Using Big Data to Solve Economic and Social Problems

Professor Raj Chetty
Head Section Leader: Gregory Bruich, Ph.D.
Part II
Education
Education and Upward Mobility

- Education is widely viewed as one of the most scalable pathways to upward mobility.

- But there is growing concern that education no longer provides a strong pathway to opportunity in the U.S.
  - U.S. students perform worse on standardized tests on average than in many European countries despite higher spending on schools.
  - Sharp differences in quality of schools across districts.
  - Rising costs of college → lack of access for low-income students.
  - Concern that some colleges (e.g., for-profit institutions) may not produce good outcomes.
How can we improve education in America?

– Traditionally, measuring impacts of education systematically was difficult

– Administrative data from colleges and school districts are giving us a more scientific understanding of the “education production function”

Start with higher education, then turn to K-12 education

How Do Colleges Shape Income Mobility in the U.S.?

- How does the higher education system affect intergenerational income mobility in the U.S.?
  - In principle, higher education can provide a pathway to upward mobility that is not directly shaped by the neighborhood where a child happens to grow up.
  - But if children from higher-income families tend to attend better colleges, higher education system may not promote mobility.
  - Colleges could actually increase intergenerational persistence of income if disparities in college attendance are sufficiently large.
Effect of Higher Education System on Mobility

- Effect of higher education system on mobility depends upon three factors:
  1. [Inputs] Parental income distributions by college
  2. [Outputs] Students’ earnings outcomes conditional on parental income by college
  3. [Causal share] Portion of variation in students’ earnings outcomes that is due to colleges’ causal effects
Estimating the Three Parameters: Data

- Chetty et al. (2017) estimate these three parameters using data covering all college students in the U.S. from 1999-2013 (30 million students)

- Combine information from three sources to construct an anonymized dataset:
  1. Parental and Student Income from income tax records
  2. College attendance from 1098-T tax data and Pell grant data
  3. SAT scores from College Board

- Note: all statistics are based on college attendance (not completion)
Parents’ Income Distributions by College: Income Segregation in the American Higher Education System
Measuring Parents’ Incomes

- Parent income: average pre-tax household income during five year period when child is aged 15-19

- Focus on percentile ranks, ranking parents relative to other parents with children in same birth cohort
Parent Household Income Distribution
For Parents with Children in 1980 Birth Cohort

20th Percentile = $25k
Median = $60k
60th Percentile = $74k
80th Percentile = $111k
99th Percentile = $512k
College Attendance Rates vs. Parent Income Percentile

Percentage in College vs. Parent Rank
Parent Income Distribution at Harvard
Classes of 2002-2004

Percent of Students

Parent Income Quintile

1  3.0%
2  5.3%
3  8.1%
4  13.2%
5  70.3%
Parent Income Distribution at Harvard
Classes of 2002-2004

- Top 1%
  - 15.4%

- Parent Income Quintile Distribution
  - Quintile 1: 3.0%
  - Quintile 2: 5.3%
  - Quintile 3: 8.1%
  - Quintile 4: 13.2%
  - Quintile 5: 70.3%
Probability of attending Harvard is **103 times** higher for children from the top 1% compared to the bottom 20%.
Parental Income Distribution of Students at Highly Selective Colleges

- Stanford
- Harvard
- Yale
- Princeton
- MIT
- Cal Tech
- Amherst
- Swarthmore

The chart shows the distribution of parental income for students at highly selective colleges, with quintiles indicating different income levels.
Parent Income Distribution by Percentile
Ivy Plus Colleges (Ivy League plus Stanford, MIT, Duke, and Chicago)

3.8% of students from bottom 20%
14.5% of students from top 1%
Parent Income Distributions by Quintile for 1980-82 Birth Cohorts
At Selected Colleges

Harvard University
Parent Income Distributions by Quintile for 1980-82 Birth Cohorts
At Selected Colleges

- Harvard University
- UC Berkeley
Parent Income Distributions by Quintile for 1980-82 Birth Cohorts
At Selected Colleges

Harvard University
UC Berkeley
SUNY-Stony Brook
Parent Income Distributions by Quintile for 1980-82 Birth Cohorts
At Selected Colleges

- Harvard University
- UC Berkeley
- SUNY-Stony Brook
- Glendale Community College

percent of students

Parent Income Quintile

1 2 3 4 5

At Selected Colleges
Parental Income Segregation Across Colleges

- Sharp differences in parental income distributions across colleges → there is significant segregation across colleges

- Useful benchmark to quantify magnitude: compare to degree of segregation across neighborhoods

- Common perception: colleges foster greater interaction between children from different socioeconomic backgrounds than places in which they grew up
Parental Income Distribution of Peers of Children from Top Quintile

![Graph showing the distribution of parental income for peers across different quintiles. The x-axis represents the quintile of peers' parent income, and the y-axis shows the share of peers (%). The bars indicate the percentage of peers in pre-college neighborhoods and colleges for each quintile.](image-url)
Parental Income Distribution of Peers of Ivy-Plus College Students from Top Quintile

![Graph showing the distribution of peers' parental income quintiles compared to pre-college neighborhoods and colleges. The x-axis represents Peers' Parent Income Quintile, with levels 1 to 5. The y-axis shows the share of peers (%). The graph indicates a significant concentration of peers in the highest income quintile, particularly in the top two quintiles compared to pre-college neighborhoods and colleges.]
Trends in Income Segregation

