Changes in the College Mobility Pipeline Since 1900

Zachary Bleemer
Princeton University and NBER

Sarah Quincy
Vanderbilt University and NBER

July 2024
College-going became regressive over the 20th century.
Introduction

Research question: when and why did the labor market returns to college-going become positively correlated with childhood parental income?

What we do:

1. Measure changes in the observational and causal return to college enrollment by parental income for men since 1900.

2. Measure and decompose the contributions of changes in the composition and value-added of collegiate majors and institutions since 1900.

3. Simulate magnitude of collegiate regressivity in mediating intergenerational income transmission.
Main Findings:

- Collegiate regressivity starts rising in the 1960s. This does not align with many high-level US higher education trends (e.g. rising enrollment, rising tuition). **Differential selection plays a secondary role.**

- Three trends explain 70% of the trend toward collegiate regressivity:
  1. The less-selective and public institutions that disproportionately enroll lower-income students have seen large declines in value-added since 1960. **Shifts between four-year institutions play a secondary role.**
  2. Lower-income students have been increasingly diverted to community colleges since the 1980s.
  3. Higher-income students have disproportionately exited humanities majors and flowed into computer science since the 2000s.

- College-going provided equitable returns before 1960, but collegiate regressivity now mediates 20 percent of intergenerational income transmission.
Introduction

Main Findings:

- Collegiate regressivity starts rising in the 1960s. This does not align with many high-level US higher education trends (e.g. rising enrollment, rising tuition). Differential selection plays a secondary role.

- Three trends explain 70% of the trend toward collegiate regressivity:
  1. The less-selective and public institutions that disproportionately enroll lower-income students have seen large declines in value-added since 1960.
     - Shifts between four-year institutions play a secondary role.
  2. Lower-income students have been increasingly diverted to community colleges since the 1980s.
  3. Higher-income students have disproportionately exited humanities majors and flowed into computer science since the 2000s.

- College-going provided equitable returns before 1960, but collegiate regressivity now mediates 20 percent of intergenerational income transmission.
Main Findings:

- Collegiate regressivity starts rising in the 1960s. This does not align with many high-level US higher education trends (e.g. rising enrollment, rising tuition). Differential selection plays a secondary role.

- Three trends explain 70% of the trend toward collegiate regressivity:
  1. The less-selective and public institutions that disproportionately enroll lower-income students have seen large declines in value-added since 1960.
     - Shifts between four-year institutions play a secondary role.
  2. Lower-income students have been increasingly diverted to community colleges since the 1980s.
  3. Higher-income students have disproportionately exited humanities majors and flowed into computer science since the 2000s.

- College-going provided equitable returns before 1960, but collegiate regressivity now mediates 20 percent of intergenerational income transmission.
Main Findings:

- Collegiate regressivity starts rising in the 1960s. This does not align with many high-level US higher education trends (e.g. rising enrollment, rising tuition). Differential selection plays a secondary role.

- Three trends explain 70% of the trend toward collegiate regressivity:
  1. The less-selective and public institutions that disproportionately enroll lower-income students have seen large declines in value-added since 1960.
     - Shifts between four-year institutions play a secondary role.
  2. Lower-income students have been increasingly diverted to community colleges since the 1980s.
  3. Higher-income students have disproportionately exited humanities majors and flowed into computer science since the 2000s.

- College-going provided equitable returns before 1960, but collegiate regressivity now mediates 20 percent of intergenerational income transmission.
Main Findings:

- Collegiate regressivity starts rising in the 1960s. This does not align with many high-level US higher education trends (e.g. rising enrollment, rising tuition). Differential selection plays a secondary role.

- Three trends explain 70% of the trend toward collegiate regressivity:
  1. The less-selective and public institutions that disproportionately enroll lower-income students have seen large declines in value-added since 1960. Shifts between four-year institutions play a secondary role.
  2. Lower-income students have been increasingly diverted to community colleges since the 1980s.
  3. Higher-income students have disproportionately exited humanities majors and flowed into computer science since the 2000s.

- College-going provided equitable returns before 1960, but collegiate regressivity now mediates 20 percent of intergenerational income transmission.
Main Findings:

- Collegiate regressivity starts rising in the 1960s. This does not align with many high-level US higher education trends (e.g. rising enrollment, rising tuition). Differential selection plays a secondary role.

- Three trends explain 70% of the trend toward collegiate regressivity:
  1. The less-selective and public institutions that disproportionately enroll lower-income students have seen large declines in value-added since 1960. Shifts between four-year institutions play a secondary role.
  2. Lower-income students have been increasingly diverted to community colleges since the 1980s.
  3. Higher-income students have disproportionately exited humanities majors and flowed into computer science since the 2000s.

