main school finance reform variables allowed to affect outcomes through both the amount of induced school spending changes and the duration of school-age years of exposure to reform-induced spending changes (i.e., models include intercept and slope terms of intensity of treatment (district

School Finance Reform

Figure 11: *Effect of Predicted Reform Induced Spending Changes Interacted with Time Relative to the First Court Ordered Reform on Annual Family Income: By Childhood Poverty Status*



Data: PSID geocode Data (1968-2011), matched with childhood school and district characteristics. Analysis sample includes all PSID individuals born 1955-1985, followed into adulthood through 2011, (N=15,353 individuals from 1,409 school districts (1,031 child counties, 50 states).
Models: Results are based on non-parametric event-study models that include: school district fixed effects, race-specific year of birth fixed effects, race*census division-specific linear cohort trends, controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, roll-out of "War on Poverty" & related safety-net programs (community health centers, county expenditures on Head Start (at age 4), food stamps, medicaid, AFDC, UI, Title-I (average during childhood years), timing of state-funded Kindergarten), controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends, and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). Standard errors are clustered at the childhood county level. Main school finance reform variables allowed to affect outcomes through both the amount of induced school spending changes and the duration of school-age years of exposure to reform-induced spending changes (i.e., models include intercept and slope terms of intensity of treatment (district spending change) and interaction terms of "school spending change*exposure years" in order to capture dose of treatment in terms

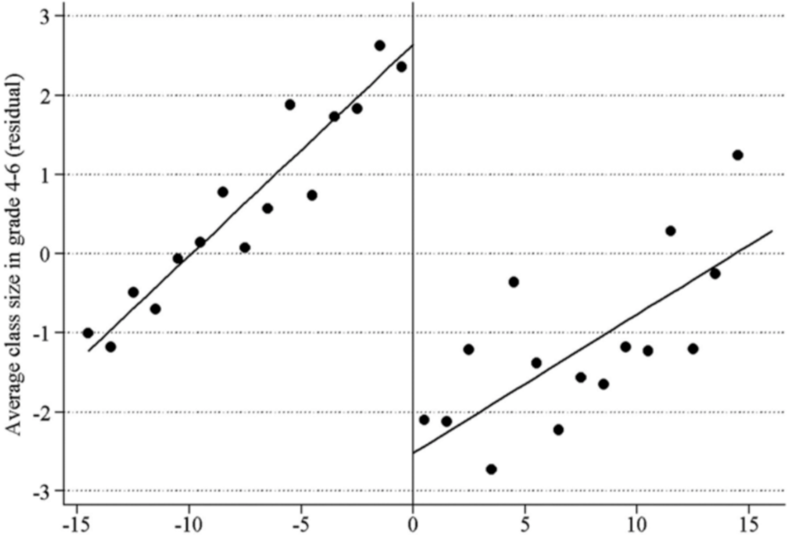
K12-Education

- Expanded school funding from mandated finance reform leads to large increases in student achievement and later-life earnings
- What about schools generates differences in outcomes?
  - Classes (and class size)
  - Teachers
  - Other factors?
- Large quasi-experimental literature analyzing these channels

