Changes in the College Mobility Pipeline Since 1900

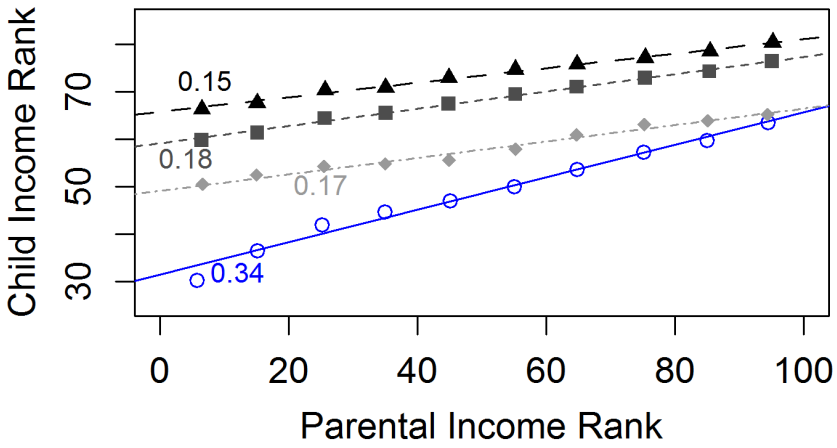
Zachary Bleemer
Princeton University and NBER

Sarah Quincy
Vanderbilt University and NBER

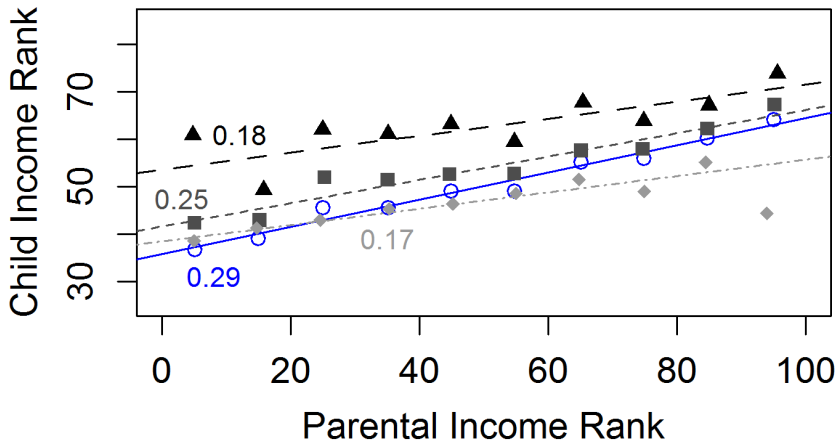
July 2024

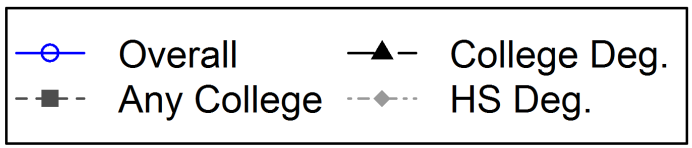
College-going became regressive over the 20th century

1905-1910 (Linked Censuses)



1980-1984 (NLSY97)





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**.

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.

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.

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.

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.

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.

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.

Contributions to literature

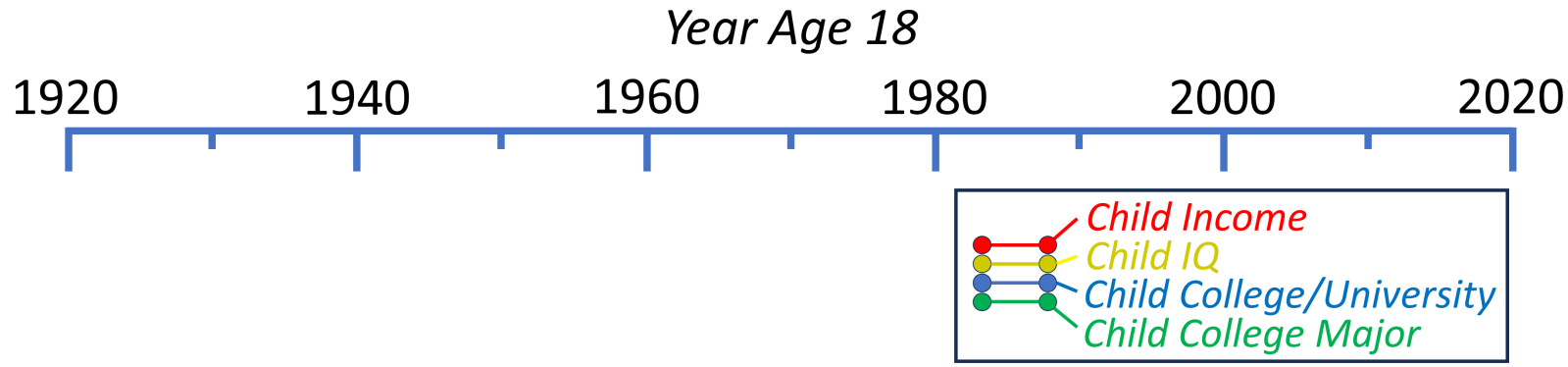
- Add university heterogeneity to the long literature on the relationship between education, inequality, and economic mobility.
  - **Rising Inequality among the College-Educated**: Goldin and Katz 1999, 2009; Lemieux 2006; Autor et al 2008, 2020; Acemoglu and Autor, 2011; Torche 2011.
  - **Changes in Education and Economic Mobility**: Aaronson and Mazumder 2008; Bailey and Dynarski 2011; Chetty *et al* 2014, 2017, 2020; Rothstein 2019; Jackson and Holzman 2020; Jácome, Kuziemko, and Naidu 2024.

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

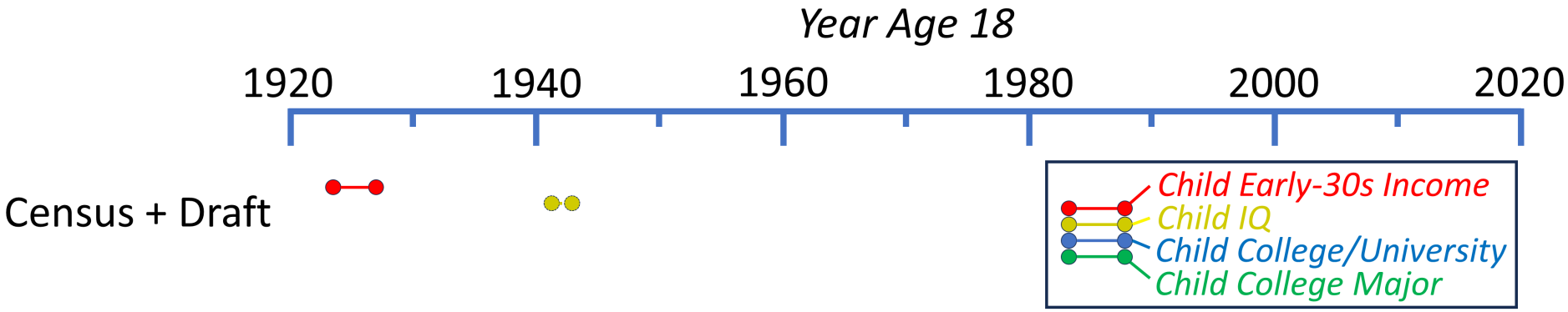
Contributions to literature

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

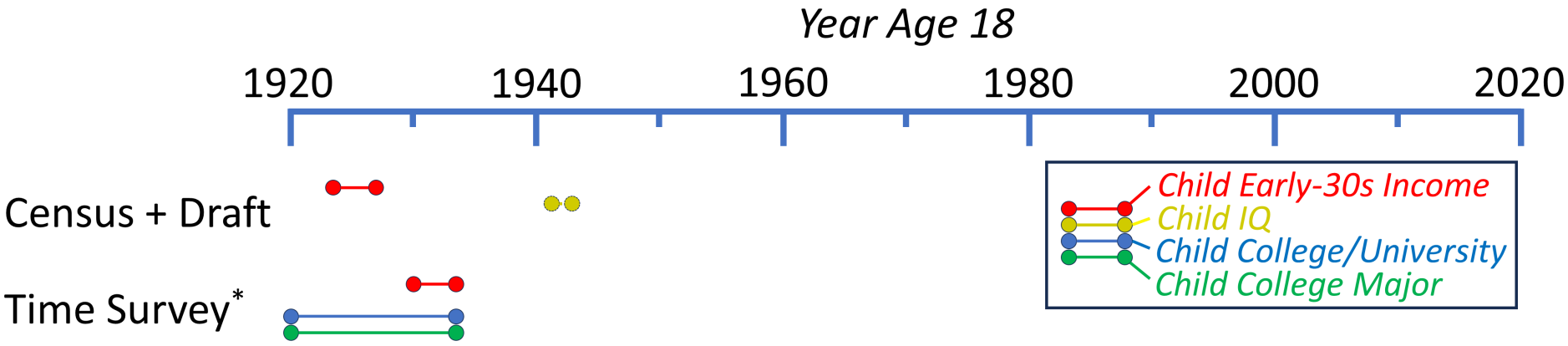
- Link the large microeconomic literature on heterogeneity in returns to higher education to long-run macro trends.
  - We extend estimates of **major** value-added (and composition by parental income) by 50 (35) years: Arcidiacono 2004; Fairlie, Hoffman, and Oreopoulos 2014; Altonji *et al* 2016; Kirkeboen, 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.



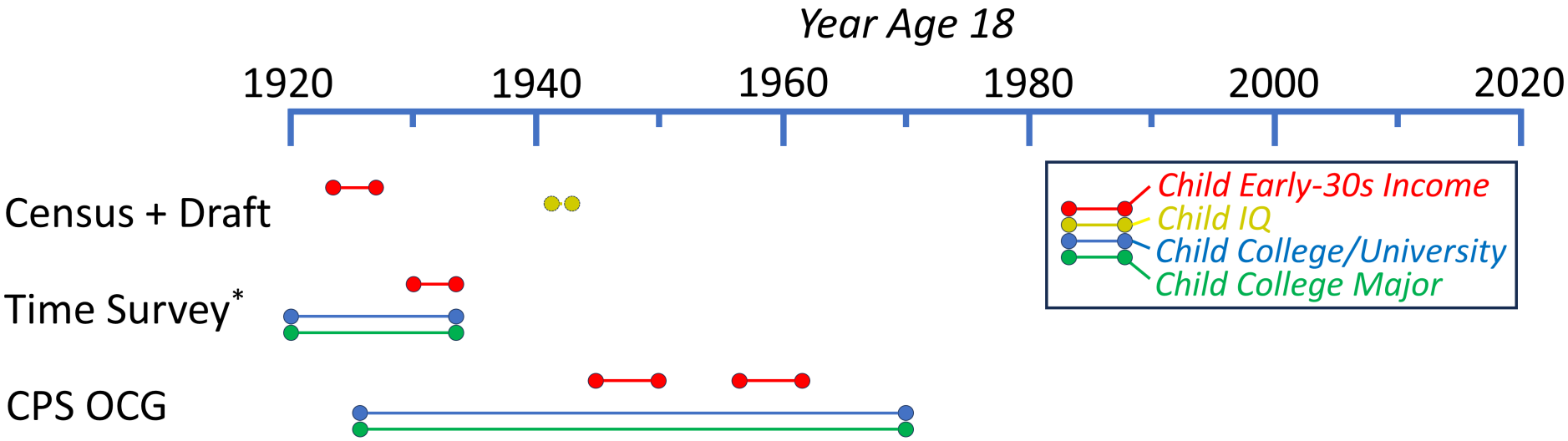
- **Data-Collection Goal**: Collect longitudinal surveys covering the past 100 years in the U.S., observing:
  1. Parental income while child is in high school
  2. Child test score in high school
  3. Child income in early 30s
  4. Child postsecondary institution (if attended college)
  5. Child college 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.



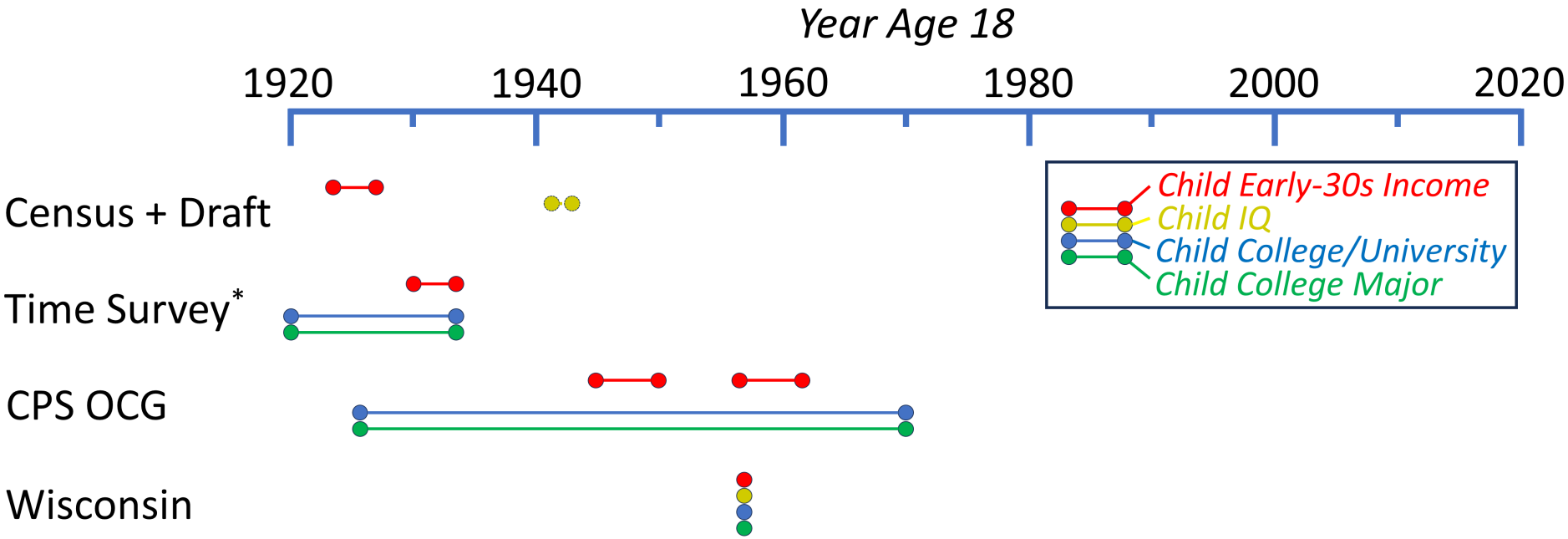
- **1920-1950 U.S. Census records** (Ruggles *et al* 2020).
  - 1920: Parental LIDO. 1940: Wage income, education.
  - Linked using the Census Linking Project (Abramitzky *et al* 2020).
- **WWII draft cards**
  - WWII enrollment cards (AGCT and education) from the National Archives.
  - Linked to 1950 education following ABE on name, birth year, and birth state.
- Sample Size: 329,000 overall; 2,804 with AGCT matches.



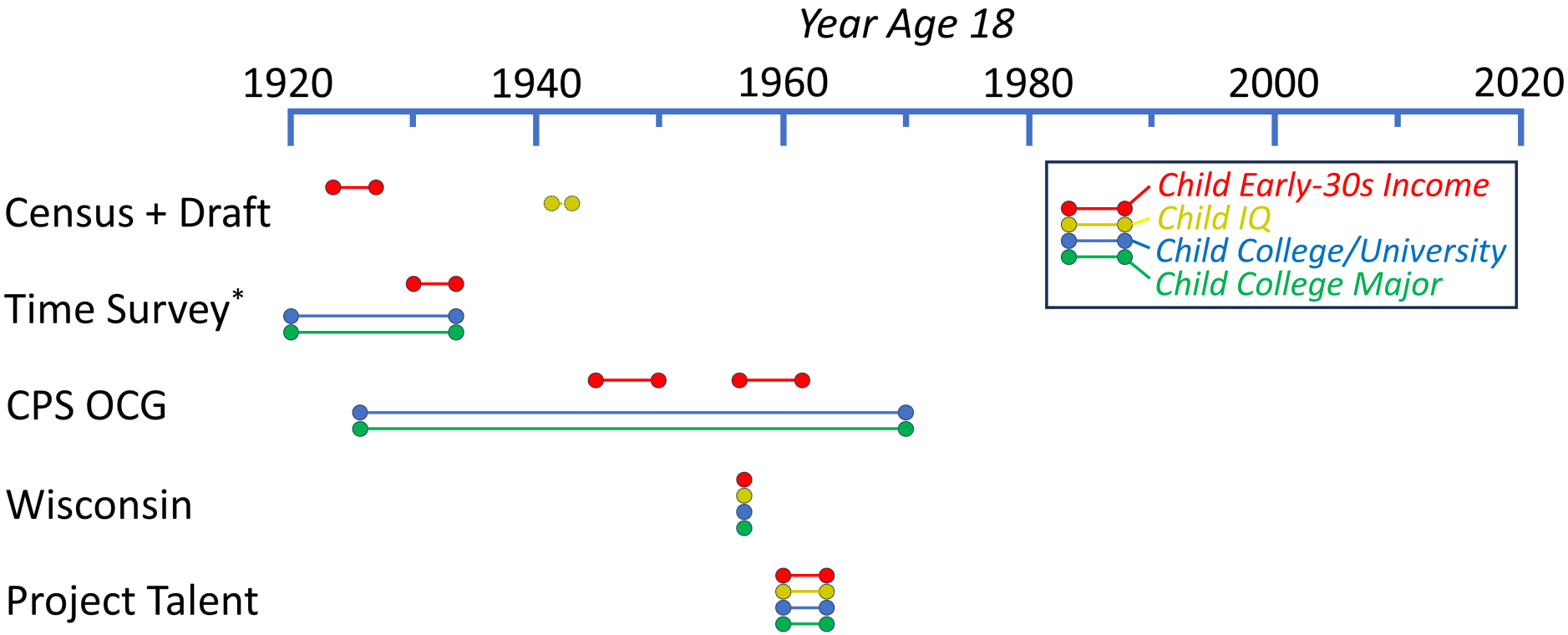
- 1947 survey of all US college graduates with last names beginning with "FA".
- Contains income, major, and institution, but not family income*.
- Sample Size: 1,818.