- College-going provided equitable returns before 1960, but collegiate regressivity now mediates 20 percent of intergenerational income transmission.
Contributions to literature

- Add university heterogeneity to the long literature on the relationship between education, inequality, and economic mobility.

- Link the large microeconomic literature on heterogeneity in returns to higher education to long-run macro trends.
Contributions to literature

- Add university heterogeneity to the long literature on the relationship between education, inequality, and economic mobility.

- Link the large microeconomic literature on heterogeneity in returns to higher education to long-run macro trends.
  
  - We extend estimates of **major** value-added (and composition by parental income) by 50 (35) years: Arcidiacono 2004; Fairlie, Hoffman, and Oreopoulos 2014; Altonji et al 2016; Kirkeboen, Leuven, and Mogstad 2016; Arcidiacono, Aucejo, and Hotz 2016; Patnaik et al 2022; Bleemer and Mehta 2023, 2024.
  
  - We extend estimates of **institutional** value-added (and composition by parental income) by 35 (50) years: Chetty et al 2020; Zimmerman 2019; Dynarski et al 2021, 2023; Bleemer 2021, 2022; Mountjoy and Hickman 2021; Mountjoy 2022; Abramitzky et al 2022; Michelman et al 2023; Black et al 2023; Chetty, Deming, and Friedman 2023.
Data-Collection Goal: Collect longitudinal surveys covering the past 100 years in the U.S., observing:

1. Parental income while child is in high school
2. Child test score in high school
3. Child income in early 30s
4. Child postsecondary institution (if attended college)
5. Child college major (if attended college)

We restrict our main sample to high school graduates to measure college returns relative to the HS baseline.

National representativeness is preferable, but we use state records when otherwise unavailable.
- Linked using the Census Linking Project (Abramitzky et al 2020).

WWII draft cards
- WWII enrollment cards (AGCT and education) from the National Archives.
- Linked to 1950 education following ABE on name, birth year, and birth state.

Sample Size: 329,000 overall; 2,804 with AGCT matches.
1947 survey of all US college graduates with last names beginning with “FA”.
- Contains income, major, and institution, but not family income*.
- Sample Size: 1,818.
CPS Occupational Change in a Generation study sampled from CPS participants.

Contains parental income proxies (occupation, education, and geography), education, and income, along with institution and major in the latter survey.

Sample Size: 1,711 (‘62) and 2,778 (‘73) for ages 30-35; 6,411 across cohorts with observable collegiate institutions and majors.
Wisconsin Longitudinal Survey sampled from 1957 Wisconsin high school graduates.

- Contains parental income, IQ, income, major, and institution.
- Sample Size: 3,297.

Contains parental income, IQ, income, major, and institution.

Sample size: 37,751, which is large enough to estimate institutional value-added for 403 institutions and 43 state-level pairs.
Seven U.S. national longitudinal surveys have collected parental income and relevant fields:

- NLSM (1,171), NLS72 (3,865), NLSY79 (1,938), and NLSY97 (2,690) have test scores, child income, and major.
- NELS (4,570) and ELS (4,212) have test scores and major.
- ADD Health (1,279) has child income.
Nationally-representative panel longitudinal survey of US households.
Sample size: 1,991.
Comprehensive University of California administrative data for six campuses: Berkeley, Davis, Irvine, Riverside, UCSB, and UCSC.

Census-linked, Census tract, or Zip code income (from Census and IRS SOI), major, and institution. Sample size: 439,719.
Institution- and major-level Pell funding, which proxies low-income enrollment. Sample size: 159,741 + 18,135.
Male college enrollment rose following WWII but has been unchanged since the late 1960s...
...and higher-income children have always been more likely to enroll, especially following the GI bill (Stanley 2003).
We estimate the observational wage return to higher education by family income:

\[ \text{Wage}_{it} = \zeta_t + \alpha_t \text{FamInc}_{it} + \beta_t \text{College}_{it} + \delta_t (\text{FamInc}_{it} \times \text{College}_{it}) + \epsilon_{it} \]  

(1)

over high school graduates \( i \) in year \( t \), where:

- \( \text{Wage}_{it} \) is measured in rank or log $ between ages 30 and 35;
- \( \text{FamInc}_{it} \) is measured in rank between ages 14 and 17 and centered;
- \( \text{College}_{it} \) indicates at least one year of college; and
- Estimation is weighted by sample weights and standard errors are robust.
Average returns to college attendance haven’t changed much in rank over time...
The Regressivity of U.S. Higher Education

Differential Return to College Enrollment by SES ($\delta_t$)

... but have become more regressive since the 1950s.
The Regressivity of U.S. Higher Education

Differential Return to College Enrollment by SES ($\delta_t$)

... but have become more regressive since the 1950s.

