I would like to thank Raj Chetty for sharing his slides on education, which comprise much of this lecture.
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
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?
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
Part of the return to education falls on the government budget.

Model setup
- \( l \) is labor effort (unobserved)
- \( y \) is an individual’s production (observed)
- \( \theta \) is an individual’s type (unobserved)
- \( h \) is human capital investment (observed)

Arbitrary production:
\[ y = f(h, l, \theta) \]

Condition for maximizing production for each \( \theta \):
\[ \frac{\partial f}{\partial h} = 1 \]

Utility
\[ u(c, l, h, \theta) \]
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

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

- Production maximized for each $\theta$ iff

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

Utility

$$u = c - \frac{l^{1+\frac{1}{\epsilon}}}{1 + \frac{1}{\epsilon}}$$
Common method for solving uni-dimensional screening problems: Use a Hamiltonian

Government chooses menu of observable variables, \( \{ c(\theta), y(\theta), h(\theta) \} \) to maximize social welfare:

\[
\int u(\theta) \psi(\theta) \, d\theta
\]

where \( u(\theta) = u(c(\theta), \frac{y(\theta)}{\theta \phi(h(\theta))}) \) 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 $h$

$$c(u, l) = u + \frac{y^{1+\frac{1}{e}}}{1 + \frac{1}{e}}$$

- Helpful to have quasilinear utility...why?
IC and Resource Constraints

- **Start with IC constraint:**
  - Define utility a type $\theta$ obtains if they say they are type $\hat{\theta}$:
    \[
    \hat{\nu}(\theta, \hat{\theta}) = u(c(\hat{\theta}), y(\hat{\theta}), h(\hat{\theta}); \theta) = c(\hat{\theta}) - \frac{\left\{ \frac{y(\hat{\theta})}{\phi(h(\hat{\theta}))\theta} \right\}^{1+\frac{1}{\epsilon}}}{1 + \frac{1}{\epsilon}}
    \]
  - IC constraint is (with abuse of notation):
    \[
    u(\theta) = \max_{\hat{\theta}} \hat{\nu}(\theta, \hat{\theta}) \quad \forall \theta
    \]
  - Each type prefers truth-telling

- **Resource Constraint:**
  \[
  \int T(y(\theta)) = \int (y(\theta) - c(\theta)) \, d\theta \geq 0
  \]
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:

\[
u'(\theta) = \frac{\partial u}{\partial \theta} = -\left(\frac{y(\theta)}{\phi(h(\theta))}\right)^{1+\frac{1}{\epsilon}} \frac{d\theta}{d\theta} \theta^{-\left(1+\frac{1}{\epsilon}\right)} \frac{1}{1 + \frac{1}{\epsilon}}\]

\[
= -\left(\frac{y(\theta)}{\phi(h(\theta))}\right)^{1+\frac{1}{\epsilon}} \frac{d\theta}{d\theta} \theta^{-\left(1+\frac{1}{\epsilon}\right)} \frac{1}{1 + \frac{1}{\epsilon}}
\]

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

\[
= \frac{1}{\theta} \left(\frac{y(\theta)}{\phi(h(\theta))}\right)^{1+\frac{1}{\epsilon}}
\]

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

- Think of \( \theta \) 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

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

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

- Now, take derivative wrt \( h \) holding \( l \) and \( v \) constant:
  
  \[
  \frac{\partial H}{\partial h} = \gamma RC \left[ \theta l(\theta) \phi'(h(\theta)) - 1 - \frac{dc}{dh} \big|_{l,v} \right] = 0
  \]
Full Deductibility

- Note that $\frac{dc}{dh} \bigg|_{l,v} = 0$, 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).
Consider general case: \( y(\theta) = \phi(h, \theta) \):

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

Then, IC constraints imply:

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

Note \( \frac{\partial \phi}{\partial \theta} \) does NOT depend on \( h \) when \( \phi = h\theta \).