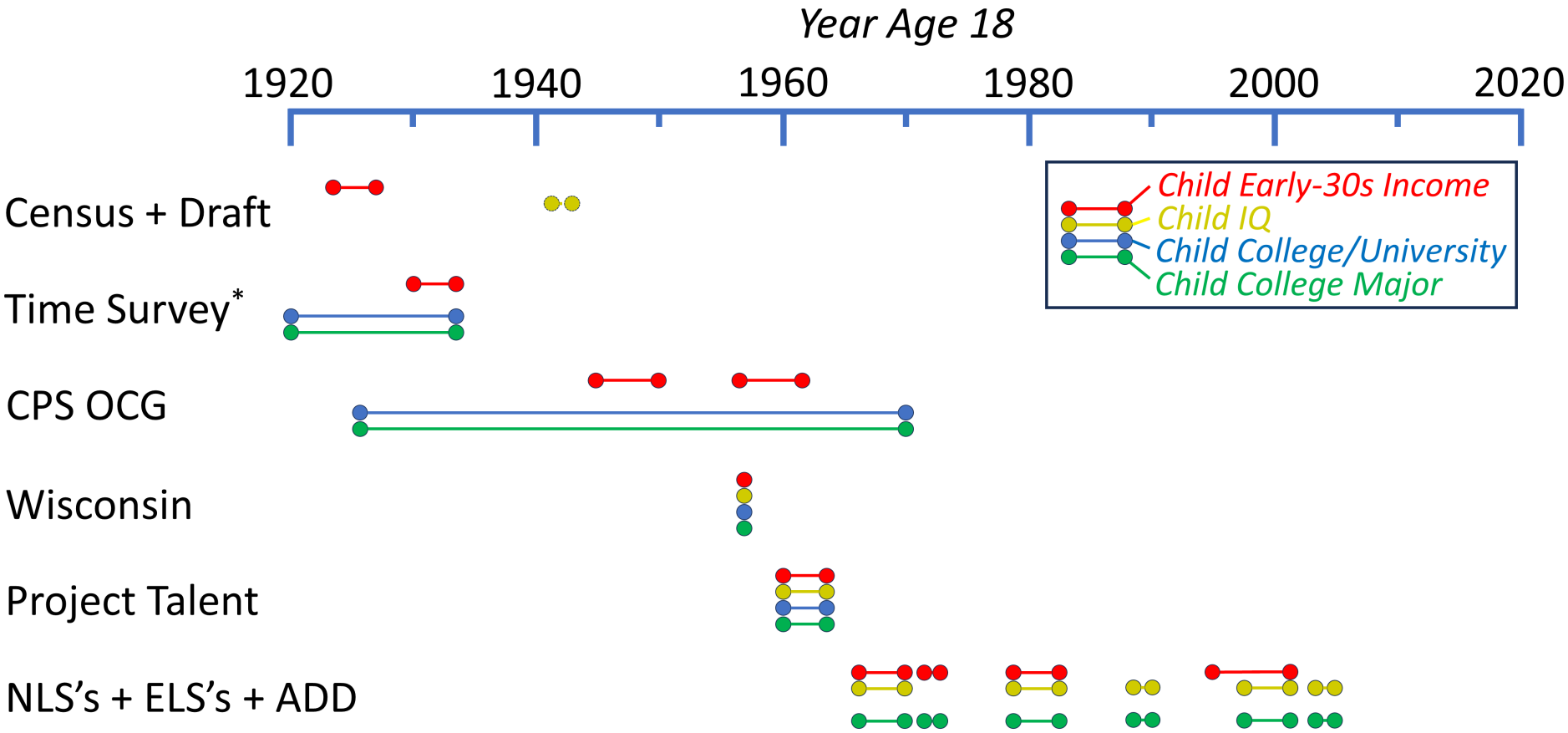
- CPS Occupational Change in a Generation study sampled from CPS participants.
- Contains parental income proxies (occupation, education, and geography), education, and income, along with institution and major in the latter survey.
- Sample Size: 1,711 ('62) and 2,778 ('73) for ages 30-35; 6,411 across cohorts with observable collegiate institutions and majors.



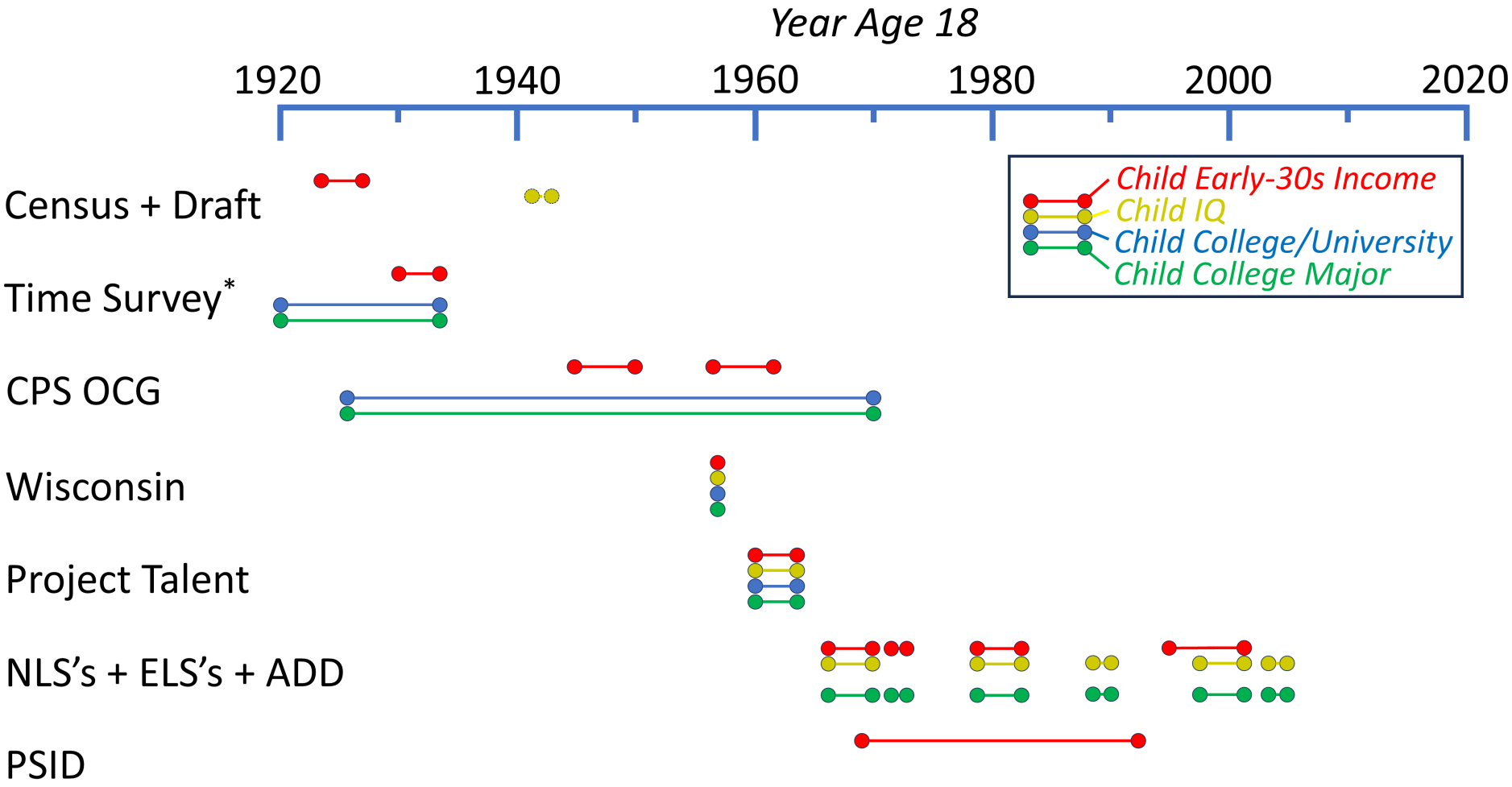
- Wisconsin Longitudinal Survey sampled from 1957 Wisconsin high school graduates.
- Contains parental income, IQ, income, major, and institution.
- Sample Size: 3,297.



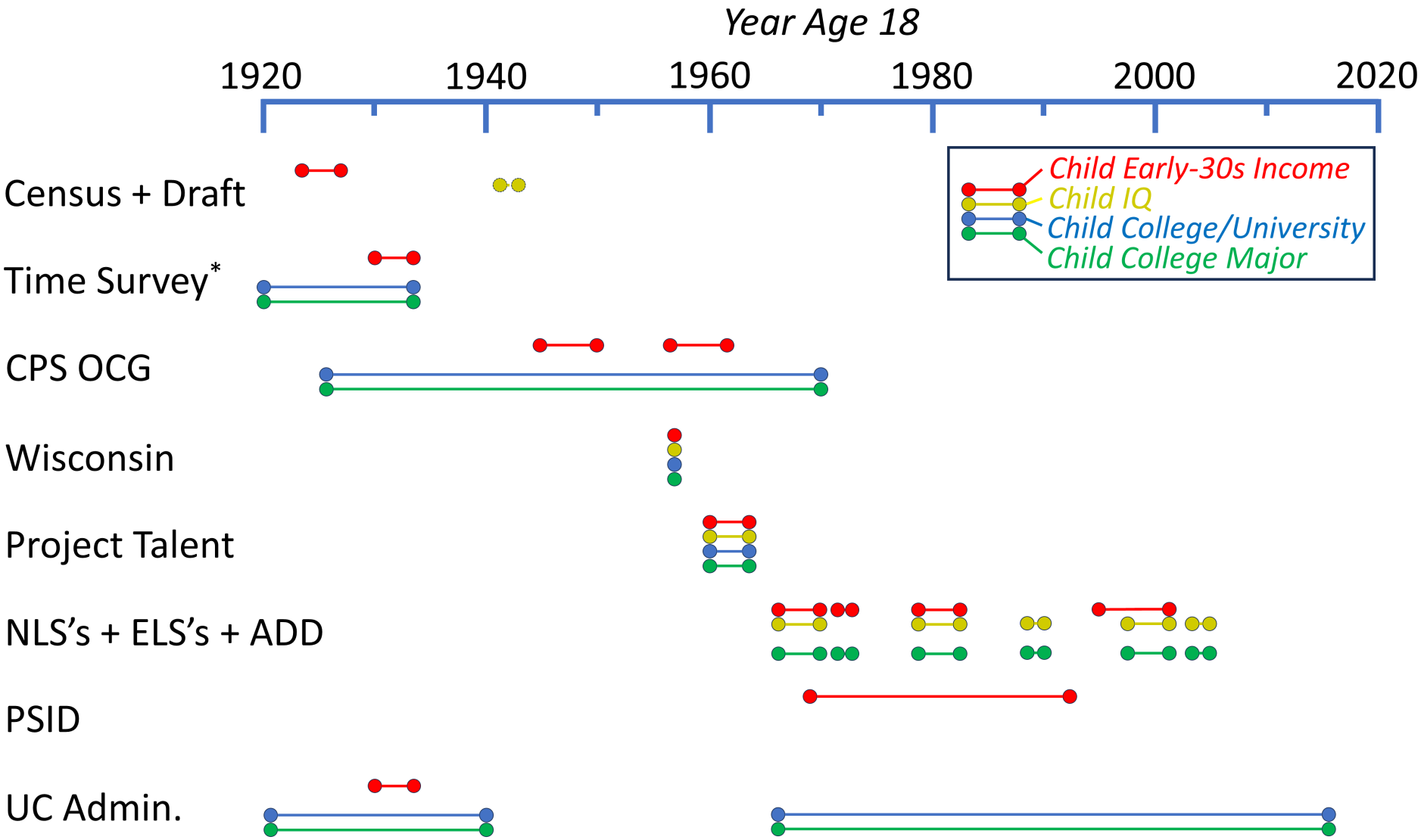
- Nationally-representative longitudinal survey of American freshmen and sophomores enrolled in 1960.
- Contains parental income, IQ, income, major, and institution.
- Sample size: 37,751, which is large enough to estimate institutional value-added for 403 institutions and 43 state-level pairs.



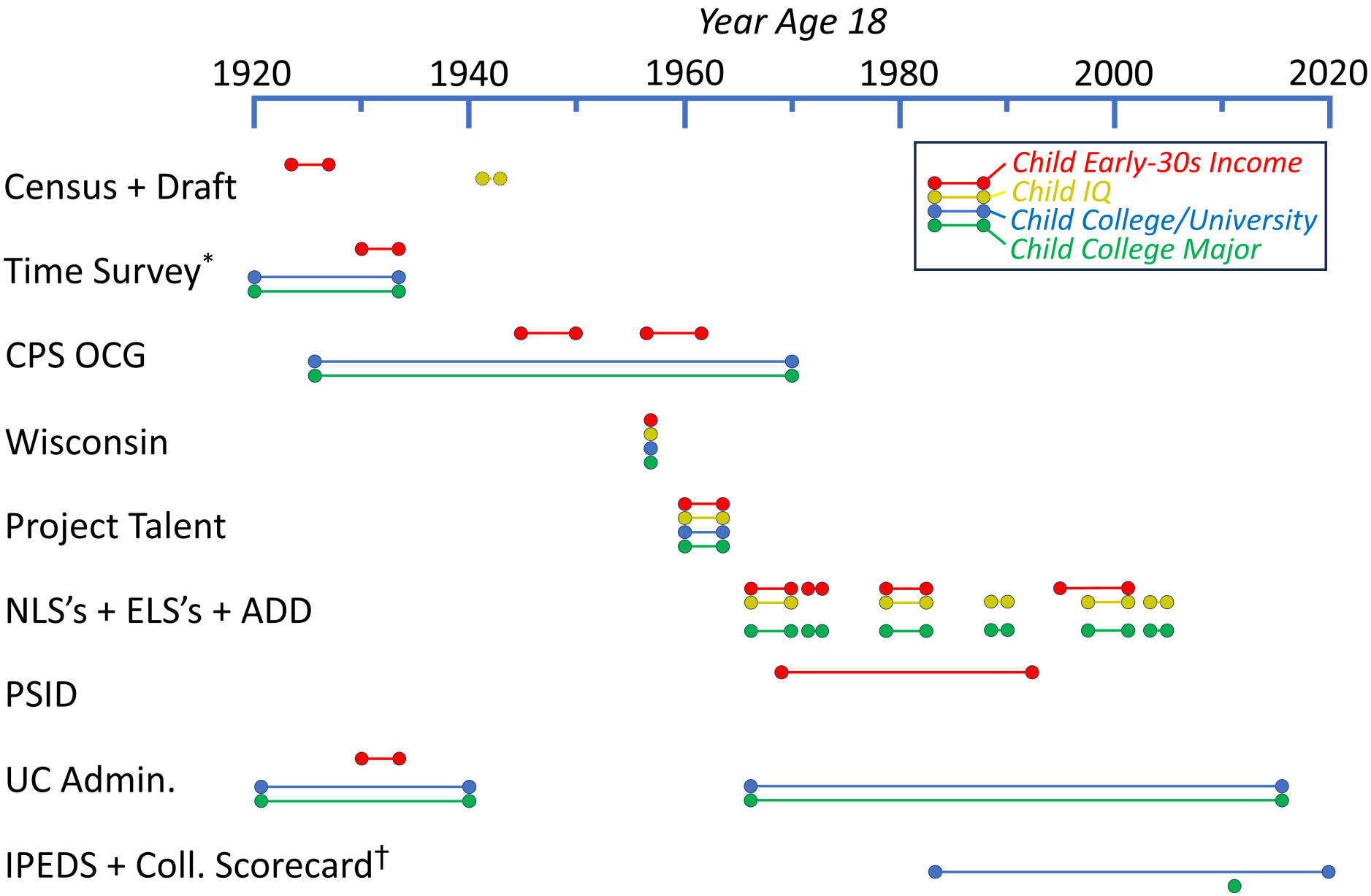
- Seven U.S. national longitudinal surveys have collected parental income and relevant fields:
  - NLSM (1,171), NLS72 (3,865), NLSY79 (1,938), and NLSY97 (2,690) have test scores, child income, and major.
  - NELS (4,570) and ELS (4,212) have test scores and major.
  - ADD Health (1,279) has child income.



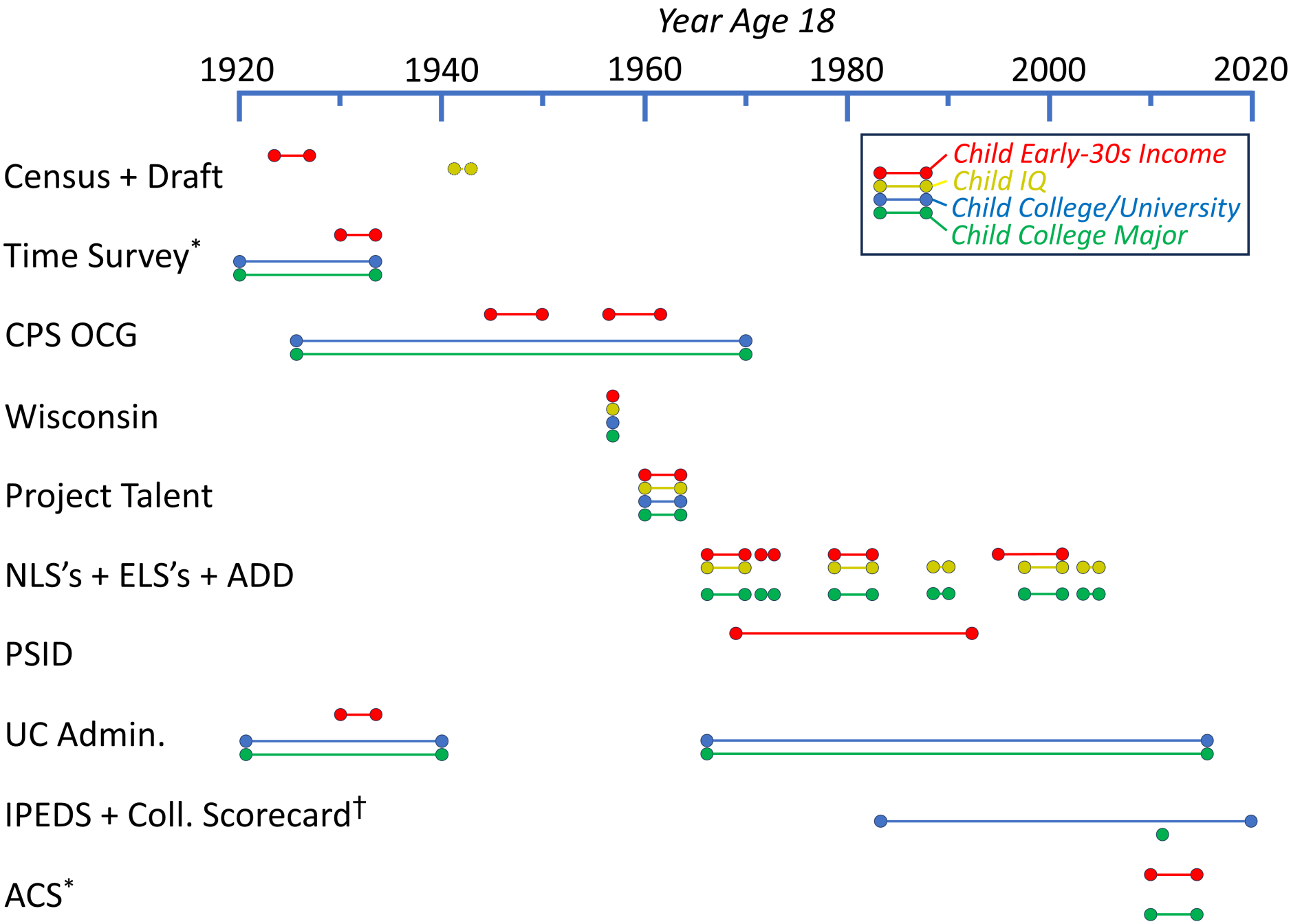
- Nationally-representative panel longitudinal survey of US households.
- Sample size: 1,991.



- 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.