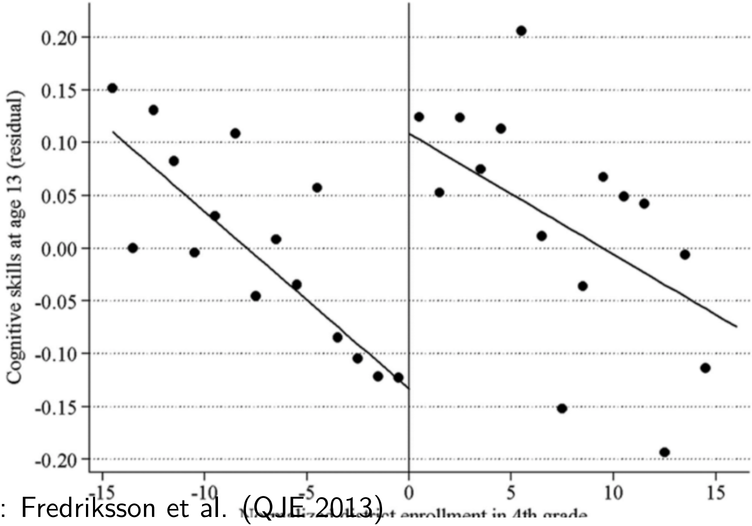
Class Size

- 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

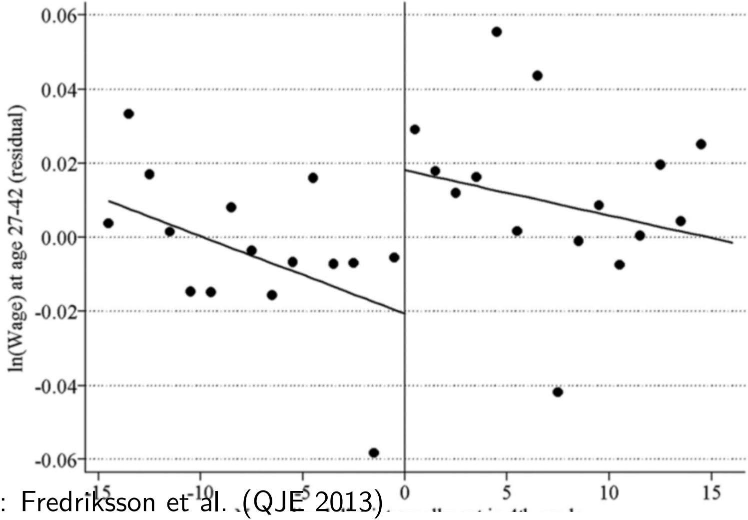


RD Evidence: Class Size vs. Enrollment in Grade 4



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

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

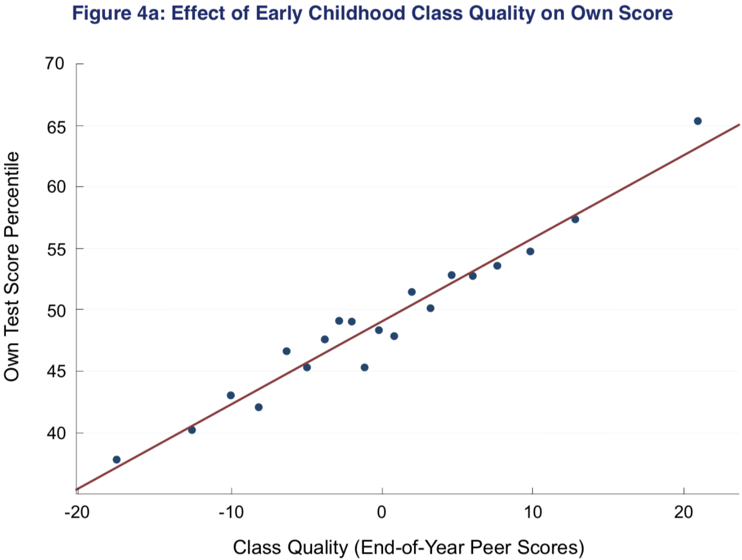
Source: Chetty et al. (QJE 2011)

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



Source: Chetty et al. (QJE 2011)