... but have become more regressive since the 1950s.
What Explains the Rise in Observational Regressivity?

Let \( p_t(i) = p_t(a_i, u_i, m_i, PI_i) \) denote \( i \)'s college-going premium, where:

- \( a_i \): pre-college aptitude
- \( u_i \): enrollment institution
- \( m_i \): college major
- \( PI_i \): parental income

Then define regressivity between top (\( T \)) and bottom (\( B \)) tercile parental incomes:

\[
D_t \equiv \Delta_q \left[ E[p_t|q] \right] = E[p_t|T] - E[p_t|B]
\]

We define \( v^x_t(j) \) as the value of a given \( j \) for \( x \in \{ A, U, M \} \) and decompose \( D_t \) into the following seven components:

\[
D_t = \sum_{x \in \{ A, U, M \}} \left( \int_j v^x_0(j) \Delta_q [P_t(j|q)] \, dj + \int_j \Delta_q [P_t(j|q)] (v^x_t(j) - v^x_0(j)) \, dj \right) + \epsilon_t
\]
What Explains the Rise in Observational Regressivity?

Let $p_t(i) = p_t(a_i, u_i, m_i, Pl_i)$ denote $i$’s college-going premium, where:

- $a_i$: pre-college aptitude
- $u_i$: enrollment institution
- $m_i$: college major
- $Pl_i$: parental income

Then define regressivity between top ($T$) and bottom ($B$) tercile parental incomes:

$$D_t = \Delta_q \left[ E[p_t|q] \right] = E[p_t|T] - E[p_t|B]$$

We define $v_t^x(j)$ as the value of a given $j$ for $x \in \{A, U, M\}$ and decompose $D_t$ into the following seven components:

$$D_t = \sum_{x \in \{A, U, M\}} \left( \int_j v_0^x(j) \Delta_q [P_t(j|q)] dj + \int_j \Delta_q [P_t(j|q)] (v_t^x(j) - v_0^x(j)) dj \right) + \epsilon_t$$
What Explains the Rise in Observational Regressivity?

Let $p_t(i) = p_t(a_i, u_i, m_i, PI_i)$ denote $i$’s college-going premium, where:

- $a_i$: pre-college aptitude
- $u_i$: enrollment institution
- $m_i$: college major
- $PI_i$: parental income

Then define regressivity between top ($T$) and bottom ($B$) tercile parental incomes:

$$D_t = \Delta_q \left[ E[p_t | q] \right] = E[p_t | T] - E[p_t | B]$$

We define $v_t^x(j)$ as the value of a given $j$ for $x \in \{A, U, M\}$ and decompose $D_t$ into the following seven components:

$$D_t = \sum_{x \in \{A, U, M\}} \left( \int_j v_0^x(j) \Delta_q [P_t(j | q)] \, dj + \int_j \Delta_q [P_t(j | q)] (v_t^x(j) - v_0^x(j)) \, dj \right) + \epsilon_t$$
The labor market value of pre-college aptitude has risen (conditional on family income rank and child education)...

Return to Pre-College Human Capital
Differential Test Score Selection into College Enrollment by SES ($\delta_t$)

...but selection into college-going has not changed since the 1960s.
Slope comes from log-dollars version of the change in regressivity over time, converting to difference between top and bottom income tercile.
Explanation I: Selection

Decomposition of University Regressivity

Test Composition (solid color): \[ \int_a \nu_0^A(a) \Delta_q \left[ P_t(a|q) \right] \, da. \]
Explanation I: Selection

Decomposition of University Regressivity

Test Wage Value (stripe): \( \int_a \Delta_q \left[ P_t(a|q) \right] \left( v_t^A(a) - v_0^A(a) \right) da \).
Explanation II: College Majors

Wage Value of College Majors, Relative to HS Grad

Year Turned 18

- Humanities
- Social Science
- Natural Science
- Engineering
- Business
- Agriculture
## Explanation II: College Majors