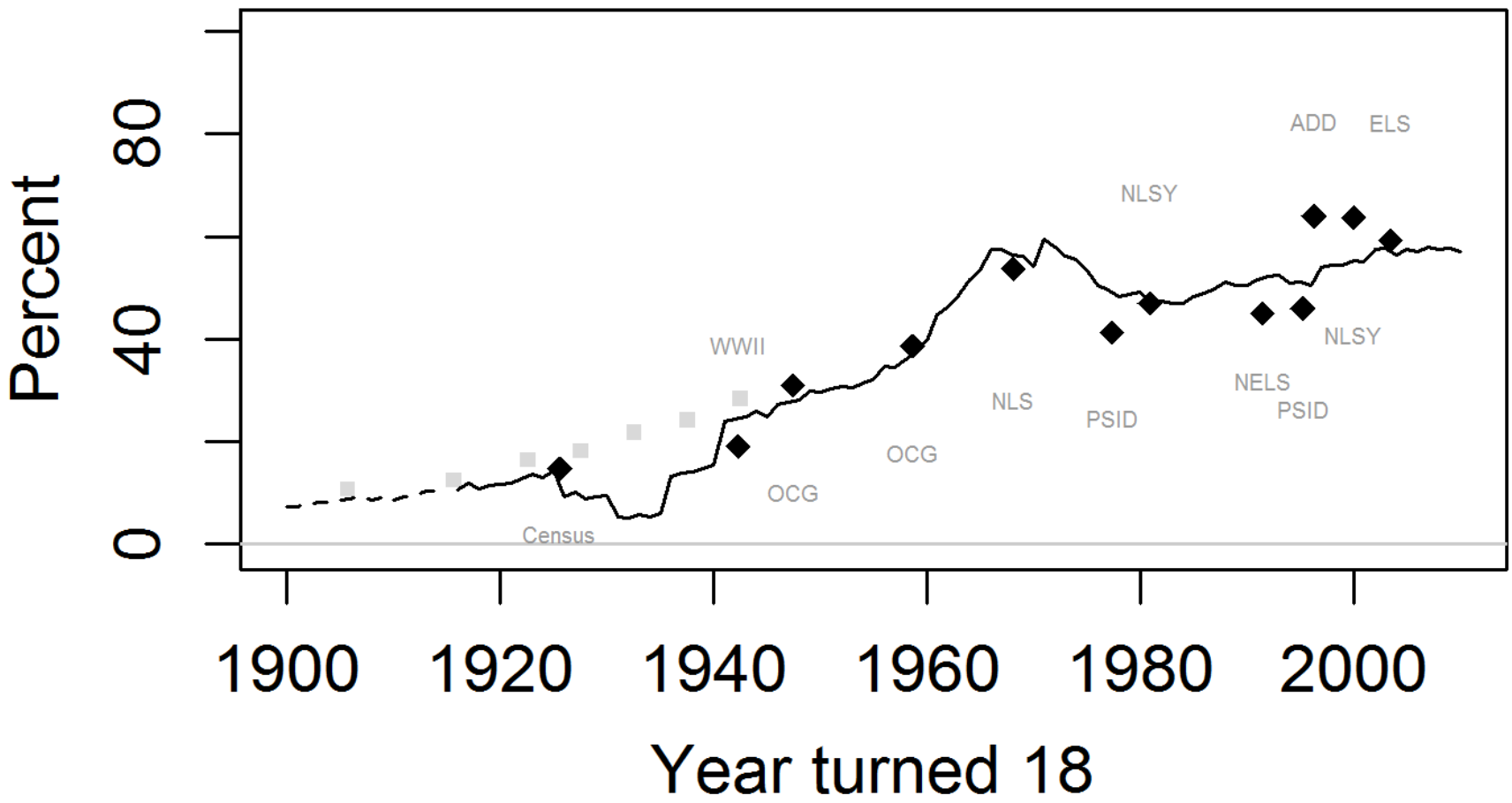


- *Institution*- and major-level Pell funding, which proxies low-income enrollment. Sample size: 159,741 + 18,135.



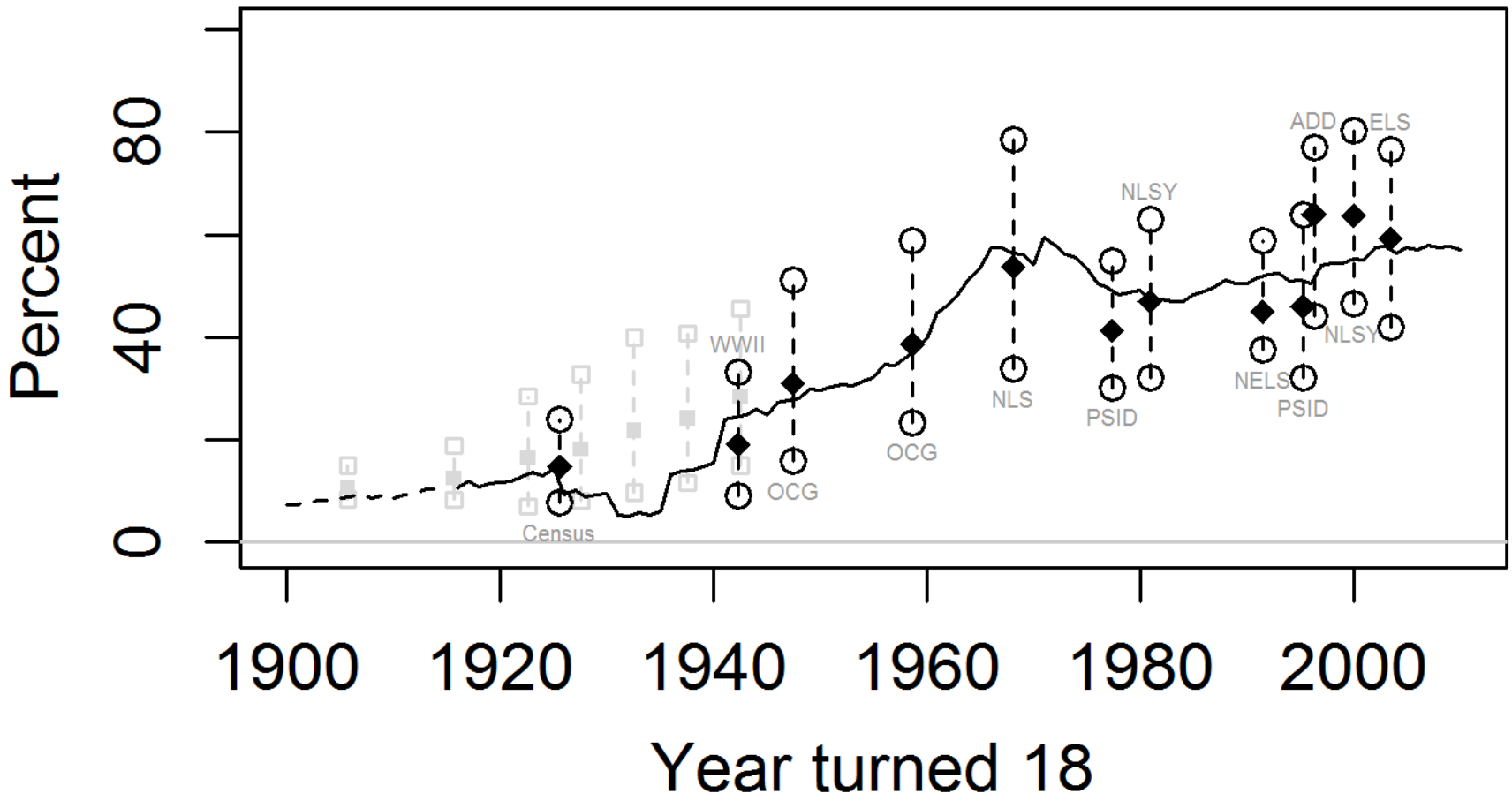
- Nationally-representative 2009-2021 survey. Sample size with major: 412,368.

College Enrollment in the United States



Male college enrollment rose following WWII but has been unchanged since the late 1960s...

College Enrollment in the United States by Income Tercile



...and higher-income children have always been more likely to enroll, especially following the GI bill (Stanley 2003).

The Regressivity of U.S. Higher Education

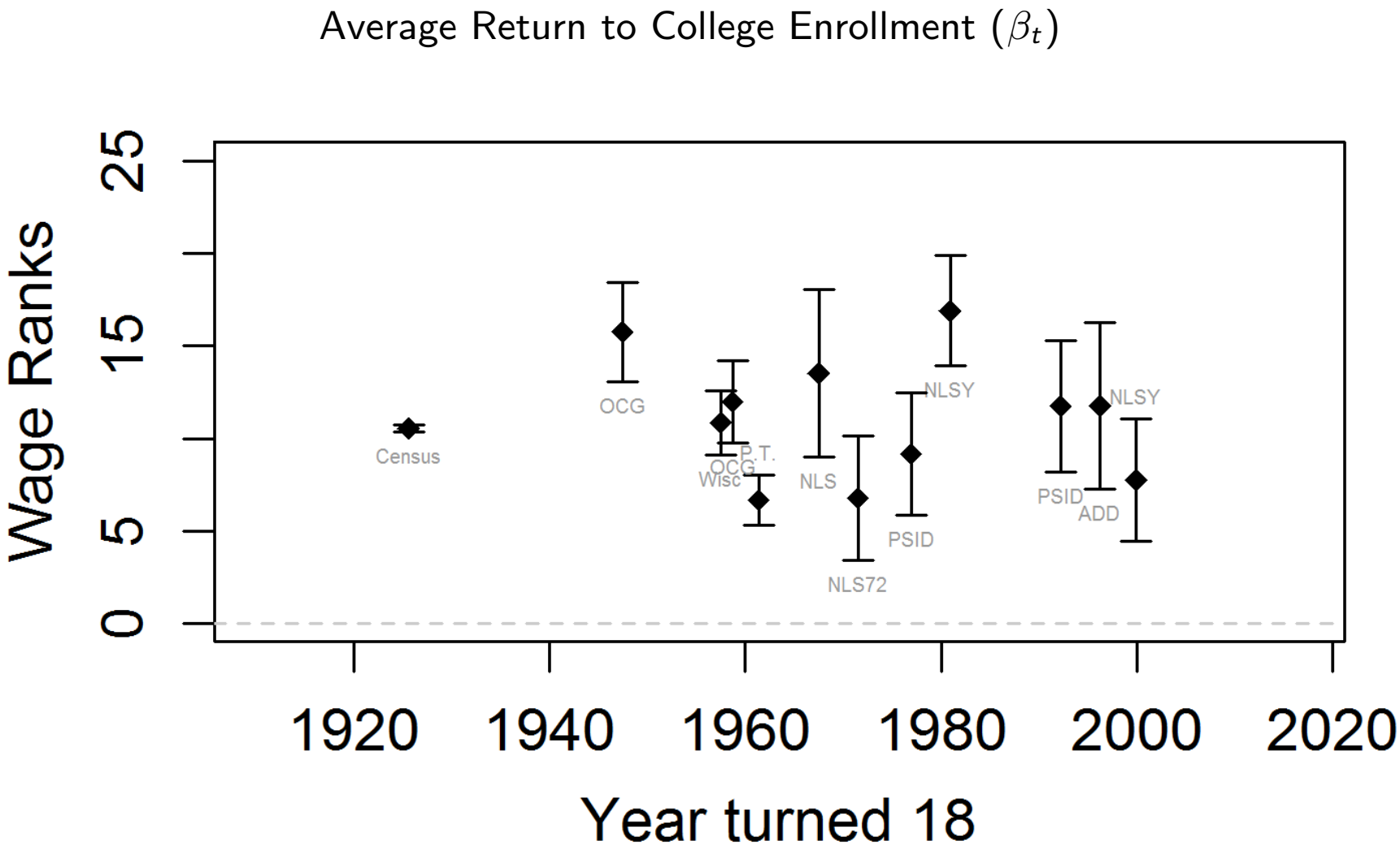
We estimate the observational wage return to higher education by family income:

$$Wage_{it} = \zeta_t + \alpha_t FamInc_{it} + \beta_t College_{it} + \boldsymbol{\delta_t}\left(FamInc_{it} \times College_{it}\right) + \epsilon_{it} \quad (1)$$

over high school graduates i in year t, where:

- $Wage_{it}$ is measured in rank or log \$ between ages 30 and 35;
- $FamInc_{it}$ is measured in rank between ages 14 and 17 and centered;
- $College_{it}$ indicates at least one year of college; and
- Estimation is weighted by sample weights and standard errors are robust.

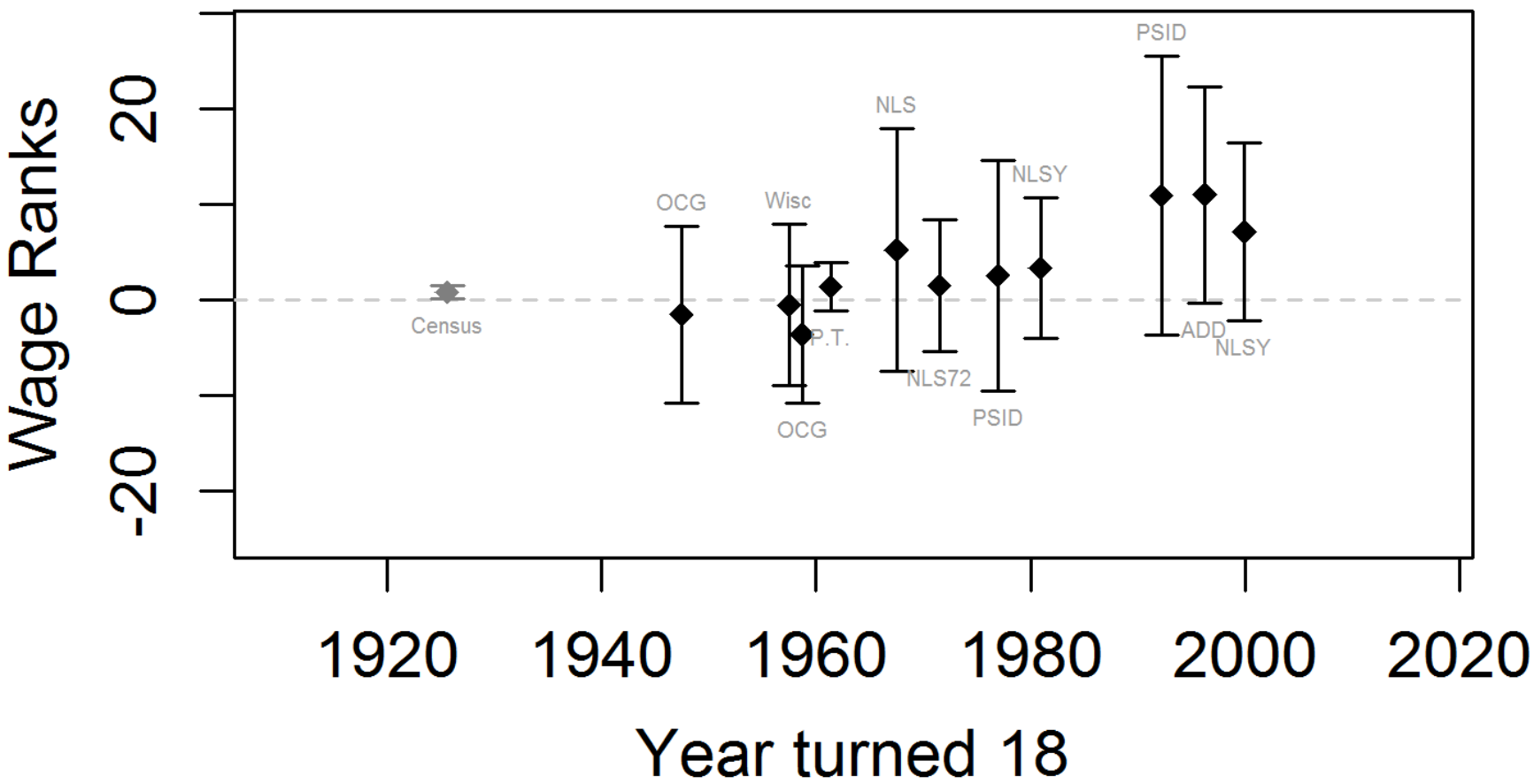
The Regressivity of U.S. Higher Education



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 (δ_t)

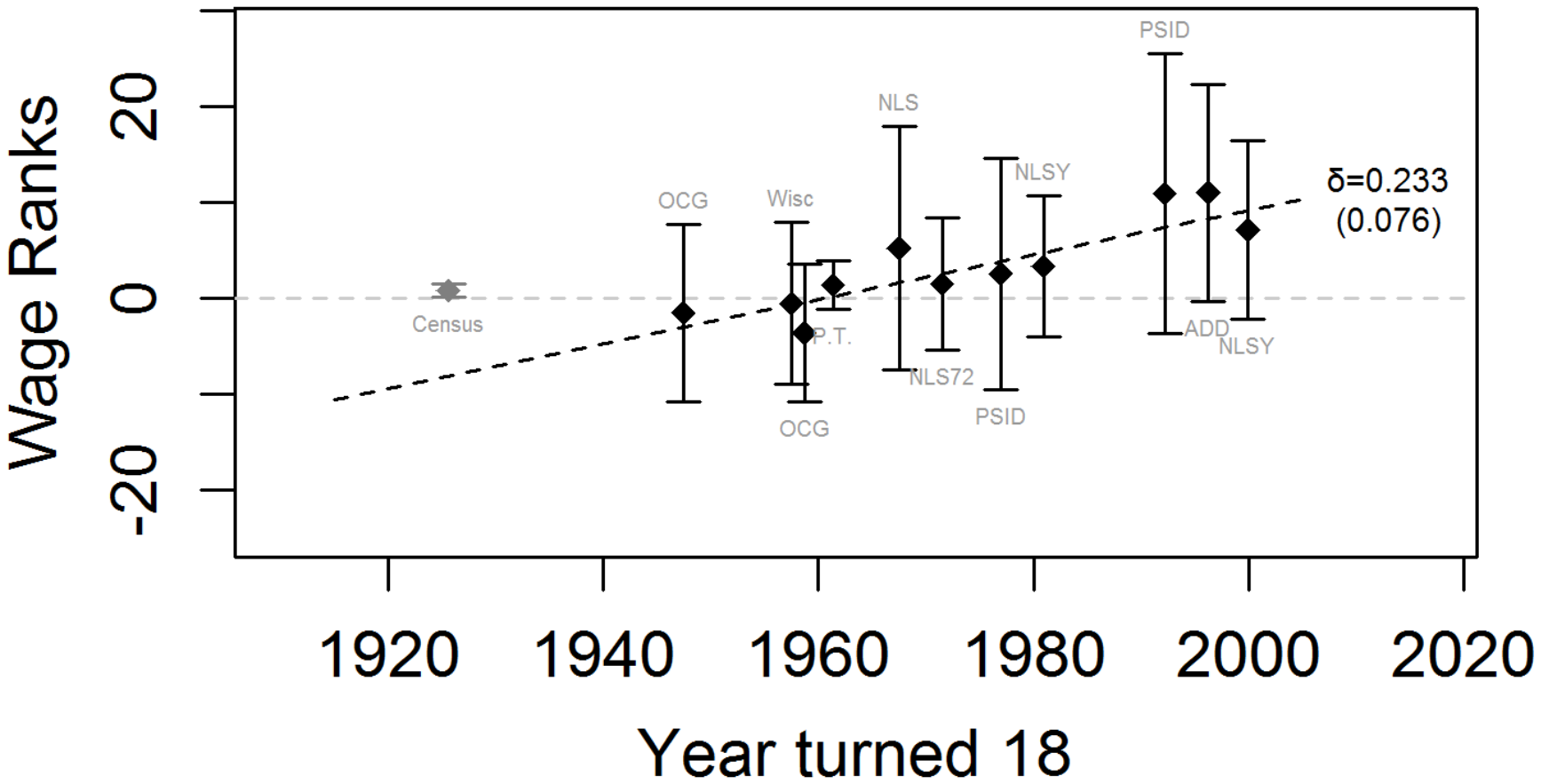


... but have become more regressive since the 1950s.