The general IC constraint now enters derivative of \( H \) wrt \( h \)

Hicksian coefficient of complementarity:

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

Subsidize human capital more (less) than taxes if \( \rho < 1 \) \( (\rho > 1) \)
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
Figure 3: The Effect of Increases in Compulsory Attendance Laws on Average Years of Schooling
## Effect of Years of Schooling on Imprisonment

<table>
<thead>
<tr>
<th></th>
<th>IV Estimates</th>
<th>Control Function</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td><strong>WHITES</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Second-Stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of Schooling</td>
<td>-0.11</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.05)</td>
</tr>
<tr>
<td><strong>First Stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compulsory Attendance = 9</td>
<td>0.278</td>
<td>0.222</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Compulsory Attendance = 10</td>
<td>0.213</td>
<td>0.199</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>Compulsory Attendance ≥ 11</td>
<td>0.422</td>
<td>0.340</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>First Stage F-test (d.o.f. = 3)</td>
<td>52.5</td>
<td>38.6</td>
</tr>
</tbody>
</table>

**Source:** 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
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 education?
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)
Child utility

\[ u_k (c_k, l_k) \]

Earnings given by

\[ y_k = f (l_k, h_k; \theta_k) \]

Budget constraint

\[ c_k \leq y_k + t \]

where \( t \) is transfers from parents
Parents altruistic utility

$$u_p(c_p, l_p, u_k)$$

Budget constraint

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

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 $t$:

$$\frac{\partial f}{\partial h} = 1$$
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
### Table 2—OLS, Effect of Eligibility for Student Benefits on Probability of Attending College by Age 23

<table>
<thead>
<tr>
<th></th>
<th>(1) Difference-in-differences</th>
<th>(2) Add covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deceased father × before</td>
<td>0.182 (0.096)</td>
<td>0.219 (0.102)</td>
</tr>
<tr>
<td>Deceased father</td>
<td>−0.123 (0.083)</td>
<td>Y</td>
</tr>
<tr>
<td>Before</td>
<td>0.026 (0.021)</td>
<td>Y</td>
</tr>
</tbody>
</table>

Source: Dynarski 2003
Many studies documenting imperfect knowledge of returns to education


- Identify high-achieving children on SATs
- Study differences in where children apply based on parental income
Figure 10. Distribution of All High-Achieving Students’ College Applications to Selective Institutions, by Student-College Match³

Percent

College’s median ACT or SAT score minus student’s score (both in percentiles)

High-income
Middle-income
Low-income
**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 \(\frac{\partial U}{\partial G}\)?
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
is uncorrelated with the unobservable determinants of outcomes. The estimating equation is

\[ Y_{ij} = \alpha + \beta_1 HS_{ij} + \beta_2 \text{PRE}_{ij} + \delta X_{ij} + \gamma_j + \varepsilon_i , \]

where \( i \) indexes individuals and \( j \) indexes the family, \( X \) is a vector of family-varying controls, and \( \gamma_j \) is the family fixed effect. \( HS_{ij} \) and \( \text{PRE}_{ij} \) are the estimated effect of Head Start and other preschools, respectively, on the outcomes \( Y_{ij} \). Threats
### Table 3—The Effect of Head Start on Cognitive Test Scores

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

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.
FIGURE 1. Effect of Early Education on High School Graduation Rates

Percentage point change in graduation rates

- **Head Start**
- **Head Start**
  - (mother education less than high school)
- **Head Start** (African American)
- **Head Start** (Hispanic)
- **Perry Preschool** Schweinhart et al. (2005)

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.
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

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
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
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
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)
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

<table>
<thead>
<tr>
<th>Largest funding source</th>
<th>Head Start (1)</th>
<th>Other centers (2)</th>
<th>Other centers attended by $c \rightarrow h$ compliers (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head Start</td>
<td>0.842</td>
<td>0.027</td>
<td>0.038</td>
</tr>
<tr>
<td>Parent fees</td>
<td>0.004</td>
<td>0.153</td>
<td>0.191</td>
</tr>
<tr>
<td>Child and adult care food program</td>
<td>0.011</td>
<td>0.026</td>
<td>0.019</td>
</tr>
<tr>
<td>State pre-K program</td>
<td>0.004</td>
<td>0.182</td>
<td>0.155</td>
</tr>
<tr>
<td>Child care subsidies</td>
<td>0.013</td>
<td>0.097</td>
<td>0.107</td>
</tr>
<tr>
<td>Other funding or support</td>
<td>0.022</td>
<td>0.118</td>
<td>0.113</td>
</tr>
<tr>
<td>No funding or support</td>
<td>0.000</td>
<td>0.003</td>
<td>0.001</td>
</tr>
<tr>
<td>Missing</td>
<td>0.105</td>
<td>0.394</td>
<td>0.375</td>
</tr>
</tbody>
</table>
Table 3: Funding Sources