- Preceding estimates are based on children born between 1980-82, who attended college in the early 2000s

- Substantial changes in higher education system since that time, e.g. substantial changes in financial aid and tuition policies

- How has income segregation across colleges changed in recent years?
Fraction of Peers from the Top Quintile for Children from the Top Quintile

- Parent income segregation across pre-college residential ZIP codes
- Parent income segregation across colleges
Trends in Low-Income Access from 2000-2011 at Selected Colleges

Year When Child was 20

Percent of Parents in the Bottom Quintile


Harvard
Stanford
UC Berkeley
SUNY Stony Brook
Glendale CC
Outcomes: Students’ Earnings Distributions
Students’ Earnings Outcomes

- Measure children’s individual earnings in their mid-30s
  - Define percentile ranks by ranking children relative to others in same birth cohort
Distribution of Children’s Individual Labor Earnings at Age 34
1980 Birth Cohort

- p20 = $1k
- p50 = $28k
- p80 = $58k
- p99 = $197k

Density

Individual Earnings ($)

0 50000 100000 150000
Students’ Earnings Outcomes
Columbia vs. SUNY-Stony Brook

- Columbia
- SUNY-Stony Brook

Percent of Students

Parent Income Quintile

Parent Income Quintile: 1, 2, 3, 4, 5

Students’ Earnings Outcomes compared between Columbia and SUNY-Stony Brook based on parent income quintiles.
Students’ Earnings Outcomes

- Key lesson: most of the gap in outcomes between children from low vs. high-income families is explained by differences between rather than within colleges.

- Raises possibility that reallocating students across colleges could potentially have a significant impact on intergenerational mobility.
  - If gap in outcomes by parental income were large even within a given college, there would be little scope to have an impact through changes in college admissions policies.
Differences in Mobility Rates Across Colleges

- We can combine data on parents’ incomes and students’ outcomes to characterize colleges’ mobility rates
  - At which colleges in America do the largest number of children come from poor families and end up in the upper middle class?
Students’ Earnings Outcomes
Columbia vs. SUNY-Stony Brook

Top-Quintile Outcome Rate: Fraction of Students who Reach Top Quintile = 51%

Fraction of Parents from Bottom Quintile (<$25K) = 16%
Measuring Mobility Rates

- Define a college’s mobility rate (MR) as the fraction of its students who come from bottom quintile and end up in top quintile.

- Mobility rate is:

\[
\text{Mobility Rate} = \text{Low-Inc. Access} \times \text{Top-Quintile Rate}
\]

At SUNY:

\[
8.4\% = 16\% \times 51\%
\]

- Frac. of Parents in Q1 and Children in Q5
- Frac. of Parents in Q1
- Frac. of Students who reach Q5 Given Parents in Q1
Mobility Rates: Top-Quintile Outcome Rate vs. Access by College

Top-Quintile Outcome Rate: \( P(\text{Child in Q5} \mid \text{Par in Q1}) \)

Access: Percent of Parents in Bottom Quintile

SUNY-Stony Brook

Columbia
Mobility Rates: Top-Quintile Outcome Rate vs. Access by College

Ivy Plus Colleges (Avg. MR = 2.2%)
Mobility Rates: Top-Quintile Outcome Rate vs. Access by College

Access: Percent of Parents in Bottom Quintile

Top-Quintile Outcome Rate: \( P(\text{Child in Q5} \mid \text{Par in Q1}) \)

- **Ivy Plus Colleges** (Avg. MR = 2.2%)
  - Princeton
  - MIT
  - Stanford
  - Columbia
  - Harvard
  - Yale
  - Brown
  - Duke
  - Chicago State
  - University Of California, Berkeley
  - University Of Michigan - Ann Arbor
  - University Of North Carolina - Chapel Hill
  - University Of New Mexico

- **Public Flagships** (Avg. MR = 1.7%)
  - University Of California, Berkeley
  - University Of Michigan - Ann Arbor
Top 10 Colleges in America By Bottom-to-Top Quintile Mobility Rate
Fraction of Students who come from Bottom Fifth and End up in Top Fifth

<table>
<thead>
<tr>
<th>College</th>
<th>Mobility Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cal State-Los Angeles</td>
<td>9.9%</td>
</tr>
<tr>
<td>Pace University</td>
<td>8.4%</td>
</tr>
<tr>
<td>SUNY-Stony Brook</td>
<td>8.4%</td>
</tr>
<tr>
<td>Technical Career Institutes</td>
<td>8.0%</td>
</tr>
<tr>
<td>U. Texas-Pan American</td>
<td>7.6%</td>
</tr>
<tr>
<td>CUNY System</td>
<td>7.2%</td>
</tr>
<tr>
<td>Glendale Comm. Coll.</td>
<td>7.1%</td>
</tr>
<tr>
<td>South Texas College</td>
<td>6.9%</td>
</tr>
<tr>
<td>Cal State Poly-Pomona</td>
<td>6.8%</td>
</tr>
<tr>
<td>U. Texas-El Paso</td>
<td>6.8%</td>
</tr>
<tr>
<td>Harvard</td>
<td>1.8%</td>
</tr>
<tr>
<td>Avg. College in the U.S.</td>
<td>1.9%</td>
</tr>
</tbody>
</table>