Log Wages | College Attainment | Kink RMSE

The Regressivity of U.S. Higher Education

Differential Return to College Enrollment by SES (δ_t)



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

Log Wages | College Attainment | Kink RMSE

What Explains the Rise in Observational Regressivity?

Let $p_t(i) = p_t\left(a_i, u_i, m_i, PI_i\right)$ 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:

$$D_t \equiv \Delta_q \Big[E[p_t|q] \Big] = E[p_t|T] - E[p_t|B]$$

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:

$$D_t = \sum_{x \in \{A,U,M\}} \left(\overbrace{\int_j v_0^x(j) \Delta_q \left[P_t(j|q) \right] dj}^{\text{Composition}} + \overbrace{\int_j \Delta_q \left[P_t(j|q) \right] \left(v_t^x(j) - v_0^x(j) \right) dj}^{\text{Wage Value}} \right) + \epsilon_t$$

Tuition Over Time | Majors Over Time | Ivy Enrollment

What Explains the Rise in Observational Regressivity?

Let $p_t(i) = p_t\left(a_i, u_i, m_i, PI_i\right)$ 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:

$$D_t \equiv \Delta_q \Big[E[p_t|q] \Big] = E[p_t|T] - E[p_t|B]$$

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:

$$D_t = \sum_{x \in \{A,U,M\}} \left(\overbrace{\int_j v_0^x(j) \Delta_q \left[P_t(j|q) \right] dj}^{\text{Composition}} + \overbrace{\int_j \Delta_q \left[P_t(j|q) \right] \left(v_t^x(j) - v_0^x(j) \right) dj}^{\text{Wage Value}} \right) + \epsilon_t$$

Tuition Over Time | Majors Over Time | Ivy Enrollment

What Explains the Rise in Observational Regressivity?

Let $p_t(i) = p_t\left(a_i, u_i, m_i, PI_i\right)$ 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:

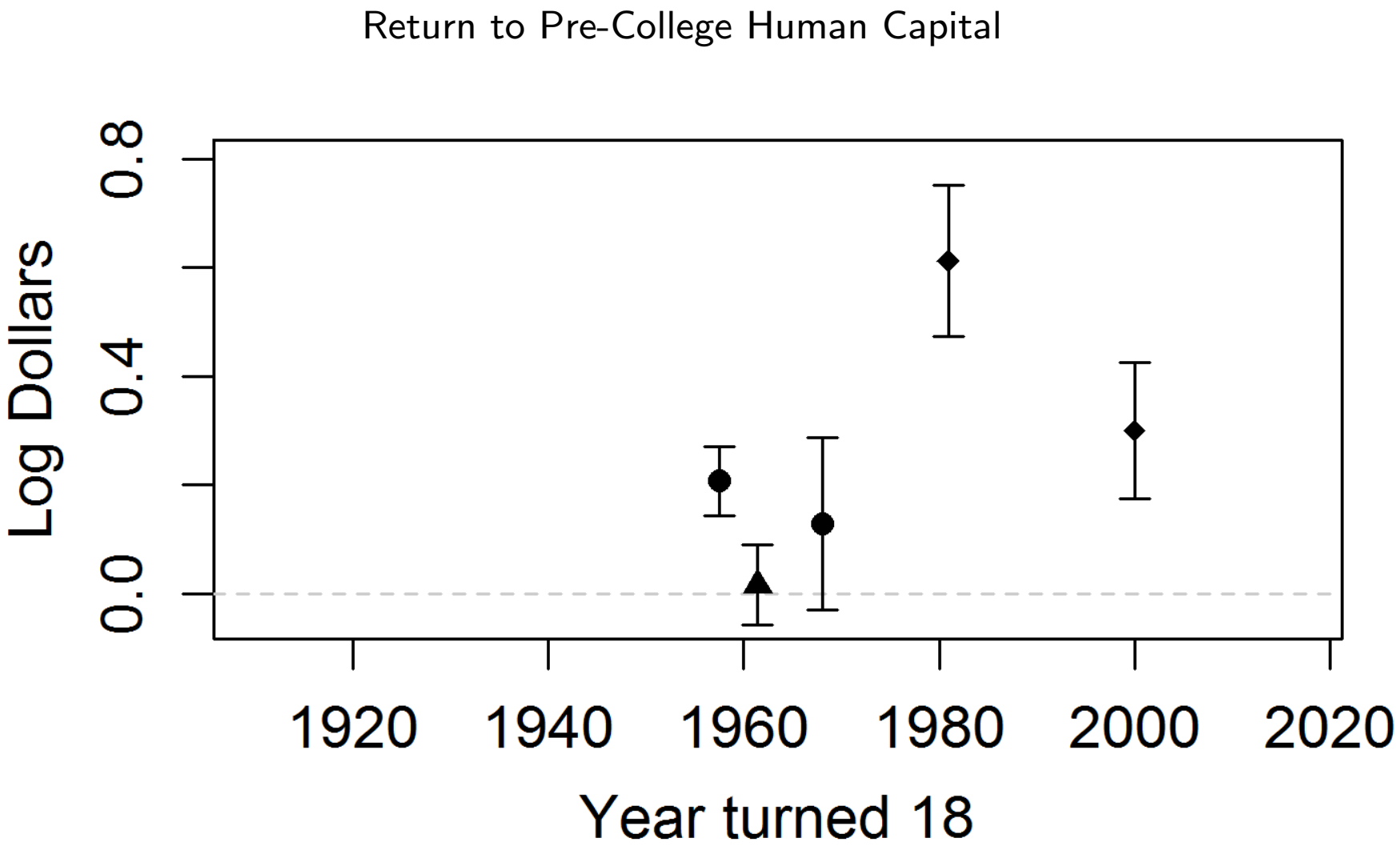
$$D_t \equiv \Delta_q \Big[E[p_t|q] \Big] = E[p_t|T] - E[p_t|B]$$

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:

$$D_t = \sum_{x \in \{A,U,M\}} \left(\overbrace{\int_j v_0^x(j) \Delta_q \left[P_t(j|q) \right] dj}^{\text{Composition}} + \overbrace{\int_j \Delta_q \left[P_t(j|q) \right] \left(v_t^x(j) - v_0^x(j) \right) dj}^{\text{Wage Value}} \right) + \epsilon_t$$

Tuition Over Time | Majors Over Time | Ivy Enrollment

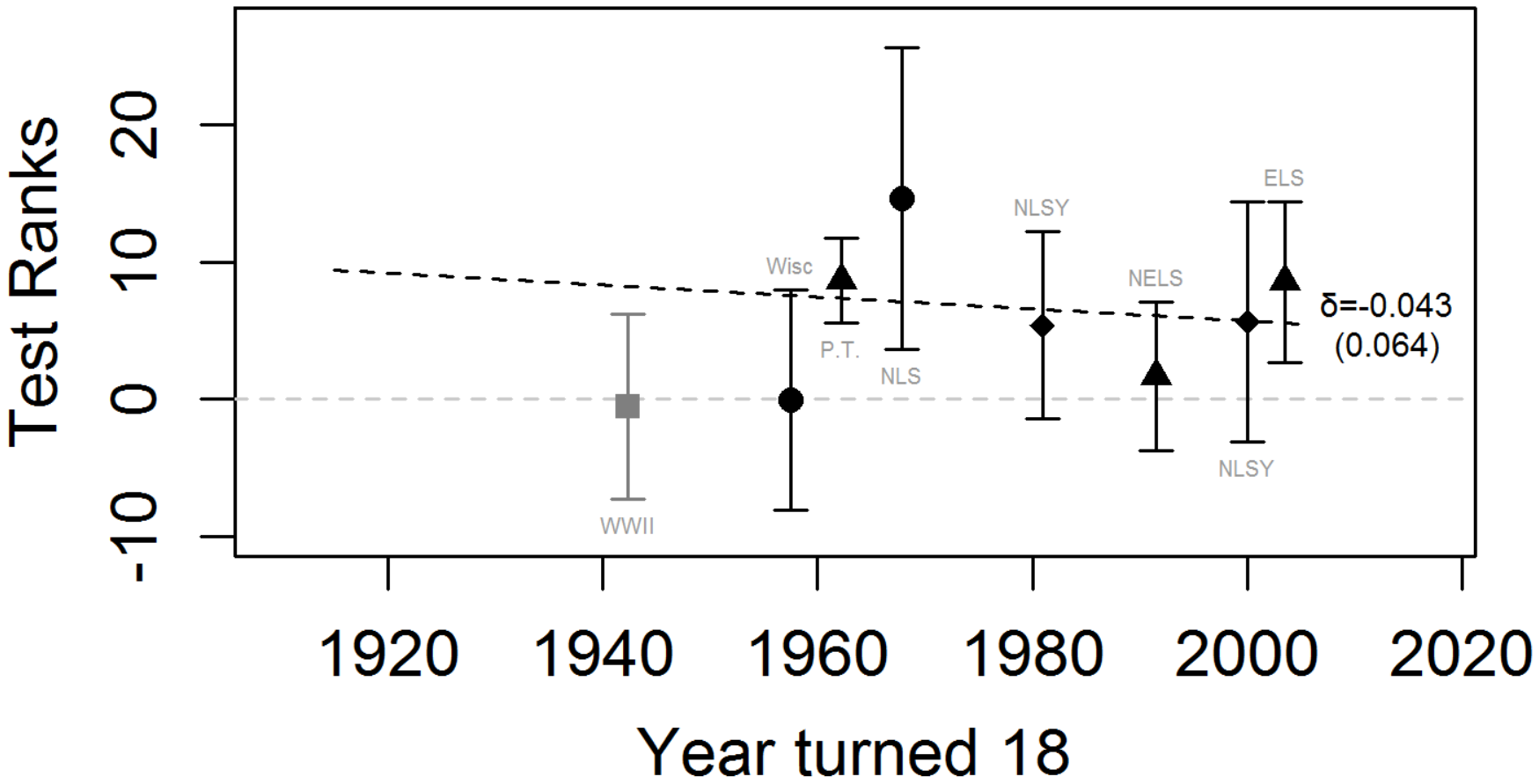
Explanation I: Selection



The labor market value of pre-college aptitude has risen (conditional on family income rank and child education)...

Explanation I: Selection

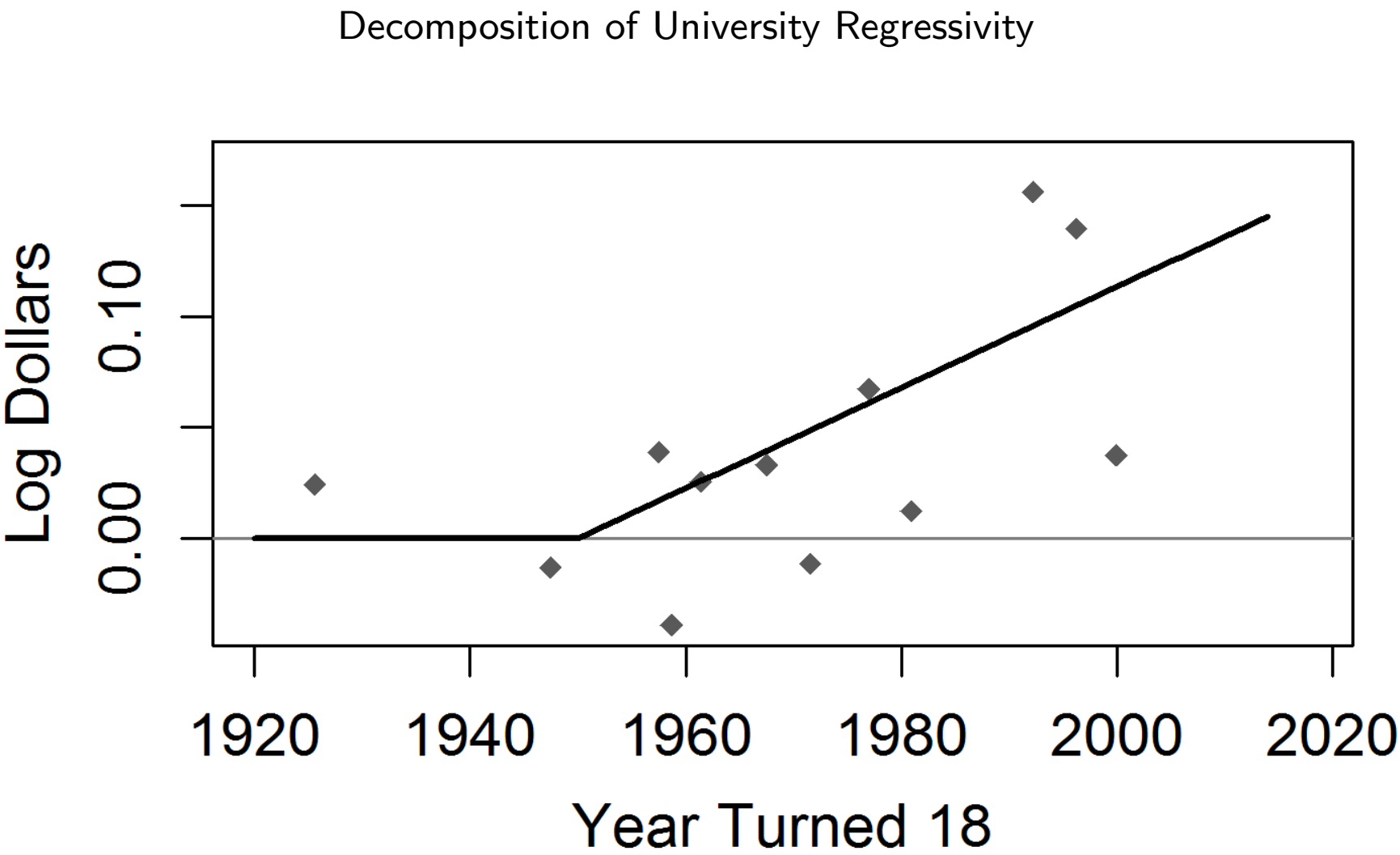
Differential Test Score Selection into College Enrollment by SES (δ_t)



...but selection into college-going has not changed since the 1960s.

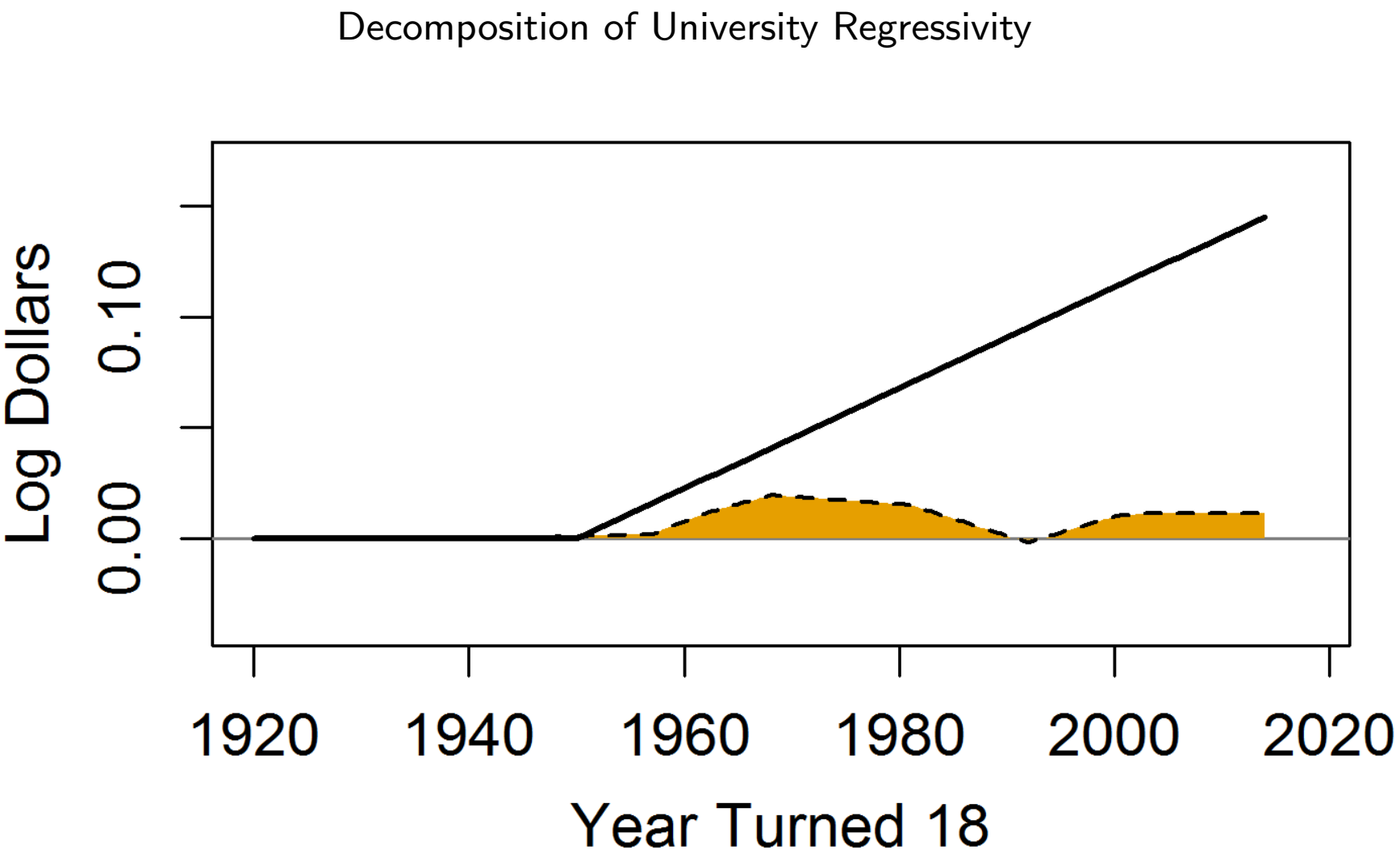
Selection into Attainment

Explanation I: Selection



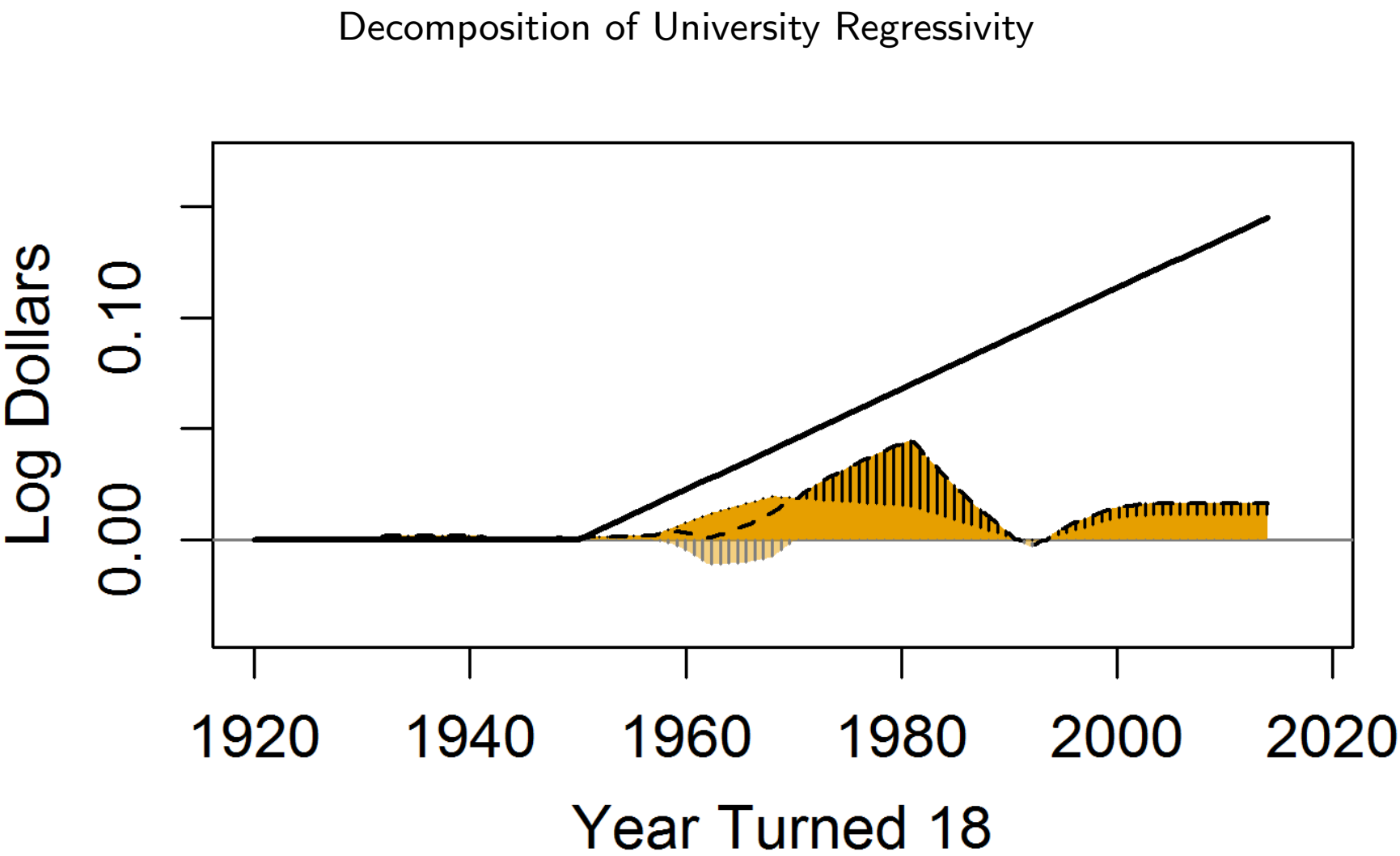
Slope comes from log-dollars version of the change in regressivity over time, converting to difference between top and bottom income tercile.