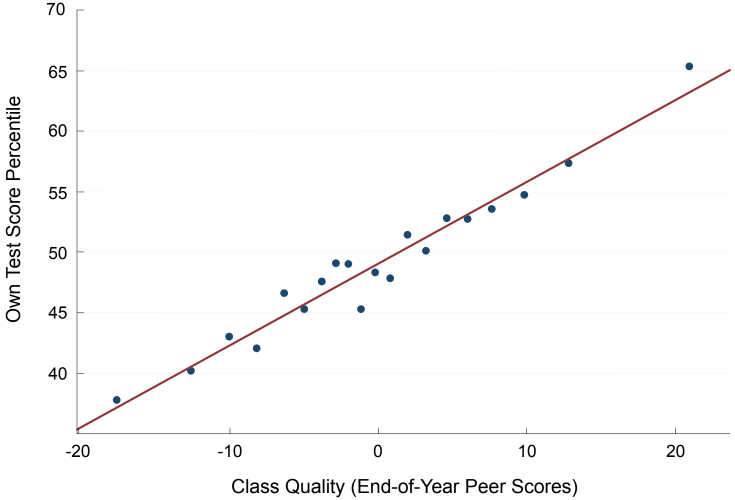
Project STAR

- 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

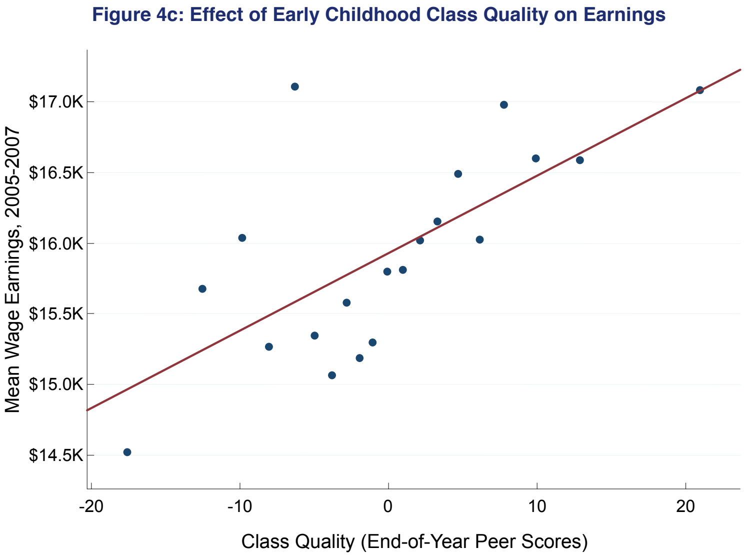
$$y_i = \alpha + \beta s_{-i,c} + \epsilon_i$$

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

Figure 4a: Effect of Early Childhood Class Quality on Own Score



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 $t-1$:

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

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 $t - s$:

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

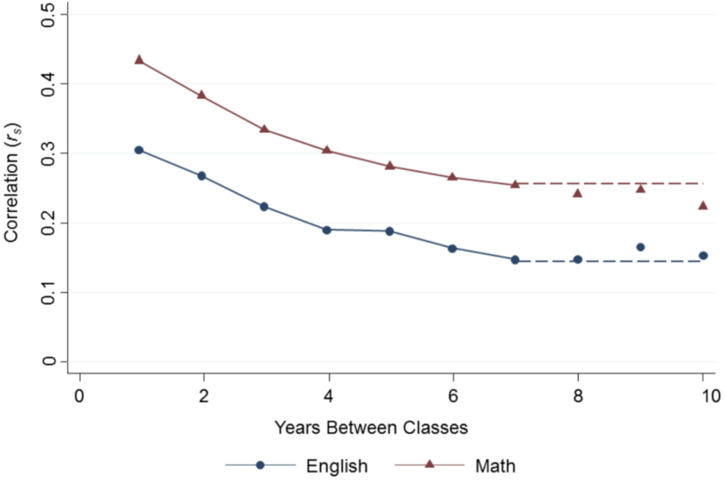
where $r_s = Corr(\bar{A}_t, \bar{A}_{t-s})$ is autocorrelation at lag s

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

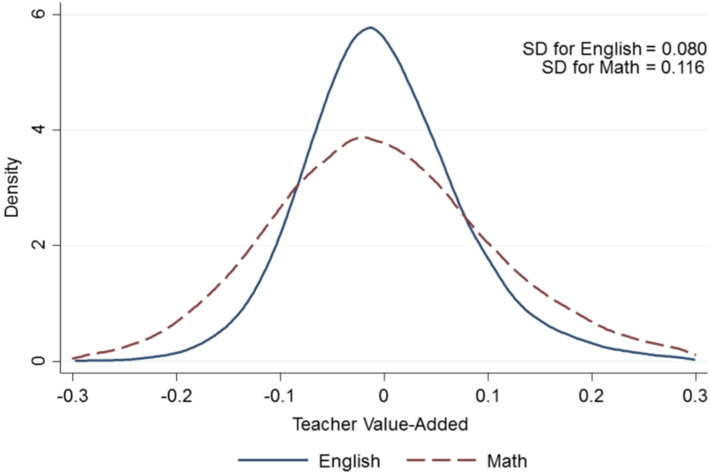
$$\hat{\mu}_{jt} = \bar{A}_j^{-t} \frac{\sigma_\mu^2}{\sigma_\mu^2 + (\sigma_\theta^2 + \sigma_{\bar{\epsilon}}^2 / n) / T}$$