<table>
<thead>
<tr>
<th>Largest funding source</th>
<th>Head Start (1)</th>
<th>Other centers (2)</th>
<th>Other centers attended by $c \rightarrow h$ compliers (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head Start</td>
<td>0.842</td>
<td>0.027</td>
<td>0.038</td>
</tr>
<tr>
<td>Parent fees</td>
<td>0.004</td>
<td>0.153</td>
<td>0.191</td>
</tr>
<tr>
<td>Child and adult care food program</td>
<td>0.011</td>
<td>0.026</td>
<td>0.019</td>
</tr>
<tr>
<td>State pre-K program</td>
<td>0.004</td>
<td>0.182</td>
<td>0.155</td>
</tr>
<tr>
<td>Child care subsidies</td>
<td>0.013</td>
<td>0.097</td>
<td>0.107</td>
</tr>
<tr>
<td>Other funding or support</td>
<td>0.022</td>
<td>0.118</td>
<td>0.113</td>
</tr>
<tr>
<td>No funding or support</td>
<td>0.000</td>
<td>0.003</td>
<td>0.001</td>
</tr>
<tr>
<td>Missing</td>
<td>0.105</td>
<td>0.394</td>
<td>0.375</td>
</tr>
<tr>
<td>Largest funding source</td>
<td>Head Start (1)</td>
<td>Other centers (2)</td>
<td>Other centers attended by $c \rightarrow h$ compliers (3)</td>
</tr>
<tr>
<td>----------------------------------------------------------</td>
<td>----------------</td>
<td>------------------</td>
<td>----------------------------------------------------------</td>
</tr>
<tr>
<td>Head Start</td>
<td>0.842</td>
<td>0.027</td>
<td>0.038</td>
</tr>
<tr>
<td>Parent fees</td>
<td>0.004</td>
<td>0.153</td>
<td>0.191</td>
</tr>
<tr>
<td>Child and adult care food program</td>
<td>0.011</td>
<td>0.026</td>
<td>0.019</td>
</tr>
<tr>
<td>State pre-K program</td>
<td>0.004</td>
<td>0.182</td>
<td>0.155</td>
</tr>
<tr>
<td>Child care subsidies</td>
<td>0.013</td>
<td>0.097</td>
<td>0.107</td>
</tr>
<tr>
<td>Other funding or support</td>
<td>0.022</td>
<td>0.118</td>
<td>0.113</td>
</tr>
<tr>
<td>No funding or support</td>
<td>0.000</td>
<td>0.003</td>
<td>0.001</td>
</tr>
<tr>
<td>Missing</td>
<td>0.105</td>
<td>0.394</td>
<td>0.375</td>
</tr>
</tbody>
</table>
Table 3: Funding Sources

<table>
<thead>
<tr>
<th>Largest funding source</th>
<th>Head Start (1)</th>
<th>Other centers (2)</th>
<th>Other centers attended by $c \rightarrow h$ compliers (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head Start</td>
<td>0.842</td>
<td>0.027</td>
<td>0.038</td>
</tr>
<tr>
<td>Parent fees</td>
<td>0.004</td>
<td>0.153</td>
<td>0.191</td>
</tr>
<tr>
<td>Child and adult care food program</td>
<td>0.011</td>
<td>0.026</td>
<td>0.019</td>
</tr>
<tr>
<td>State pre-K program</td>
<td>0.004</td>
<td>0.182</td>
<td>0.155</td>
</tr>
<tr>
<td>Child care subsidies</td>
<td>0.013</td>
<td>0.097</td>
<td>0.107</td>
</tr>
<tr>
<td>Other funding or support</td>
<td>0.022</td>
<td>0.118</td>
<td>0.113</td>
</tr>
<tr>
<td>No funding or support</td>
<td>0.000</td>
<td>0.003</td>
<td>0.001</td>
</tr>
<tr>
<td>Missing</td>
<td>0.105</td>
<td>0.394</td>
<td>0.375</td>
</tr>
</tbody>
</table>
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
- Reduced participation in transfer programs
- Savings from reduced grade repetition / Special Ed
Benefits