Explanation I: Selection



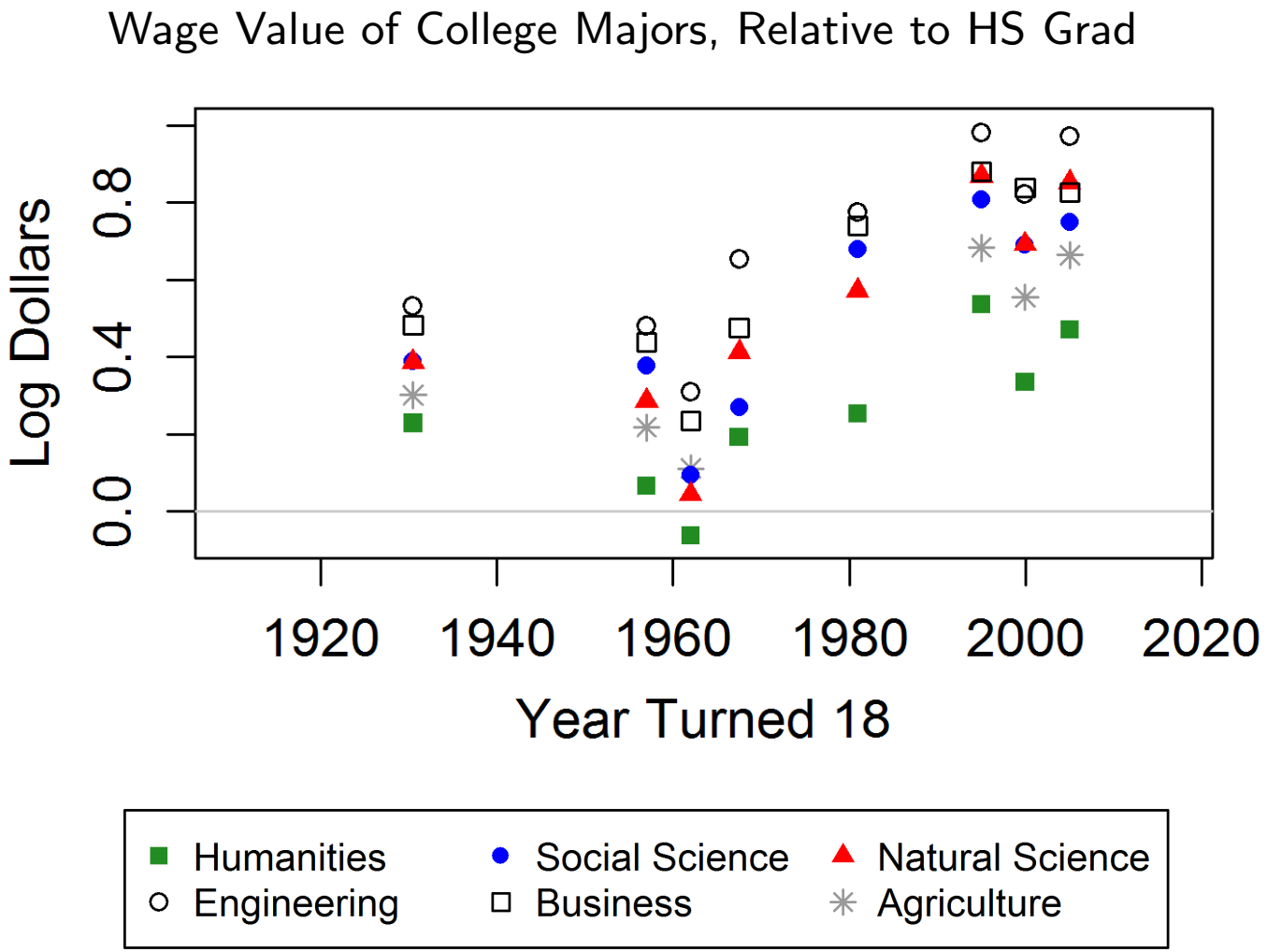
Test Composition (solid color): $\int_a v_0^A(a)\Delta_q\left[P_t(a|q)\right]\,da$.

Explanation I: Selection



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



Explanation II: College Majors

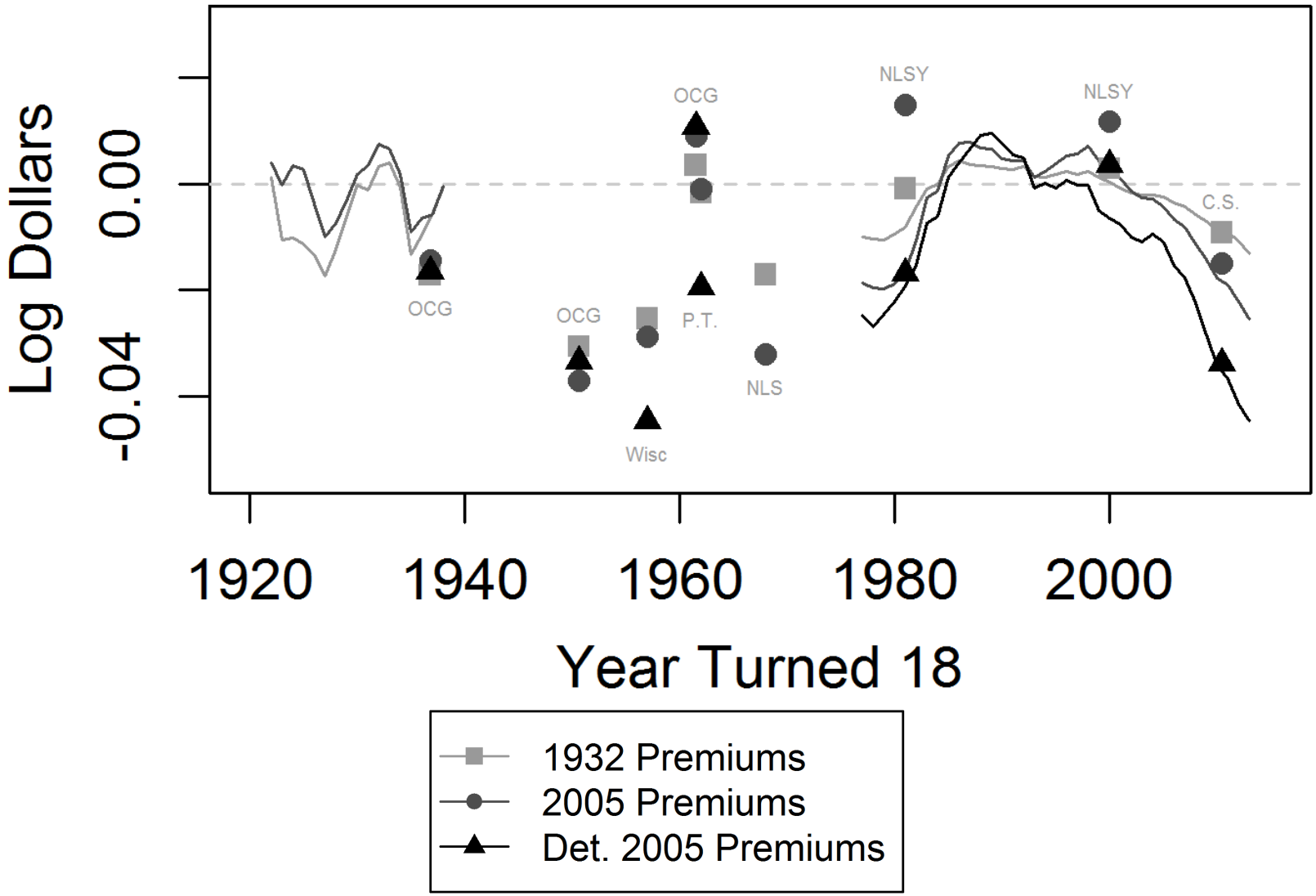
Selection-on-Observables Forecast Coefficients of Average Wages by Major

Sample:	Same Sample			Split Sample			
Add'l Cov.:	Fam. Inc.	+AFQT	+Race	None	Fam. Inc.	+AFQT	+Race
A. Discipline Premiums	1.02 (0.02)	1.03 (0.02)	1.03 (0.03)	0.83 (0.09)	0.88 (0.09)	0.92 (0.09)	0.91 (0.09)
Obs.		7			7		
1st Stg. Obs.		842			418		
B. Det. Major Premiums	1.00 (0.06)	1.01 (0.08)	1.00 (0.08)	0.66 (0.28)	0.68 (0.27)	0.68 (0.28)	0.67 (0.28)
Obs.		14			14		
1st Stg. Obs.		753			372		

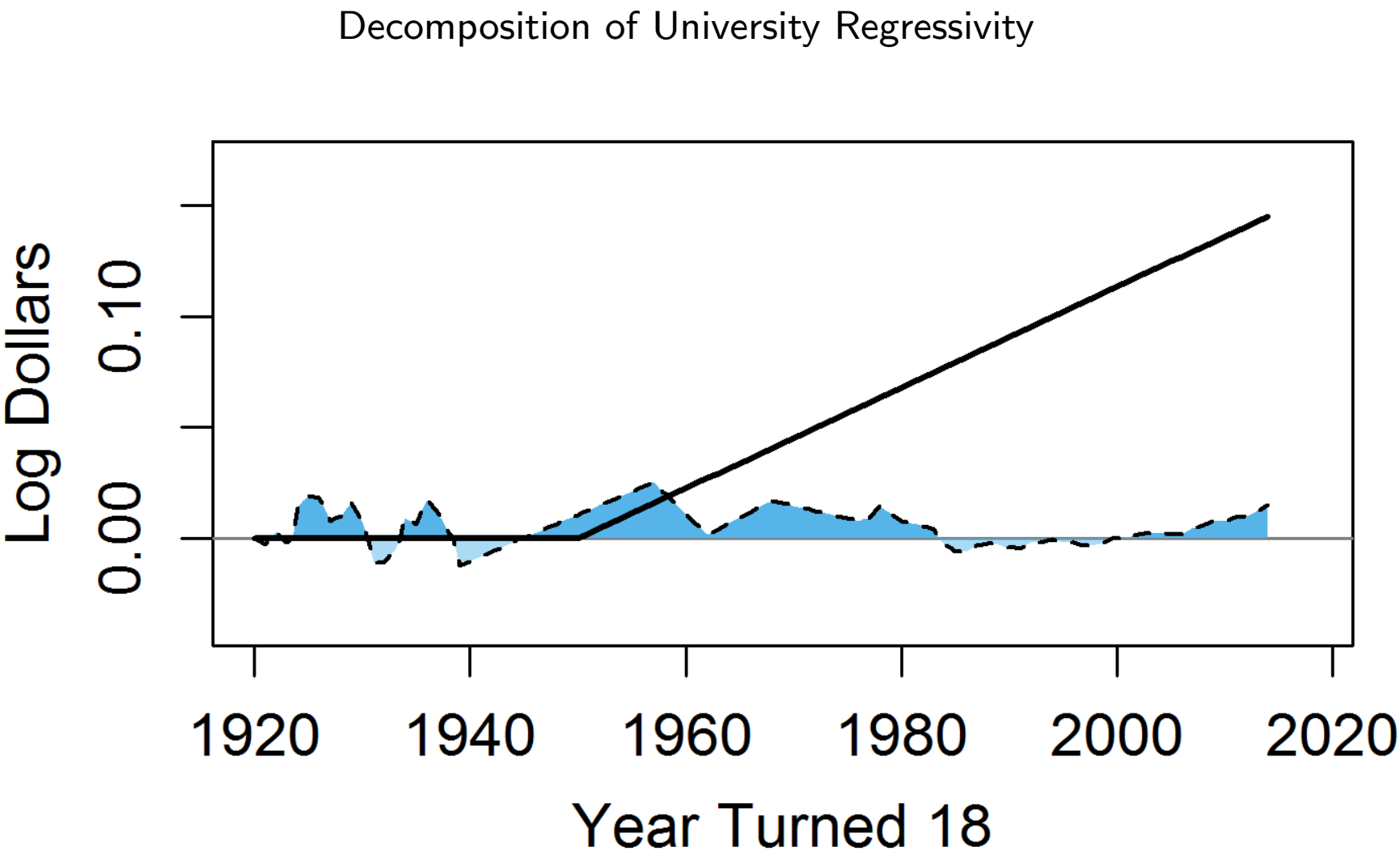
Takeaway: $\bar{w}_m$ seems to be a reasonable proxy for ATT_m.

Explanation II: College Majors

Difference in Major Premium Between Students from Bottom- and Top-Tercile Par. Inc.