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

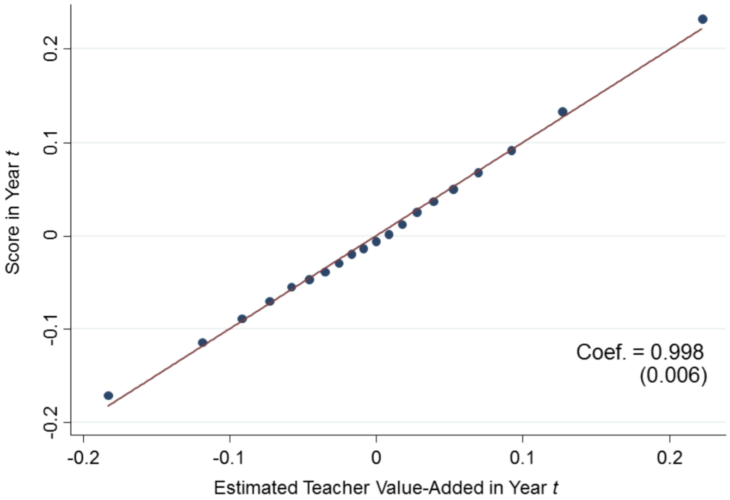
Autocorrelation Vector in Elementary School



Distribution of VA Estimates



Test Score Residuals vs. VA in Cross-Section



Are VA Estimates Biased?

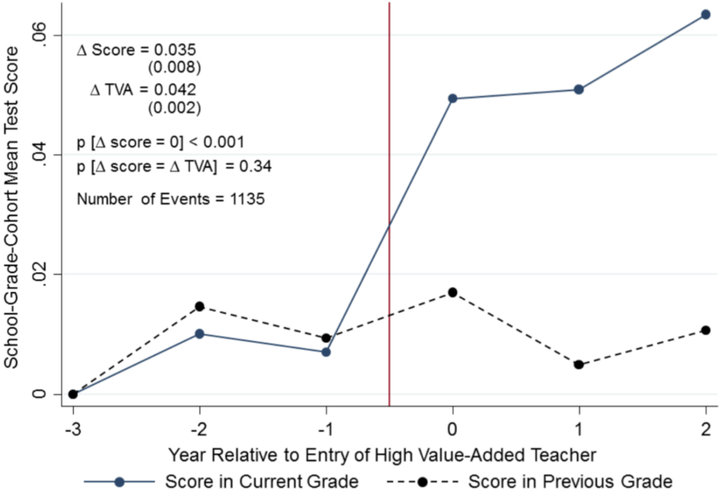
- Let γ denote causal impact of 1 unit increase in teacher's estimated VA on student's test score
    - Define forecast bias as $B = 1 - \gamma$
- Ideal experiment to estimate forecast bias (Kane and Staiger 2008): randomly assign students to teachers with different VA estimates
    - Does a student 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

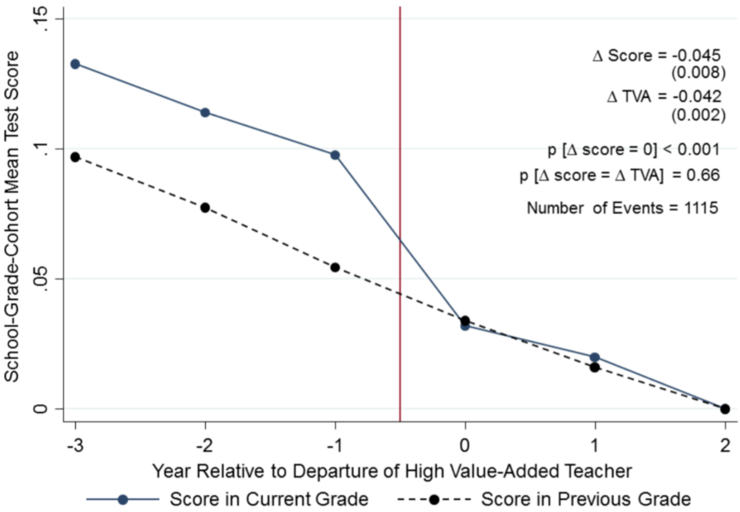
School	Grade	Subject	Year	Teachers	Mean Score	Mean Age 28 Earnings
1	5	math	1992	Jones, Heckman, …	-.09	$15K
1	5	math	1993	Jones, Heckman, …	-.04	$17K
1	5	math	1994	Jones, Heckman, …	-.05	$16K
1	5	math	1995	Katz, Heckman, …	0.01	$18K
1	5	math	1996	Katz, Heckman, …	0.04	$17K
1	5	math	1997	Katz, Heckman,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