Increased earnings
Benefits

Increased earnings

Tuition / time savings for parents

Reductions in crime

Health improvements
### 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
Reduced participation in transfer programs
Savings from reduced grade repetition / Special Ed
## Benefits and Costs of Head Start

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Net Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased earnings</td>
<td>Administrative costs</td>
</tr>
<tr>
<td>Tuition / time savings for parents</td>
<td>Reduced funding of competing preschool programs</td>
</tr>
<tr>
<td>Reductions in crime</td>
<td>Extra tax revenue from more productive children</td>
</tr>
</tbody>
</table>
### 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
- Reduced participation in transfer programs
- Savings from reduced grade repetition / Special Ed
### Table 1: Summary of Cost-Benefit Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Tulsa Full-Day Preschool</th>
<th>Tulsa Half-Day Preschool</th>
<th>Oklahoma &amp; Georgia Preschool</th>
<th>Head Start</th>
<th>Perry Preschool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value of earnings gains per child</td>
<td>$27,897</td>
<td>$16,683</td>
<td>$24,094</td>
<td>$14,459</td>
<td>$180,257$</td>
</tr>
<tr>
<td>Value of total benefits per child</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$92,020</td>
</tr>
<tr>
<td>Cost of program per child</td>
<td>$9,118</td>
<td>$4,559</td>
<td>$4,086</td>
<td>$9,173</td>
<td>$20,948</td>
</tr>
<tr>
<td>Net benefit per child</td>
<td>$18,779</td>
<td>$12,124</td>
<td>$20,008</td>
<td>$5,286</td>
<td>$159,309$</td>
</tr>
<tr>
<td>Benefit to cost ratio (earnings only)</td>
<td>3.06</td>
<td>3.66</td>
<td>5.90</td>
<td>1.58$</td>
<td>4.39</td>
</tr>
<tr>
<td>Benefit to cost ratio (all benefits)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>8.60$</td>
</tr>
</tbody>
</table>
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 \( f = 0 \).
Class Size

- Robust evidence that smaller class sizes improve outcomes

  - Angrist and Lavy: test score impacts in Israel
  - Fredriksson et al.: long-term impacts in Sweden
c. Fifth Grade (Math)

- Predicted class size
- Average test scores
- Math score residual
- Enrollment count

Size - function residual

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

<table>
<thead>
<tr>
<th>Dep Var:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Class</td>
<td>4.81</td>
<td>2.02%</td>
<td>$119</td>
<td>-$4</td>
<td>5.06%</td>
</tr>
<tr>
<td></td>
<td>(1.05)</td>
<td>(1.10%)</td>
<td>($97)</td>
<td>($327)</td>
<td>(2.16%)</td>
</tr>
<tr>
<td>Observations</td>
<td>9,939</td>
<td>10,992</td>
<td>10,992</td>
<td>10,992</td>
<td>10,992</td>
</tr>
<tr>
<td>Mean of Dep. Var.</td>
<td>48.67</td>
<td>26.4%</td>
<td>$27,115</td>
<td>$15,912</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Source: Chetty et al. (QJE 2011)
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

$$y_i = \alpha + \beta s_{-i,c} + \epsilon_i$$
Figure 4a: Effect of Early Childhood Class Quality on Own Score

- Own Test Score Percentile
- Class Quality (End-of-Year Peer Scores)

Nathaniel Hendren (Harvard and NBER)
Figure 4c: Effect of Early Childhood Class Quality on Earnings

Mean Wage Earnings, 2005-2007

Class Quality (End-of-Year Peer Scores)
STAR: Impacts of Class Size (Chetty et al 2011)