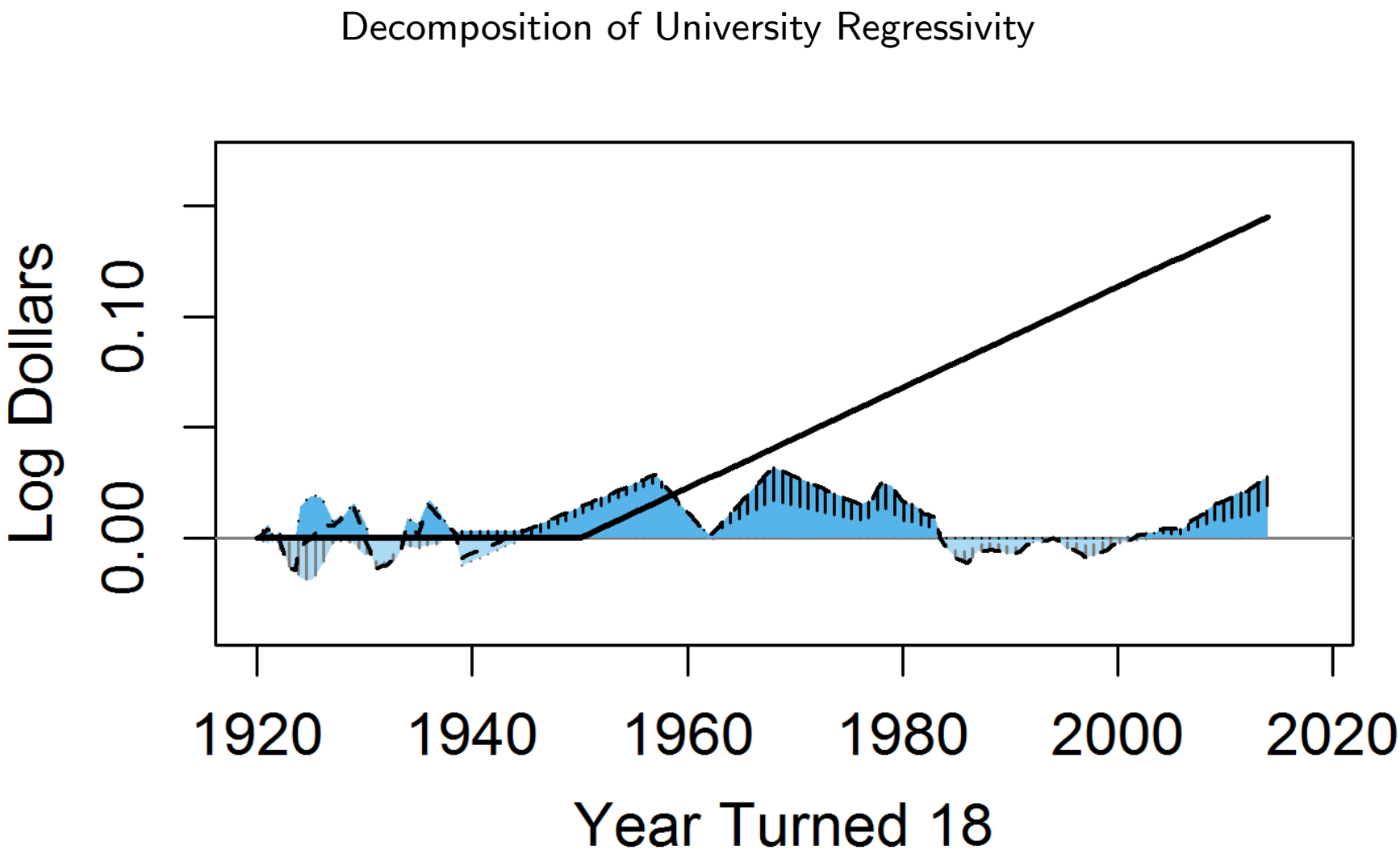


Explanation II: College Majors



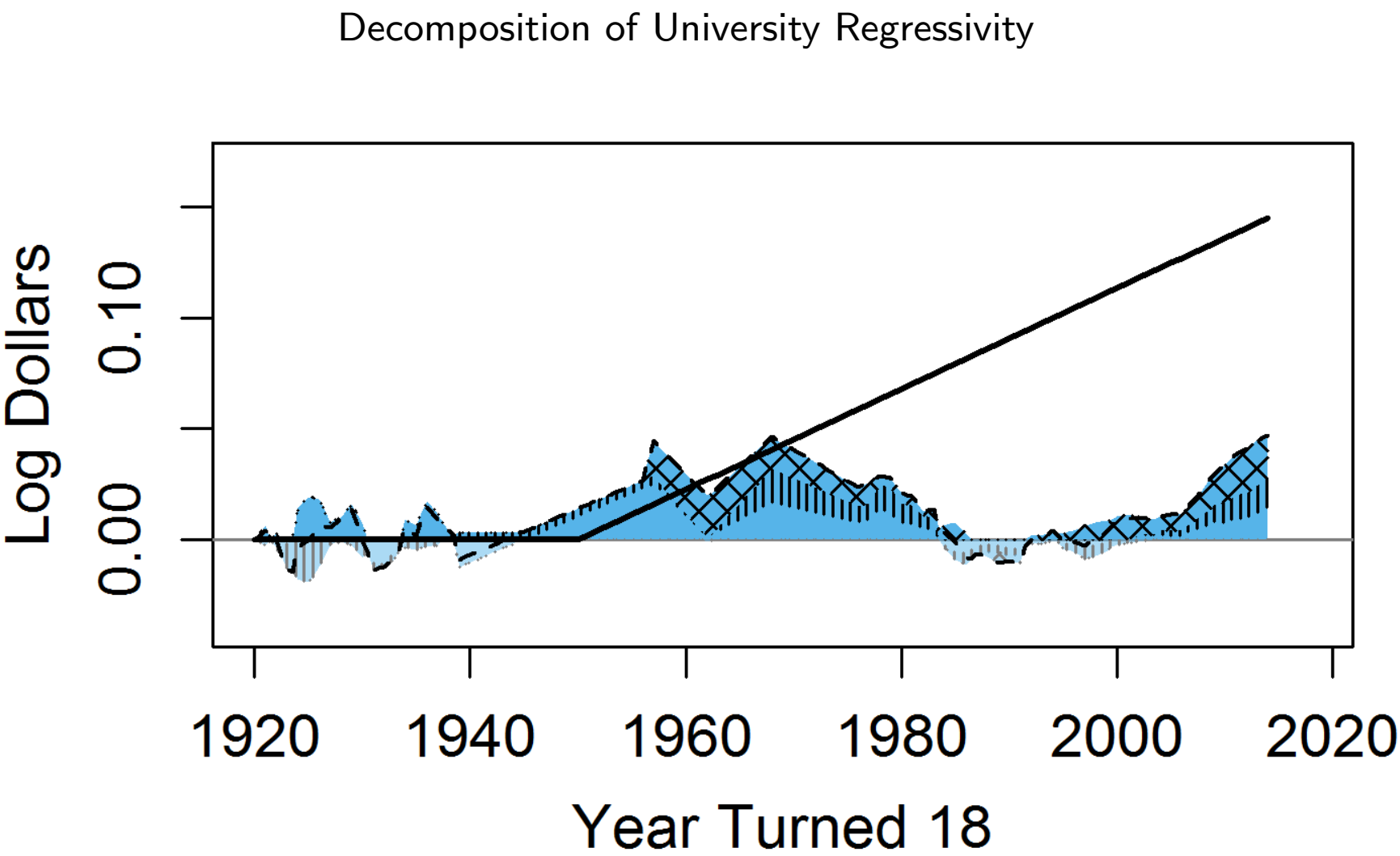
Discipline Composition (solid color): $\sum_m v_0^M(m)\Delta_q\left[P_t(m|q)\right]$.

Explanation II: College Majors



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

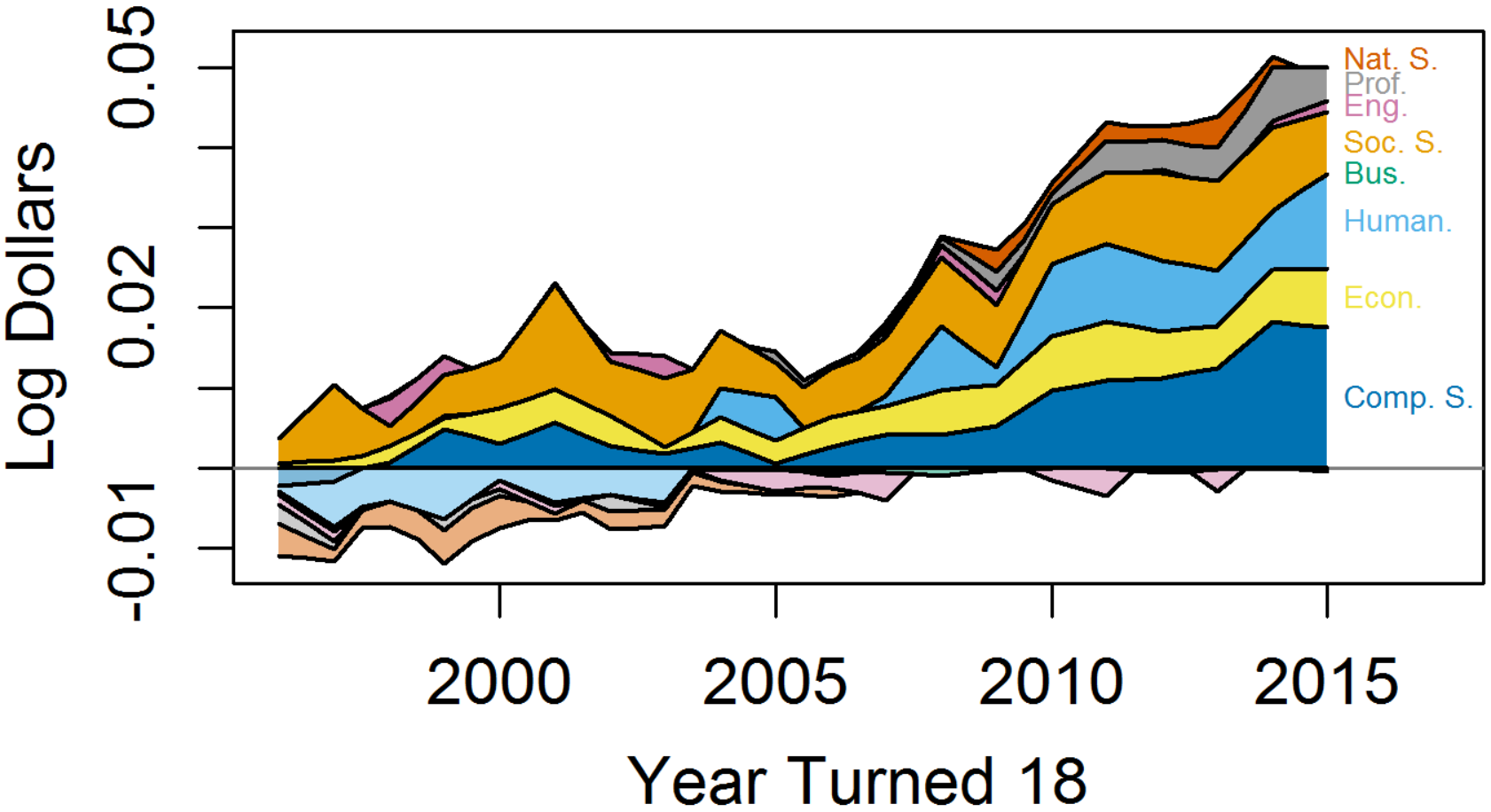
Explanation II: College Majors



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

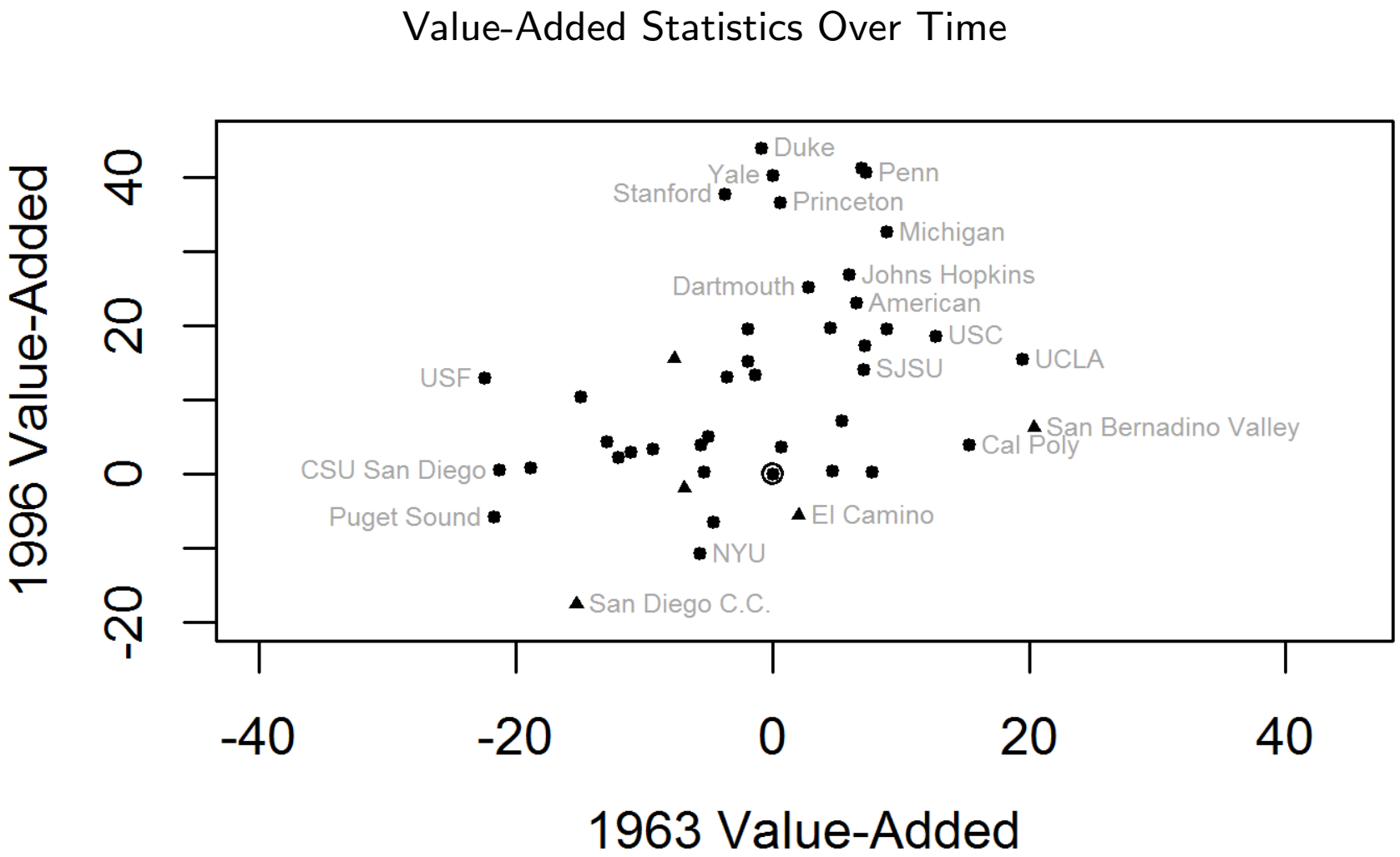
Explanation II: College Majors

Decomposition of Recent Rising Major Regressivity by Discipline



Majors by Parent Income

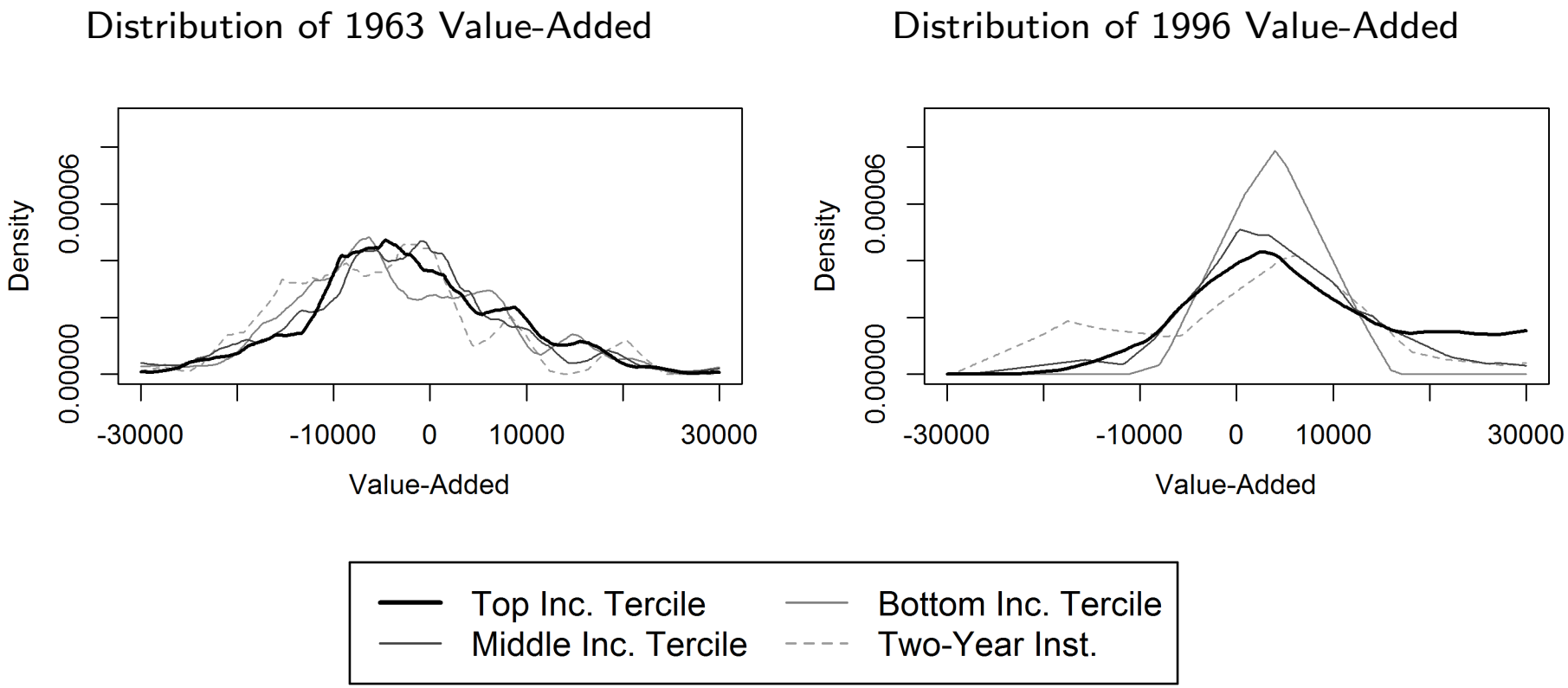
Explanation III: Institutions



Slope: 0.29

1963 Value-Added Map

Explanation III: Institutions



Relationship between average parental income and log wage value-added goes from -0.002 to 0.02 per 10 ranks.

Distribution by Avg. Test Scores

Explanation III: Institutions

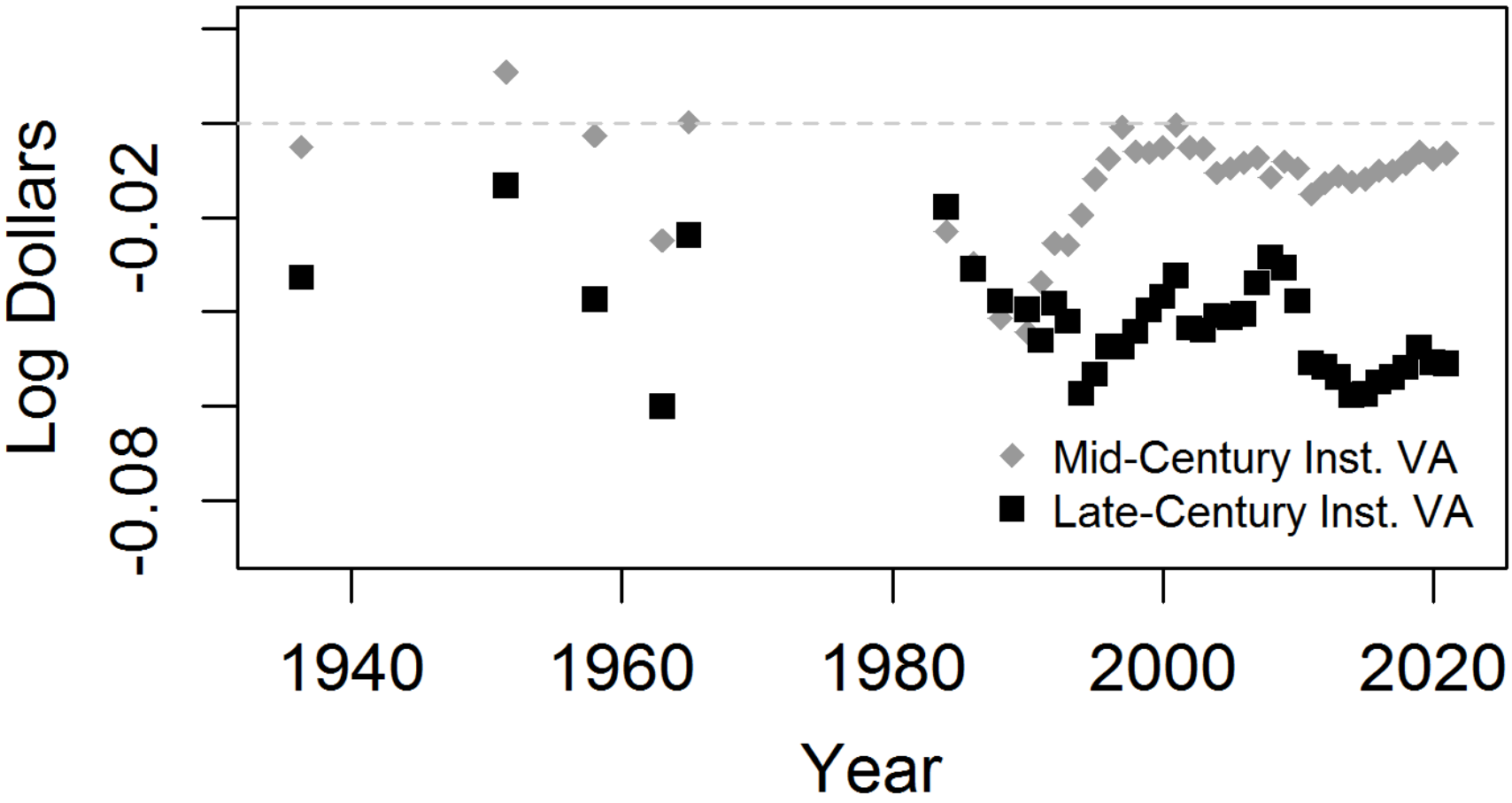
Selection-on-Observables Forecast Coefficients of Institutional Value-Added

Add'l Cov.:	Base.	+Tests	+Grades	+HS FE	+Extra.
		Same Sample			
Inst. FE		0.97	0.95	0.82	0.82
		(0.004)	(0.006)	(0.027)	(0.026)
Obs.			396		
1st Stg. Obs.			22,099		
		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.