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

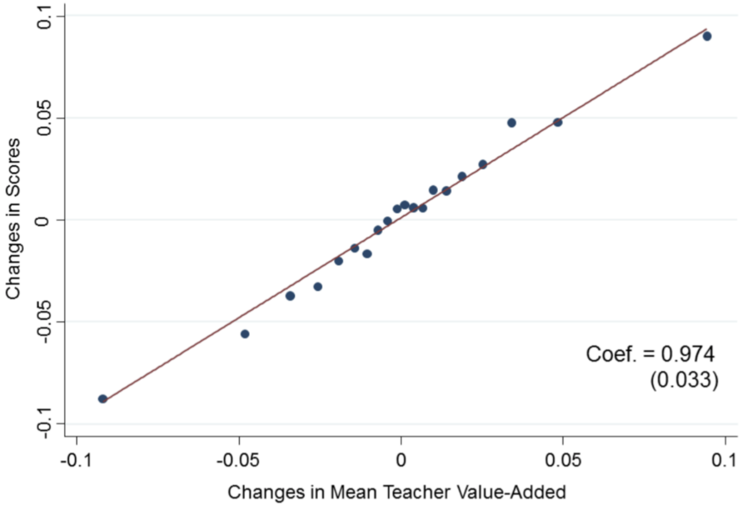
Impact of High VA Teacher Entry on Cohort Test Scores



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

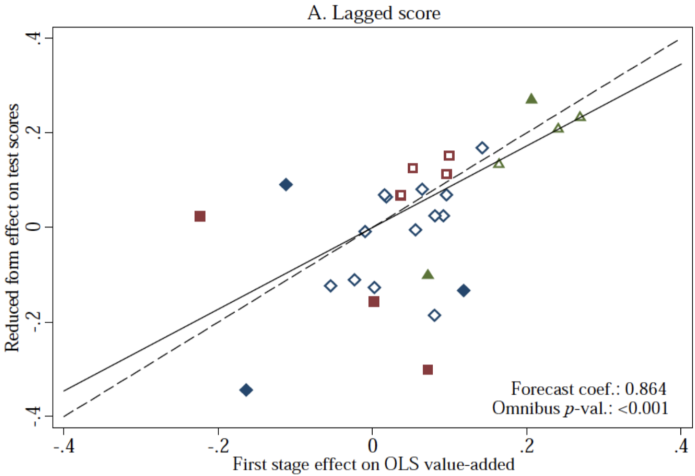
Control Vector	Quasi-Experimental Estimate of Bias (%)
Baseline	2.58
	(3.34)
Student-level lagged scores	4.83
	(3.29)
Non-score controls only	45.39
	(2.26)
No controls	65.58
	(3.73)

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)

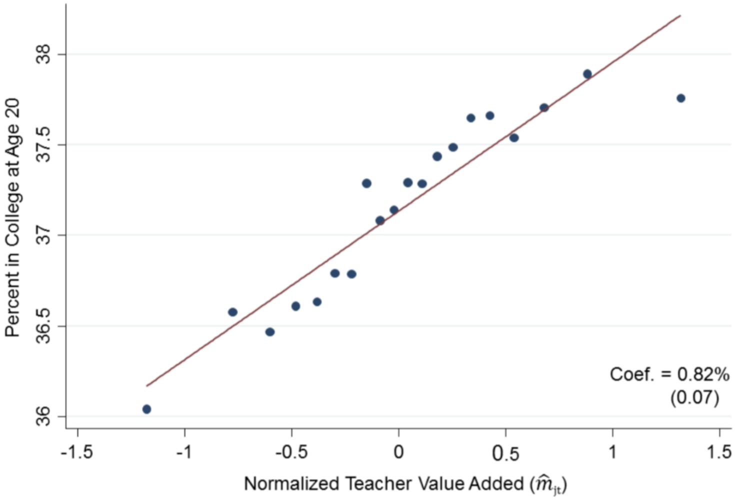
Figure 2: Visual instrumental variables tests for bias