### Selection-on-Observables Forecast Coefficients of Average Wages by Major

<table>
<thead>
<tr>
<th>Add’l Cov.:</th>
<th>Same Sample</th>
<th>Split Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fam. Inc.</td>
<td>+AFQT</td>
</tr>
<tr>
<td>A. Discipline</td>
<td>1.02 (0.02)</td>
<td>1.03 (0.02)</td>
</tr>
<tr>
<td>Premiums</td>
<td>7 842</td>
<td>7 418</td>
</tr>
<tr>
<td>B. Det. Major</td>
<td>1.00 (0.06)</td>
<td>1.01 (0.08)</td>
</tr>
<tr>
<td>Premiums</td>
<td>14 753</td>
<td>14 372</td>
</tr>
</tbody>
</table>

**Takeaway:** $\tilde{w}_m$ seems to be a reasonable proxy for $ATT_m$. 
Explanation II: College Majors

Difference in Major Premium Between Students from Bottom- and Top-Tercile Par. Inc.
Explanation II: College Majors

Decomposition of University Regressivity

Discipline Composition (solid color): \( \sum_{m} v_0^M(m) \Delta_q \left[ P_t(m|q) \right] \).
Explanation II: College Majors

Decomposition of University Regressivity

Disc. Value-Added (stripe): \( \sum_m \Delta_q \left[ P_t(m|q) \right] \left( v_t^M(m) - v_0^M(m) \right) \)

Zachary Bleemer and Sarah Quincy
Explanation II: College Majors

Decomposition of University Regressivity

Major Composition II (cross-hatches): Add detailed majors.
Explanation II: College Majors

Decomposition of Recent Rising Major Regressivity by Discipline

Majors by Parent Income

Zachary Bleemer and Sarah Quincy

Changes in the College Mobility Pipeline Since 1900

July 2024

19 / 26
Explanation III: Institutions

Value-Added Statistics Over Time

Slope: 0.29

1963 Value-Added

1996 Value-Added

1963 Value-Added Map

Zachary Bleemer and Sarah Quincy

Changes in the College Mobility Pipeline Since 1900

July 2024
Relationship between average parental income and log wage value-added goes from -0.002 to 0.02 per 10 ranks.
### Explanation III: Institutions

**Selection-on-Observables Forecast Coefficients of Institutional Value-Added**

<table>
<thead>
<tr>
<th>Add'l Cov.:</th>
<th>Base.</th>
<th>+Tests</th>
<th>+Grades</th>
<th>+HS FE</th>
<th>+Extra.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inst. FE</td>
<td>0.97</td>
<td>0.95</td>
<td>0.82</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.006)</td>
<td>(0.027)</td>
<td>(0.026)</td>
<td></td>
</tr>
<tr>
<td>Obs.</td>
<td>396</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Stg. Obs.</td>
<td>22,099</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Split Sample**

| Inst. FE    | 0.552 | 0.523 | 0.511 | 0.385 | 0.389 |
|             | (0.041)| (0.041)| (0.040)| (0.051)| (0.049)|
| Obs.        | 396   |        |       |       |
| 1st Stg. Obs.| 10,956|        |       |       |

**Takeaway:** With large-enough samples, 60-80 percent of institutional VA is treatment effect.
Explanation III: Institutions

Difference in Inst. VA Between Students from Bottom- and Top-Tercile Par. Inc.

Graph showing the difference in institutional VA between students from bottom- and top-tercile parental income. The x-axis represents the year from 1940 to 2020, and the y-axis represents the log of dollars with values from -0.08 to -0.02. The graph includes two lines: one for mid-century institutional VA and another for late-century institutional VA.
Explanation III: Institutions

Decomposition of University Regressivity

Institution Composition (solid color): \[ \sum_u \nu^U_0(u) \Delta_q \left[ P_t(u|q) \right]. \]
Explanation III: Institutions

Decomposition of University Regressivity

Inst. Value-Added (stripe): \( \sum_u \Delta_q \left[ P_t(u|q) \right] (v_t^U(u) - v_0^U(u)) \)
Explanation III: Institutions

Decomposition of Rising Institutional Stratification

Log Dollars

1940 1960 1980 2000 2020
Year Turned 18

4-Year Composition 2-Year Comp. + Returns
4-Year Returns 2/4-Year Transitions

CC Enrollment
Zachary Bleemer and Sarah Quincy
Changes in the College Mobility Pipeline Since 1900
July 2024 24 / 26
Discussion

- US higher education has become regressive since 1960 for men and women.

- Three key factors explain 70% of the trend:
  1. Less-selective and public institutions’ decreased value-added since 1960.
  2. Lower-income students’ diversion to community college since the 1980s.
  3. Declining access to high-value majors within universities since the 2000s.