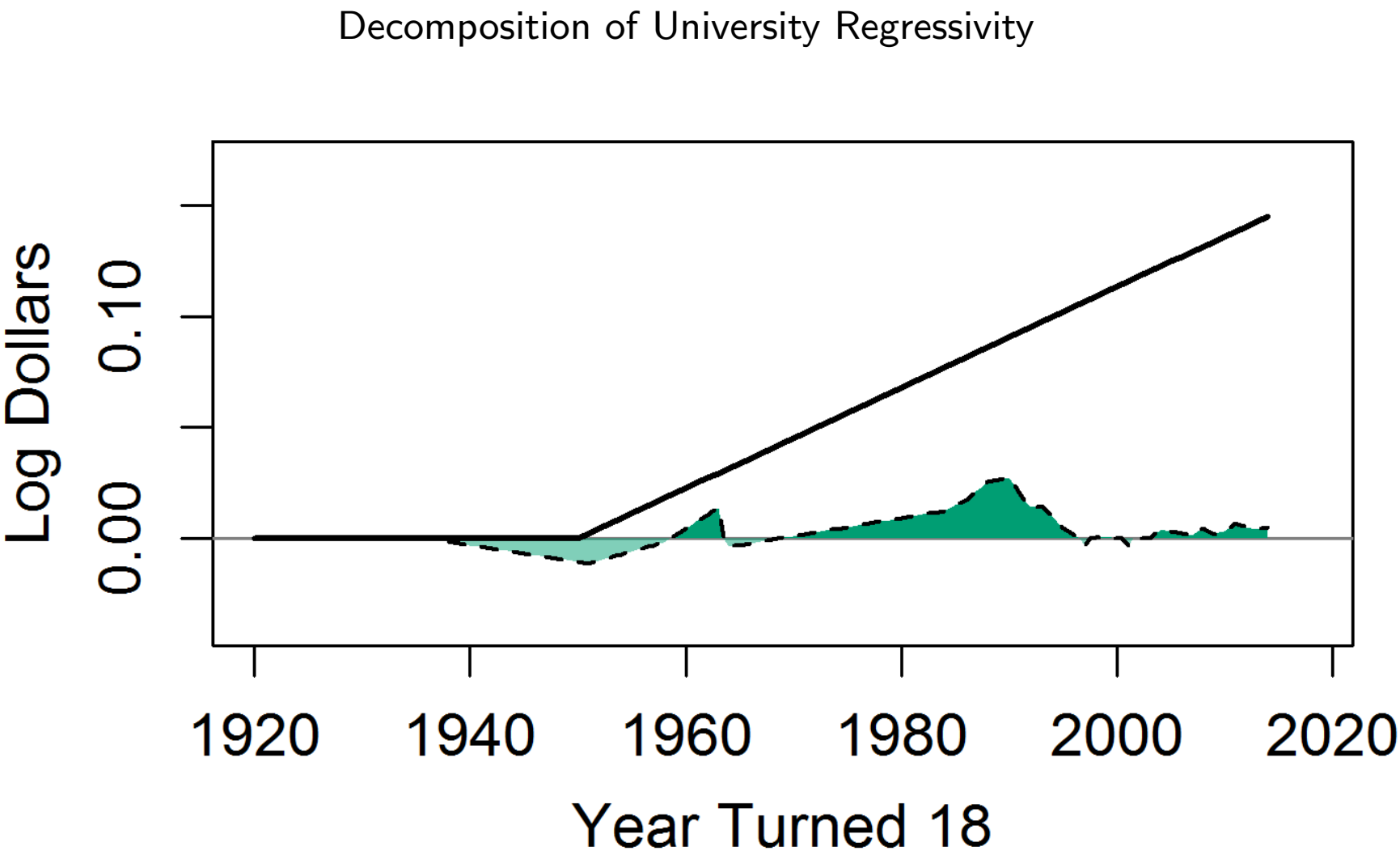
Explanation III: Institutions

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



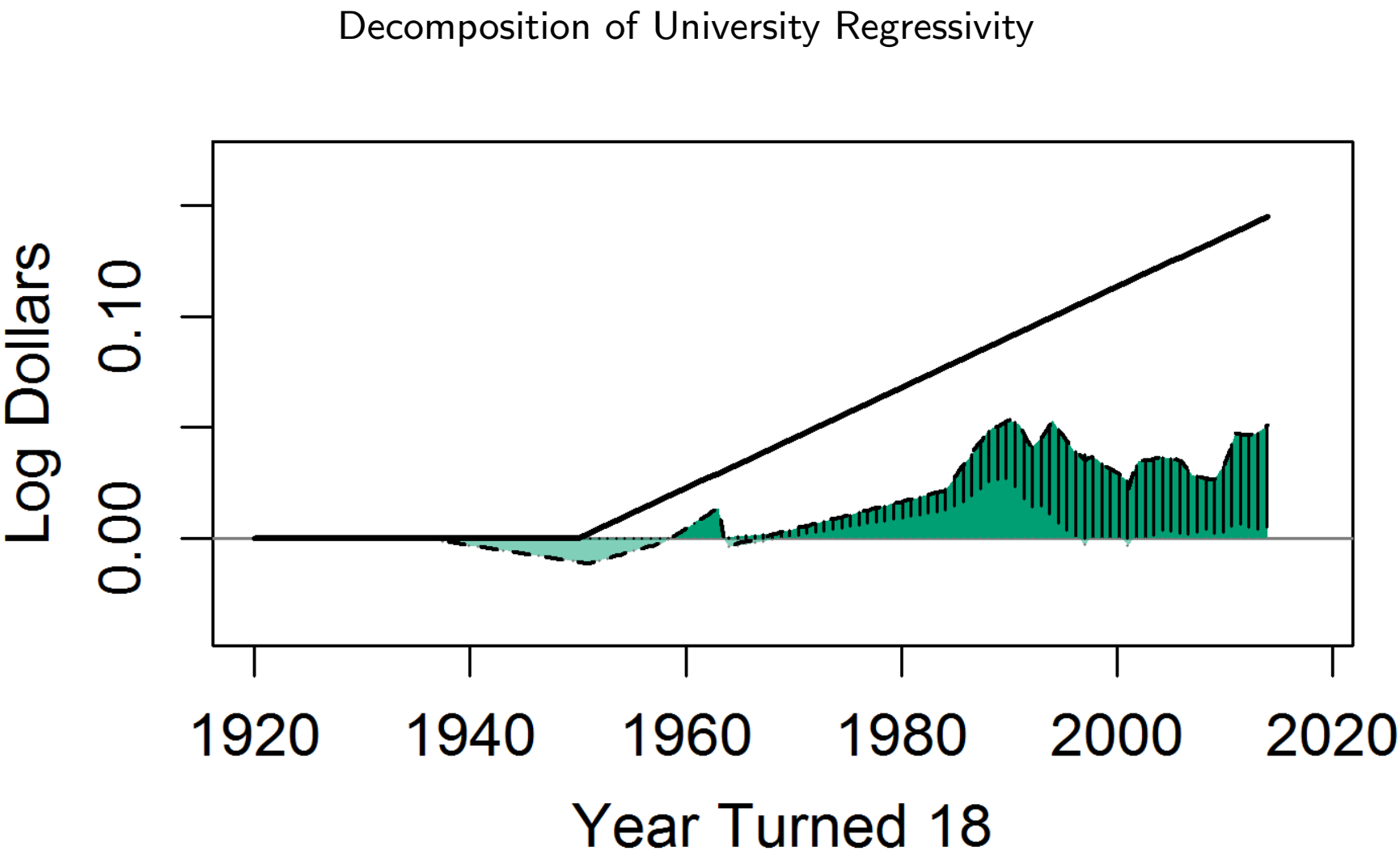
Avg. Income of Pell | Avg. Wage vs. VA

Explanation III: Institutions



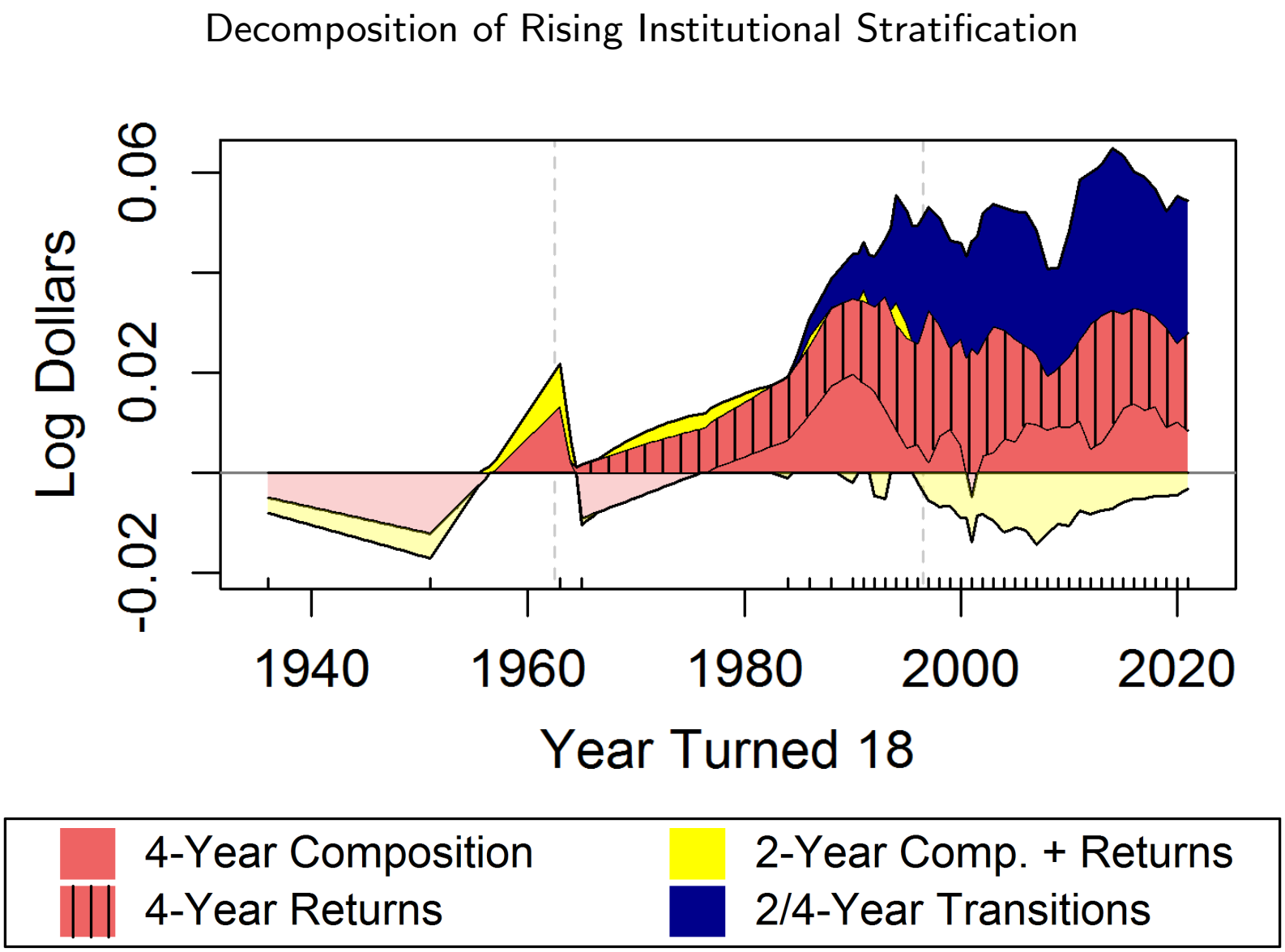
Institution Composition (solid color): $\sum_u v_0^U(u)\Delta_q\left[P_t(u|q)\right]$.

Explanation III: Institutions



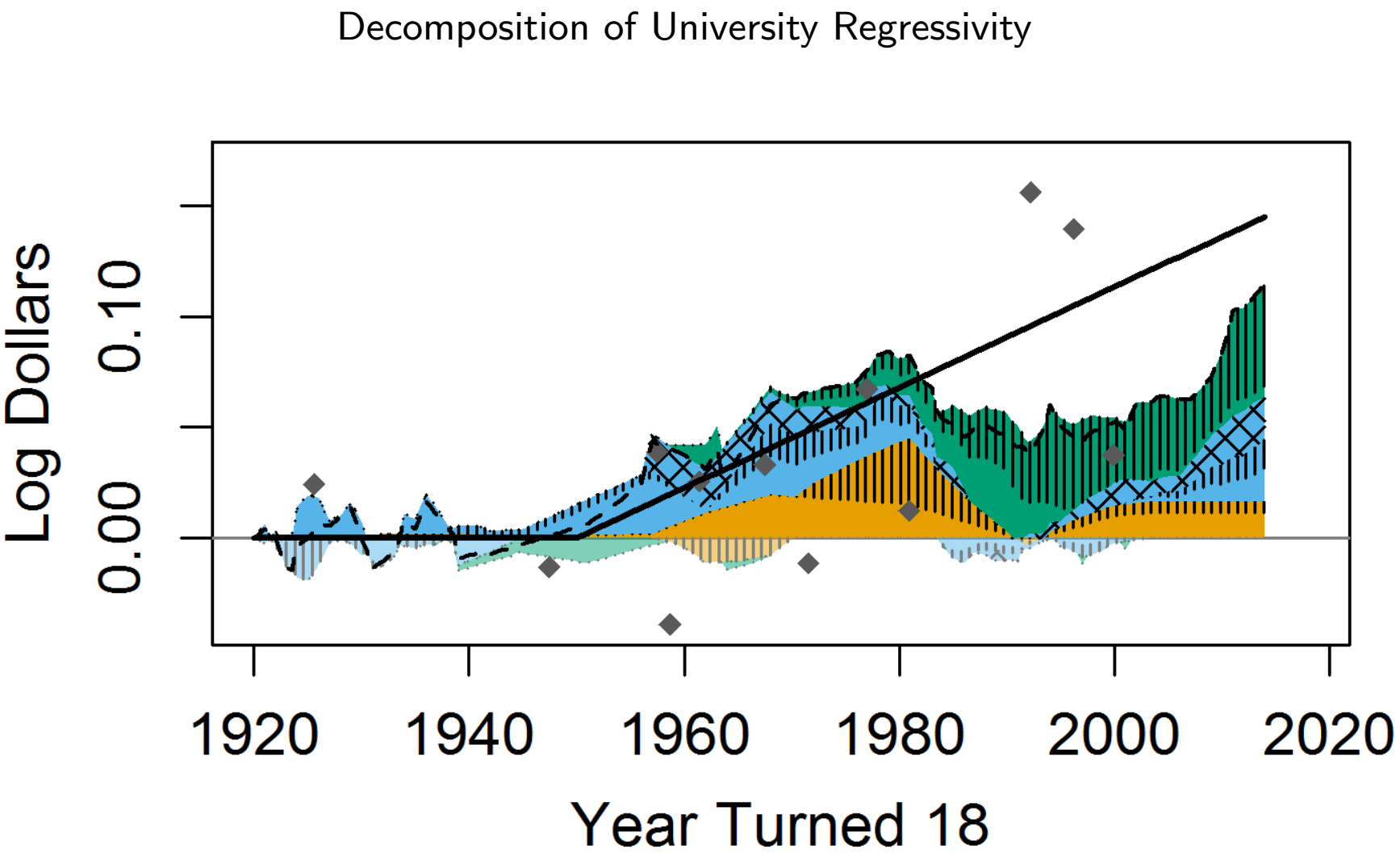
Inst. Value-Added (stripe): $\sum_u \Delta_q \left[P_t(u|q) \right] \left(v_t^U(u) - v_0^U(u) \right)$

Explanation III: Institutions



CC Enrollment

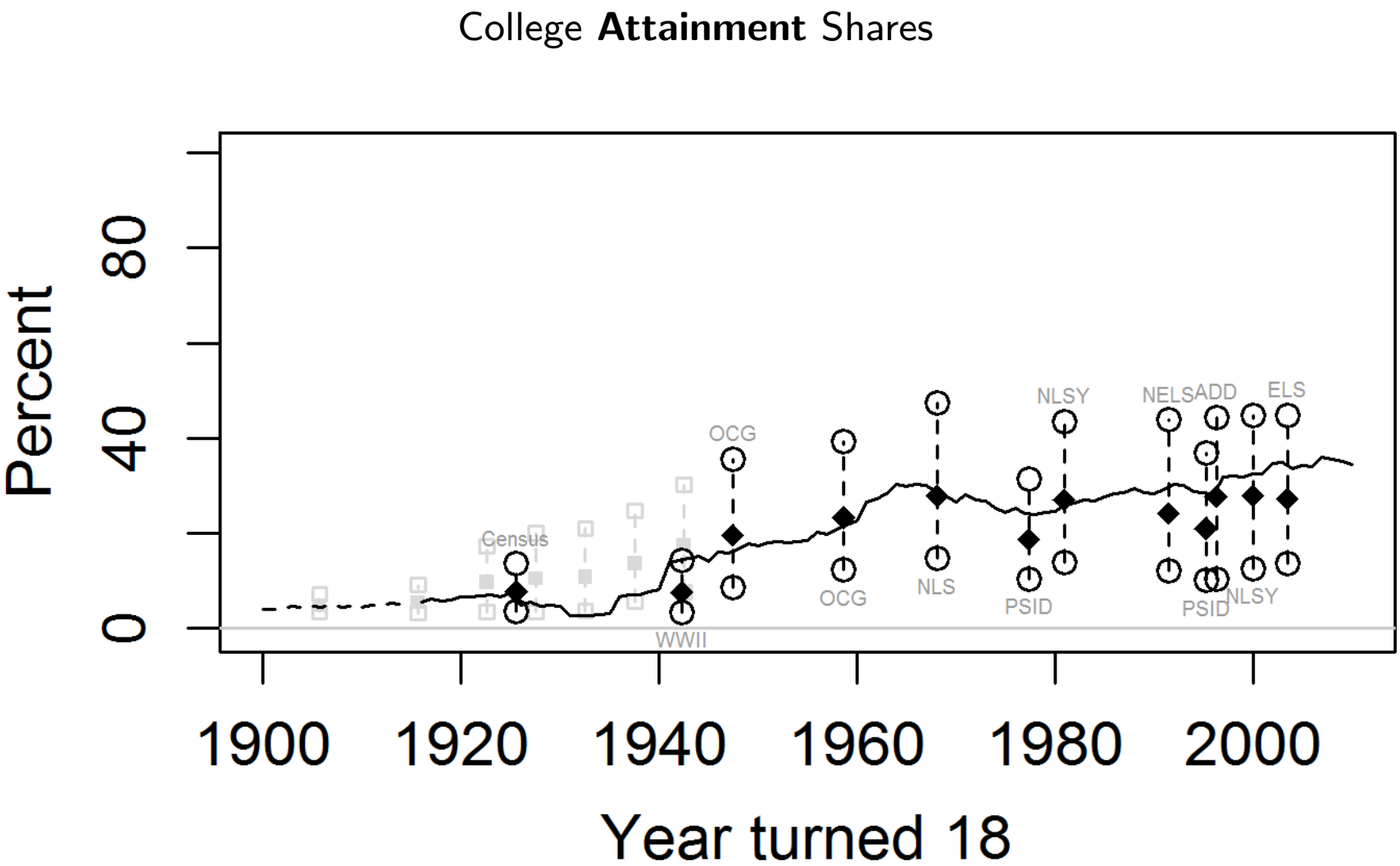
Overall Decomposition



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.