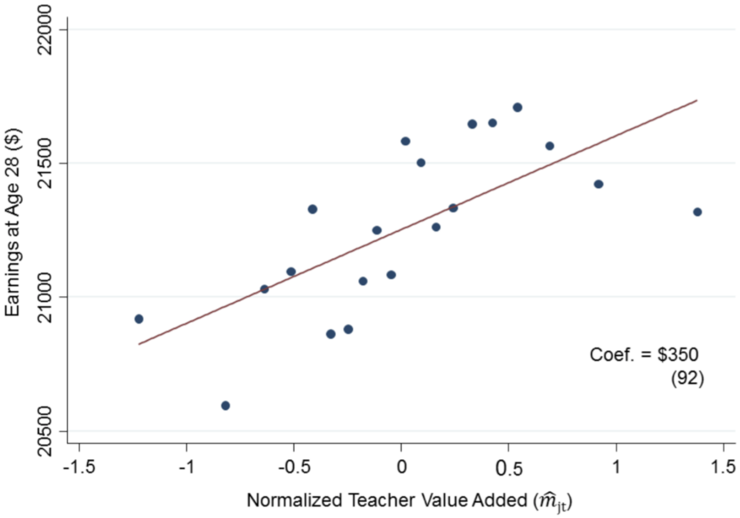
Impacts on Outcomes in Adulthood

- 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

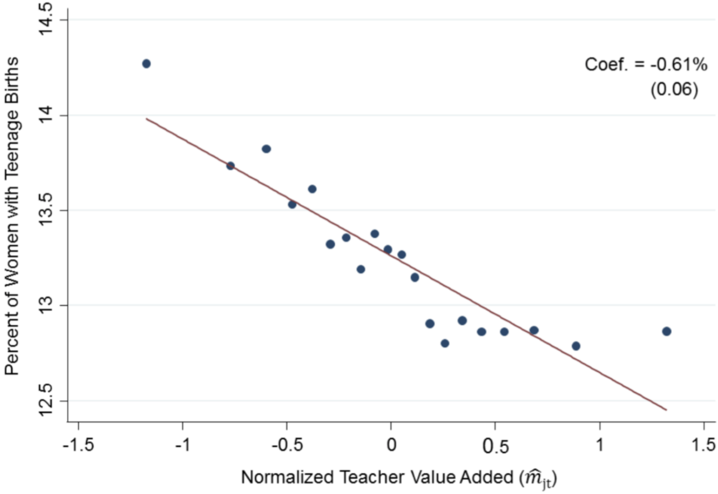
College Attendance at Age 20 vs. Teacher Value-added



Earnings at Age 28 vs. Teacher Value-Added



Women with Teenage Births vs. Teacher Value-Added



Valuing Value-Add?

- Do parents value schools that produce higher test scores/value added?
- Abdulkadiroglu et al. (2019 AER) use rank - ordered school choice lists in NYC
- Individuals report their preferred list of schools
- Use TVA model to estimate causal effects of each school
- Construct "Peer quality" as a school's average predicted test score given the characteristics of its students
- Regress parent rankings on the actual VA

Preferences Regressed on VA + Characteristics

Table 8. Preferences for peer quality and Regents math effects

	Value-added models				Control function models			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A. No controls for school characteristics							
Peer quality	0.416		0.438	0.406	0.407		0.439	0.437
	(0.061)		(0.063)	(0.067)	(0.057)		(0.059)	(0.059)
ATE		0.244	-0.033	-0.022		0.219	-0.051	-0.047
		(0.047)	(0.046)	(0.047)		(0.046)	(0.043)	(0.043)
Match effect				-0.072				-0.172
				(0.047)				(0.054)
N	21684							
	Panel B. With controls for school characteristics							
Peer quality	0.310		0.314	0.286	0.299		0.303	0.308
	(0.060)		(0.059)	(0.060)	(0.056)		(0.056)	(0.056)
ATE		0.157	-0.005	0.005		0.144	-0.008	-0.003
		(0.042)	(0.039)	(0.040)		(0.040)	(0.035)	(0.035)
Match effect				-0.068				-0.142
				(0.039)				(0.044)
N	20200							

Notes: This table reports estimates from regressions of school popularity on peer quality and school effectiveness. School popularity is measured as the estimated mean utility for each school and covariate cell in the choice model from Table 4. Covariate cells are defined by borough, gender, race, subsidized lunch status, an indicator for students above the median of census tract median income, and tercile of the average of eighth grade math and reading scores. Peer quality is constructed as the average predicted Regents math score for enrolled students. Treatment effect estimates are empirical Bayes posterior mean predictions of Regents math effects. Mean utilities, peer quality, and treatment effects are scaled in standard deviation units. Columns (1)-(4) report results from value-added models, while columns (5)-(8) report results from control function models. All regressions include cell indicators and weight by the inverse of the squared standard error of the mean utility estimates. Panel A includes no additional controls, while Panel B controls for the school environment score, violent and

Conclusions

- Education Matters.
- In particular, low-income children tend to systematically benefit from greater public spending
- Evidence suggests information constraints may prevent efficient investment
- Implications for policy?:
  - Limitations of school choice for leading to efficient outcomes?
  - Information interventions complementary to choice-based policy?
  - Lower value to relaxing financial constraints of parents if there are also informational constraints?