- The current rank-rank correlation is 0.29. We conduct two simulations in NLSY97:
  1. Equalize return to college, assuming 15% of collegiate regressivity is selection. Resulting rank-rank correlation: 0.23
  2. Equalize college attainment. Resulting rank-rank correlation: 0.18

- About 20 percent of intergenerational income persistence is mediated by differential value to college-going. In 1960, that share was 0.
Similar pattern when looking at four-year college attainment.
College Enrollment in the United States by Income Tercile

Top-Tercile Students’ Percent Higher College-Going Relative to Bottom-Tercile Students

![Graph showing changes in college enrollment by income tercile from 1900 to 2000.](image-url)
College Enrollment in the United States by Income Tercile

College Enrollment Among Female Students

Percent

<table>
<thead>
<tr>
<th>Year turned 18</th>
</tr>
</thead>
<tbody>
<tr>
<td>1900</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>80</td>
</tr>
</tbody>
</table>
The Regressivity of U.S. Higher Education

Average Return to College **Attainment** ($\beta_t$)

![Graph showing average return to college attainment over time](image)

- **Wage Ranks**
- **Year turned 18**
- **Log Wages**
The Regressivity of U.S. Higher Education

Differential Return to College **Attainment** by SES ($\delta_t$)

![Graph showing the differential return to college attainment by SES over time. The graph plots wage ranks against years turned 18, with data points for different years (1920, 1940, 1960, 1980, 2000, 2020). The trend line indicates an overall increase in wage ranks, with a slope of $\delta=0.214$ (0.088).](graph.png)
The Regressivity of U.S. Higher Education

Average Return to College Enrollment in Log $ (\beta_t)$
The Regressivity of U.S. Higher Education

Differential Return to College Enrollment by SES in Log $ (\delta_t)$
The Regressivity of U.S. Higher Education

Average Return to College **Attainment in Log $ (β_t)**

![Graph showing changes in college mobility pipeline since 1900](image-url)
The Regressivity of U.S. Higher Education

Differential Return to College \textit{Attainment} by SES in Log $\delta_t$

![Graph showing changes in college mobility pipeline since 1900](image-url)
The Regressivity of U.S. Higher Education

Root Mean Squared Error from Regressivity Models with Kink Points
High-Level Trends in US Higher Education

Sticker and Net Cost of College Over Time

- Private Sticker/Net Cost
- Public Sticker/Net Cost
- C.C. Sticker/Net Cost

Matriculation Year

1920 1940 1960 1980 2000 2020

$'000s (2022)

0 20 40 60
Explanation I: Selection

College Major Attainment Over Time

Percent

Graduation Year

1920 1940 1960 1980 2000 2020

Ivy and Ivy+ Enrollment Over Time

- **Blue Book**
- **Other Admin.**
- **HEGIS**
- **IPEDS**


Percent: 0, 2, 4, 6

Zachary Bleemer and Sarah Quincy
Changes in the College Mobility Pipeline Since 1900
July 2024
Pre-College Human Capital-Based Selection into College Attainment

Test Ranks vs. Year turned 18


Test Ranks: -20, 0, 10

\[ \delta = -0.220 \quad (0.073) \]
Explanation II: College Majors

Annual Share of Declared College Majors by Parental Income Tercile

Year Turned 18

Percent


Comp. S.
Human.
Econ.
Explanation III: Institutions

Geographic Distribution of Mid-Century Institutional Value-Added

-0.10 -0.05 0.00 0.05 0.10
Institutional Value-Added by testing tercile in the 1960s and 1990s

Distribution of 1963 Value-Added

Distribution of 1996 Value-Added
**Explanation III: Institutions**

Average Parental Income of Pell and Non-Pell Students (NPSAS)

![Graph showing average parental income of Pell and Non-Pell students from 1990 to 2020.](image)

- **Fam. Inc. Percentile**: 0, 40, 80
- **Enrollment Year**: 1990, 2000, 2010, 2020

Zachary Bleemer and Sarah Quincy

Changes in the College Mobility Pipeline Since 1900

July 2024
Explanation III: Institutions

Institutional Enrollment by Pell Eligibility Indexed by Average Wage or Value-Added

![Graph showing institutional enrollment by Pell eligibility indexed by average wage or value-added over time from 1980 to 2020.](image)

- All Colleges, Avg. Wage
- VA Colleges, Avg. Wage
- VA Colleges, Avg. VA
Explanation III: Institutions

Enrollment Share in Two-Year Institutions by Income Tercile or Pell

Year Turned 18

Percent

0 10 20 30 40

1940 1960 1980 2000 2020

Pell Students  Non-Pell Students