Econ 2450B, Topic 4: Education

Nathaniel Hendren

Harvard

Spring, 2023
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
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 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

$$u(c, l, h, \theta)$$
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
  \[
  y = \theta l \phi (h)
  \]
  - Production maximized for each \( \theta \) iff
  \[
  \theta l \phi' (h) = 1
  \]

- Utility
  \[
  u = c - \frac{l^{1+\frac{1}{\varepsilon}}}{1 + \frac{1}{\varepsilon}}
  \]
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 \( u(\theta) = u \left( c(\theta), \frac{y(\theta)}{\theta \phi(h(\theta))} \right) \) 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 $h$

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

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

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

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

- IC constraint is (with abuse of notation):

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

- 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:

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

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

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

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

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

- 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 $\frac{dc}{dh}\big|_{l,v} = 0$, 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).
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:

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

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:

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

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
  \[
  \max_{c,y,h} u(c, y, h; \theta) \quad s.t. \quad B(c, y, h) \leq 0
  \]
  where \( B \) is the same for all people but utility varies with \( \theta \).

- Show that if
  \[
  u(c, y, h; \theta) = \tilde{u}(v(c, h), y; \theta) \implies h, c \text{ same tax rate}
  \]
  implies tax \( h \) as consumption

- Similarly if
  \[
  u(c, y, h; \theta) = \tilde{u}(c, v(y, h); \theta) \implies h, y \text{ same tax rate}
  \]
  implies tax \( y \) as pre-tax income (i.e. don’t tax it!)

- Turns out we didn’t need the Hamiltonian at all!
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)
- Imply 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?
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
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
Sample,

You are an academically excellent student who has worked hard for your considerable achievements. Congratulations!

As you are thinking about life after your senior year, we hope that you will consider and apply to the University of Michigan. You can put your grand imagination to work here in Ann Arbor—an attractive college town—where you will be surrounded by the professors and resources to help you on your journey.

Today, I’m excited to make you an outstanding offer. If you apply to U-M and are admitted, we are prepared to cover the full cost of your in-state tuition and fees for four years of study at U-M’s Ann Arbor campus. That’s an approximate $100,000 value to you and your family. Furthermore, after a review of your financial aid applications, you will likely be eligible for additional aid to cover costs of housing, textbooks, and other expenses.

Take a look at these materials and discuss them with your family. Explore more about the University at admissions.umich.edu.

We are eager to receive your application by our Early Action deadline of November 1 or Regular Decision deadline of February 1.

Go Blue!

Mark Schlissel
President
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

![Bar chart showing the average fraction of students who applied to UM by decile of school-level rate of application to UM in 2015, with bars for control schools and treatment schools.](chart.png)
(b) Enrolled

![Bar chart showing average fraction of students enrolled at UM by decile of school-level rate of application to UM in 2015. The chart compares control schools (yellow bars) and treatment schools (blue bars). There is a notable difference in enrollment rates between the two groups, with treatment schools having higher enrollment rates in the top deciles.](image-url)
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

**Average Quarterly Earnings ($)**

- **Impacts**: +1,794
- **Follow-up Quarter**: Q0, Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q11
- **Quarters**: Q-2, Q-1, Q0, Q1, Q2, Q3

**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.
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
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

Actual Change in Per Pupil Spending by Predicted Reform Induced Change in Spending (years 3 through 8)

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 spending changes) and duration of "school age", respectively).
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
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 of household individual's school-age years of exposure to school finance reform and the district's change in per-pupil spending induced by reform).
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
Robust evidence that smaller class sizes improve outcomes

  - Angrist and Lavy: test score impacts in Israel
  - Fredriksson et al.: long-term impacts in Sweden
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)

### Table

<table>
<thead>
<tr>
<th>Dep Var:</th>
<th>Test Score</th>
<th>College in 2000</th>
<th>College Quality</th>
<th>Wage Earnings</th>
<th>Summary Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<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)
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
Figure 4a: Effect of Early Childhood Class Quality on Own Score
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
Figure 4c: Effect of Early Childhood Class Quality on Earnings

Mean Wage Earnings, 2005-2007

Class Quality (End-of-Year Peer Scores)

$14.5K
$15.0K
$15.5K
$16.0K
$16.5K
$17.0K

-20 -10 0 10 20
Figure 5a: Effect of Class Quality on Earnings by Year

Wage Earnings

$875

$8K

$10K

$12K

$14K

$16K

$18K


Year

Below-Average Class Quality  Above-Average Class Quality

Nathaniel Hendren (Harvard)
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 \( \left( \hat{\mu}_{j,t+1} \right) \) 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 \( j \):
   \[
   \hat{\mu}_{j,t+1} = \sum_{s=1}^{t} \psi_s \bar{A}_{js}
   \]
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:

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

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

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

SD for English = 0.080
SD for Math = 0.116

Density vs. Teacher Value-Added

- English
- Math

Teacher Value-Added:
-0.3 to 0.3
Density:
0 to 6
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
### Teacher Switchers in School-Grade-Subject-Year Data

<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

\[ \Delta \text{Score} = 0.035 \]
\[ \Delta \text{TVA} = 0.042 \]

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

Number of Events = 1135

School-Grade-Cohort Mean Test Score

Year Relative to Entry of High Value-Added Teacher

- Score in Current Grade
- Score in Previous Grade
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

Coef. = 0.974
      (0.033)
### Estimates of Forecast Bias with Alternative Control Vectors

<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>
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

![Graph showing the relationship between percent in college at age 20 and normalized teacher value added. The coefficient is 0.82% (0.07).]
Women with Teenage Births vs. Teacher Value-Added

![Graph showing the relationship between percent of women with teenage births and normalized teacher value added. The graph includes data points and a trend line, with a coefficient of -0.61% (0.06).]
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

<table>
<thead>
<tr>
<th></th>
<th>Value-added models</th>
<th>Control function models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td><strong>Panel A. No controls for school characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer quality</td>
<td>0.416 (0.061)</td>
<td>0.438 (0.063)</td>
</tr>
<tr>
<td>ATE</td>
<td>0.244 (0.047)</td>
<td>-0.033 (0.046)</td>
</tr>
<tr>
<td>Match effect</td>
<td></td>
<td>-0.072 (0.047)</td>
</tr>
<tr>
<td>N</td>
<td>21684</td>
<td></td>
</tr>
<tr>
<td><strong>Panel B. With controls for school characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer quality</td>
<td>0.310 (0.060)</td>
<td>0.314 (0.059)</td>
</tr>
<tr>
<td>ATE</td>
<td>0.157 (0.042)</td>
<td>-0.005 (0.039)</td>
</tr>
<tr>
<td>Match effect</td>
<td></td>
<td>-0.068 (0.039)</td>
</tr>
<tr>
<td>N</td>
<td>20200</td>
<td></td>
</tr>
</tbody>
</table>

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?