Figure 5a: Effect of Class Quality on Earnings by Year

Below-Average Class Quality
Above-Average Class Quality

Wage Earnings

Year


$8K $10K $12K $14K $16K $18K

$875

Nathaniel Hendren (Harvard and NBER)
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 (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
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 $t - 1$:

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

  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 $t - s$:

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

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

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

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

Bayesian interpretation: shrinkage based on signal-noise ratio (Kane and Staiger 2008)

Why does this deal with measurement error in $\tilde{A}_{j,t}$?
Distribution of VA Estimates

SD for English = 0.080
SD for Math = 0.116
Test Score Residuals vs. VA in Cross-Section

Score in Year $t$

Estimated Teacher Value-Added in Year $t$

Coef. = 0.998 (0.006)
Let $\gamma$ 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
<table>
<thead>
<tr>
<th>School</th>
<th>Grade</th>
<th>Subject</th>
<th>Year</th>
<th>Teachers</th>
<th>Mean Score</th>
<th>Mean Age 28 Earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>math</td>
<td>1992</td>
<td>Jones, Heckman, ...</td>
<td>-.09</td>
<td>$15K</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>math</td>
<td>1993</td>
<td>Jones, Heckman, ...</td>
<td>-.04</td>
<td>$17K</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>math</td>
<td>1994</td>
<td>Jones, Heckman, ...</td>
<td>-.05</td>
<td>$16K</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>math</td>
<td>1995</td>
<td>Katz, Heckman, ...</td>
<td>0.01</td>
<td>$18K</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>math</td>
<td>1996</td>
<td>Katz, Heckman, ...</td>
<td>0.04</td>
<td>$17K</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>math</td>
<td>1997</td>
<td>Katz, Heckman, ...</td>
<td>0.02</td>
<td>$18K</td>
</tr>
</tbody>
</table>

- Jones switches to a different school in 1995; Katz replaces him
Impact of High VA Teacher Entry on Cohort Test Scores

\[ \Delta \text{Score} = 0.035 \]
\[ (0.008) \]

\[ \Delta \text{TVA} = 0.042 \]
\[ (0.002) \]

\[ p [\Delta \text{score} = 0] < 0.001 \]
\[ p [\Delta \text{score} = \Delta \text{TVA}] = 0.34 \]

Number of Events = 1135
Impact of High VA Teacher Exit on Cohort Test Scores

\[ \Delta \text{Score} = -0.045 \quad (0.008) \]
\[ \Delta \text{TVA} = -0.042 \quad (0.002) \]

\[ p [\Delta \text{score} = 0] < 0.001 \]
\[ p [\Delta \text{score} = \Delta \text{TVA}] = 0.66 \]

Number of Events = 1115
Changes in Mean Scores vs. Changes in Mean Teacher VA

![Graph showing the relationship between Changes in Mean Scores and Changes in Mean Teacher Value-Added. The coefficient is 0.974 with a standard error of 0.033.](image)

Coef. = 0.974
(0.033)
<table>
<thead>
<tr>
<th>Control Vector</th>
<th>Quasi-Experimental Estimate of Bias (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>2.58 (3.34)</td>
</tr>
<tr>
<td>Student-level lagged scores</td>
<td>4.83 (3.29)</td>
</tr>
<tr>
<td>Non-score controls only</td>
<td>45.39 (2.26)</td>
</tr>
<tr>
<td>No controls</td>
<td>65.58 (3.73)</td>
</tr>
</tbody>
</table>
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
Figure 2: Visual instrumental variables tests for bias

A. Lagged score

Forecast coef.: 0.864
Omnibus p-val.: <0.001
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

Normalized Teacher Value Added ($\hat{m}_{jt}$) vs. Percent in College at Age 20

Coef. = 0.82% (0.07)
Earnings at Age 28 vs. Teacher Value-Added

Coef. = $350 (92)
Women with Teenage Births vs. Teacher Value-Added

Coefficient: -0.61% (0.06)

Percent of Women with Teenage Births vs. Normalized Teacher Value Added ($\hat{m}_{jt}$)
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))
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