College Attainment in the United States by Income Tercile

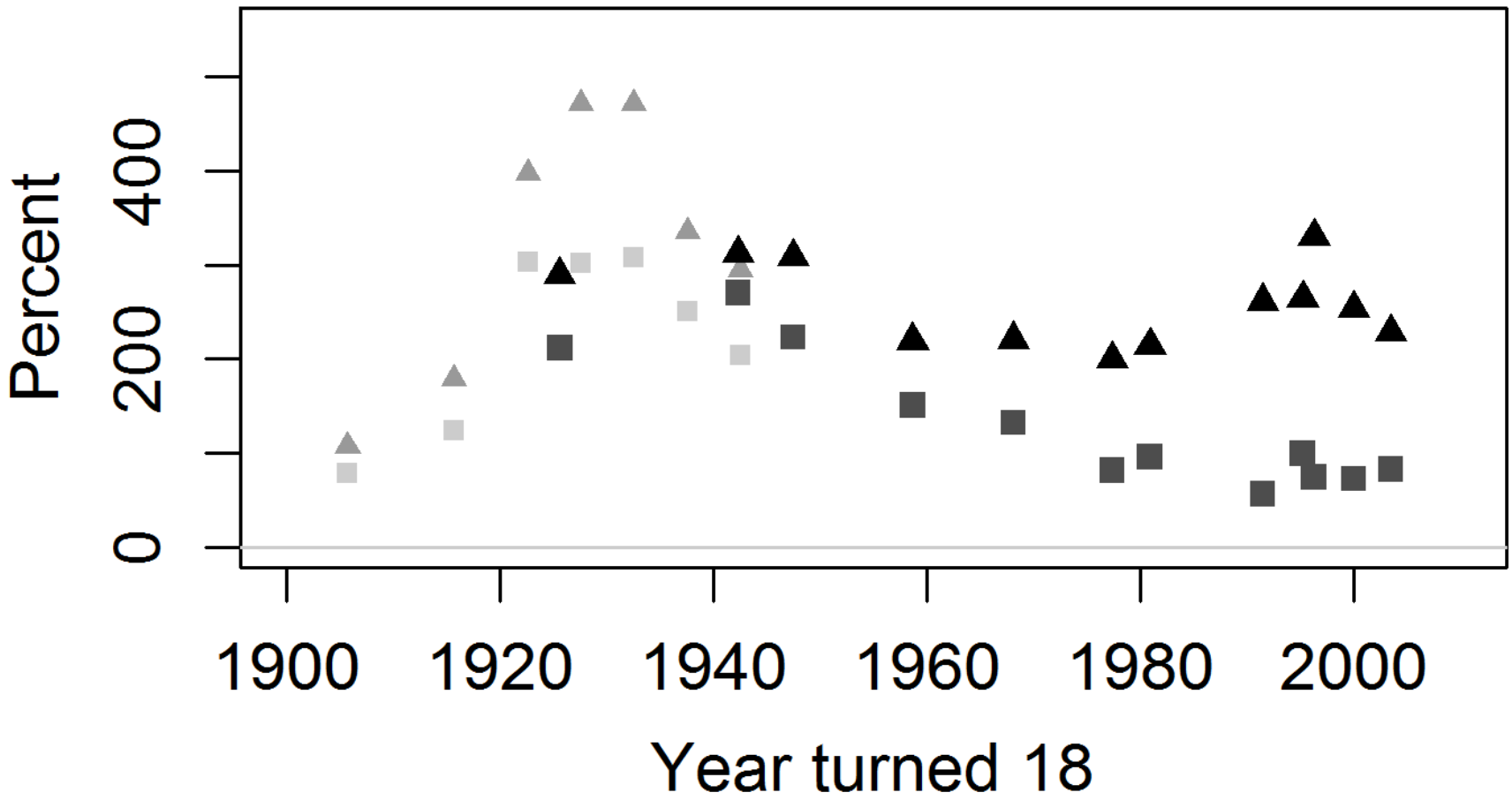


Similar pattern when looking at four-year college attainment.

Back

College Enrollment in the United States by Income Tercile

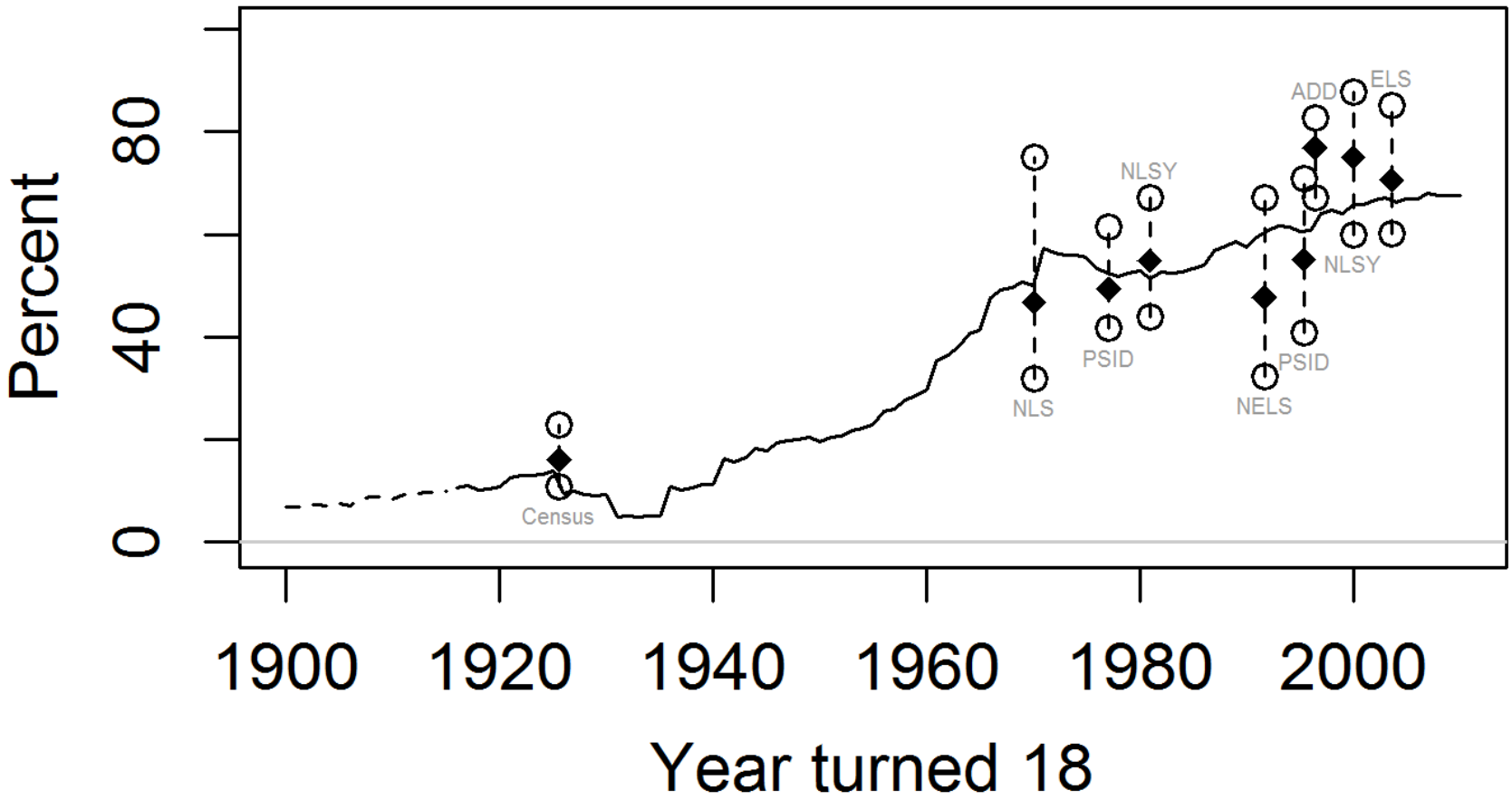
Top-Tercile Students' Percent Higher College-Going Relative to Bottom-Tercile Students



Back

College Enrollment in the United States by Income Tercile

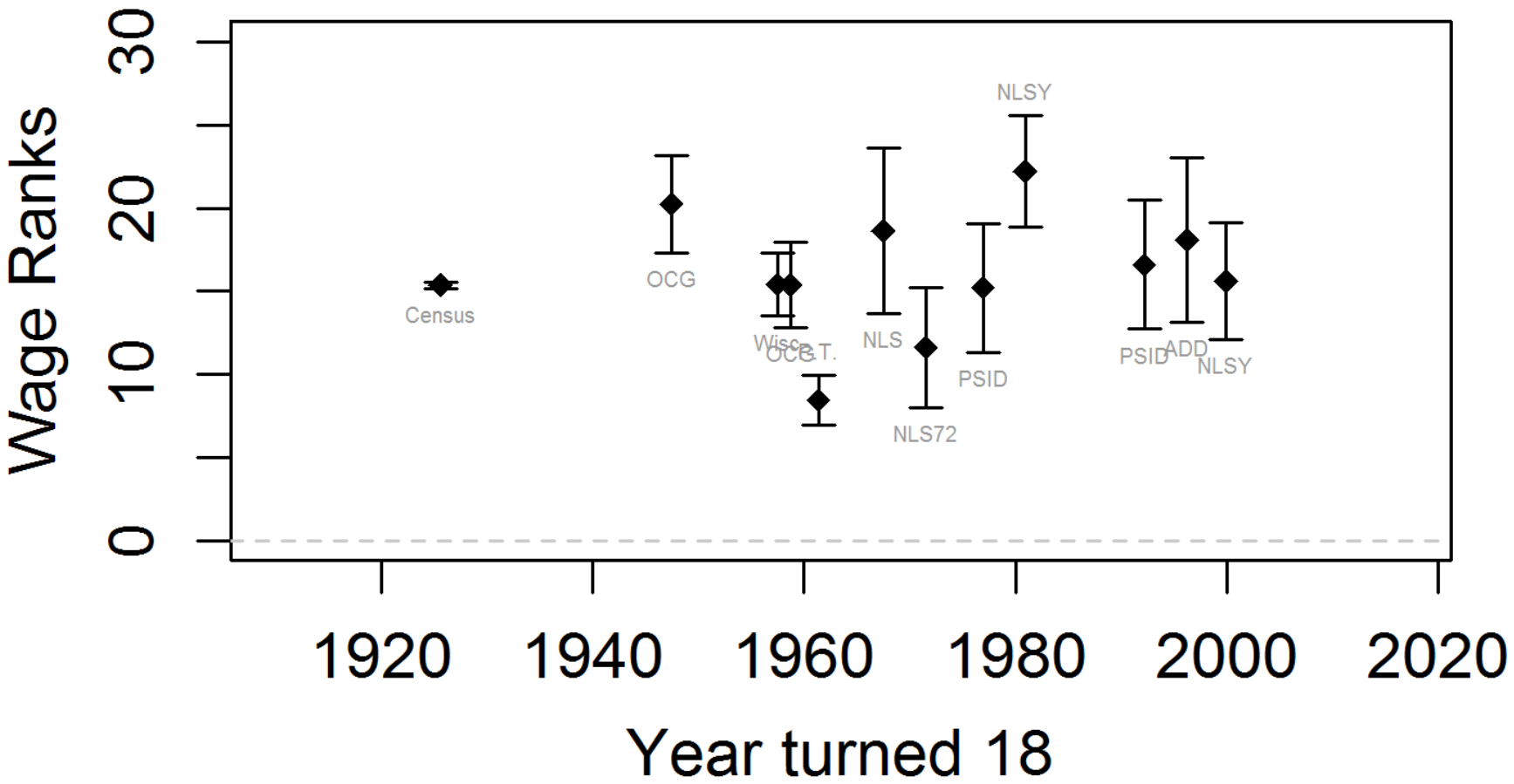
College Enrollment Among Female Students



Back

The Regressivity of U.S. Higher Education

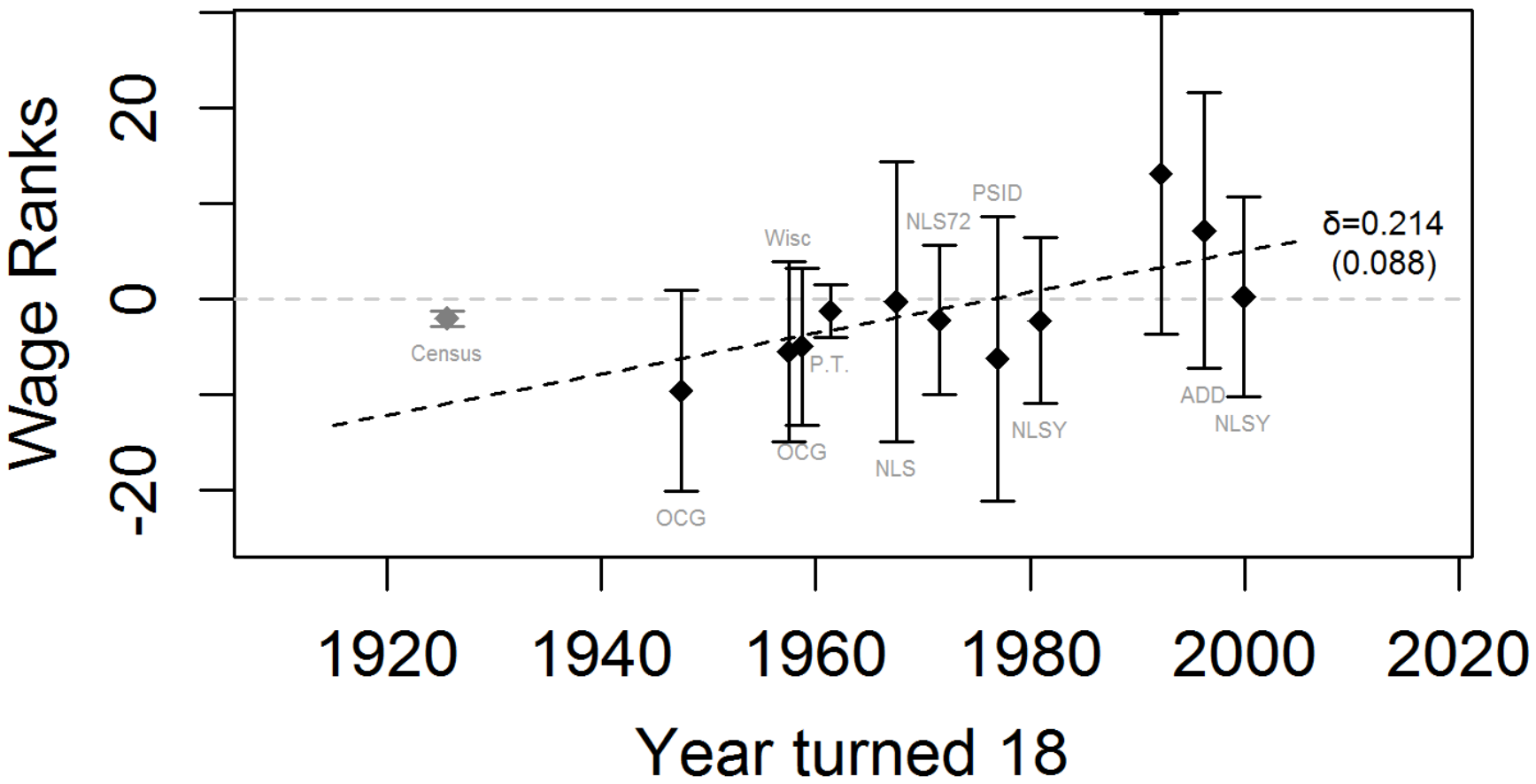
Average Return to College **Attainment** (β_t)



Log Wages

The Regressivity of U.S. Higher Education

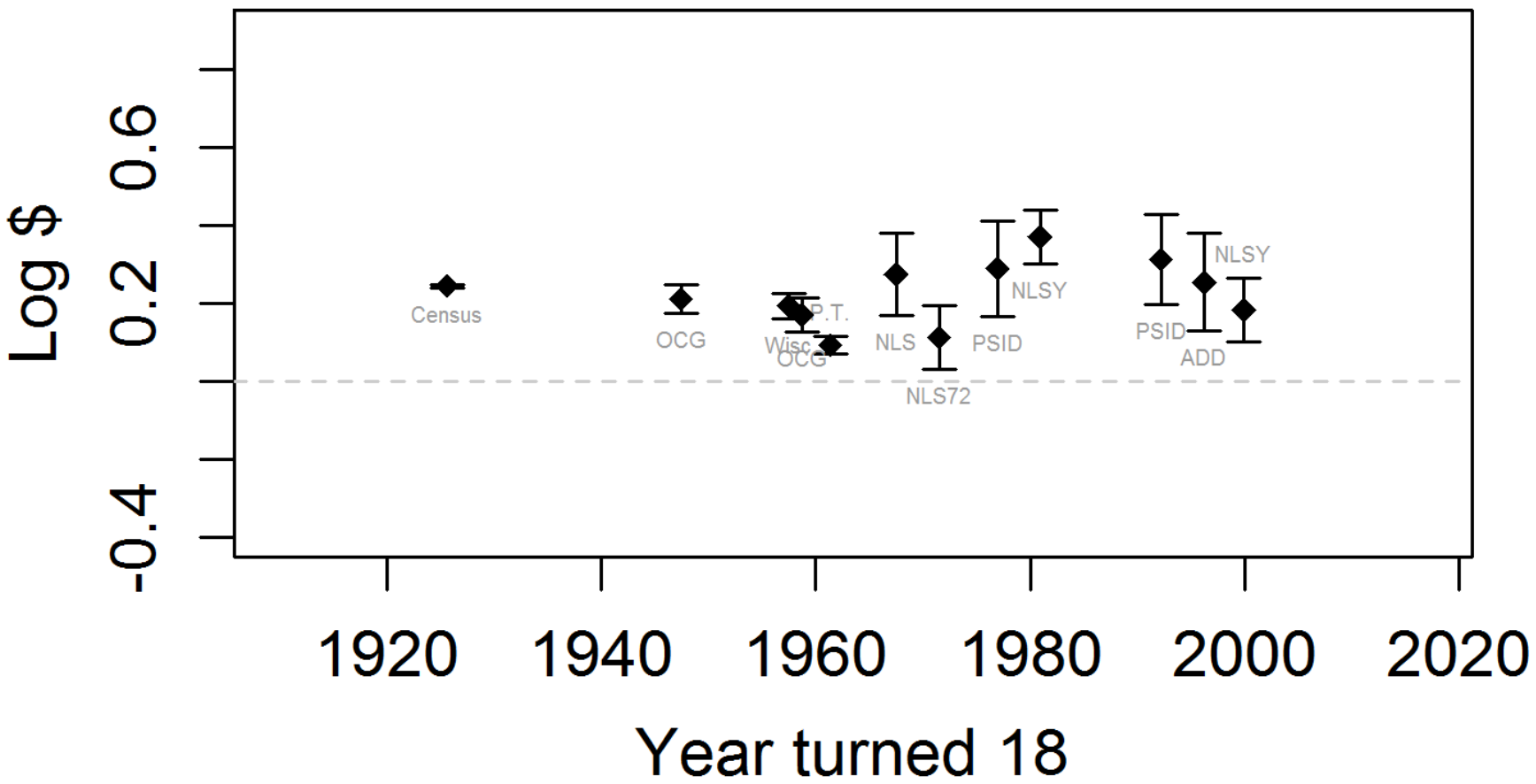
Differential Return to College **Attainment** by SES (δ_t)



Back

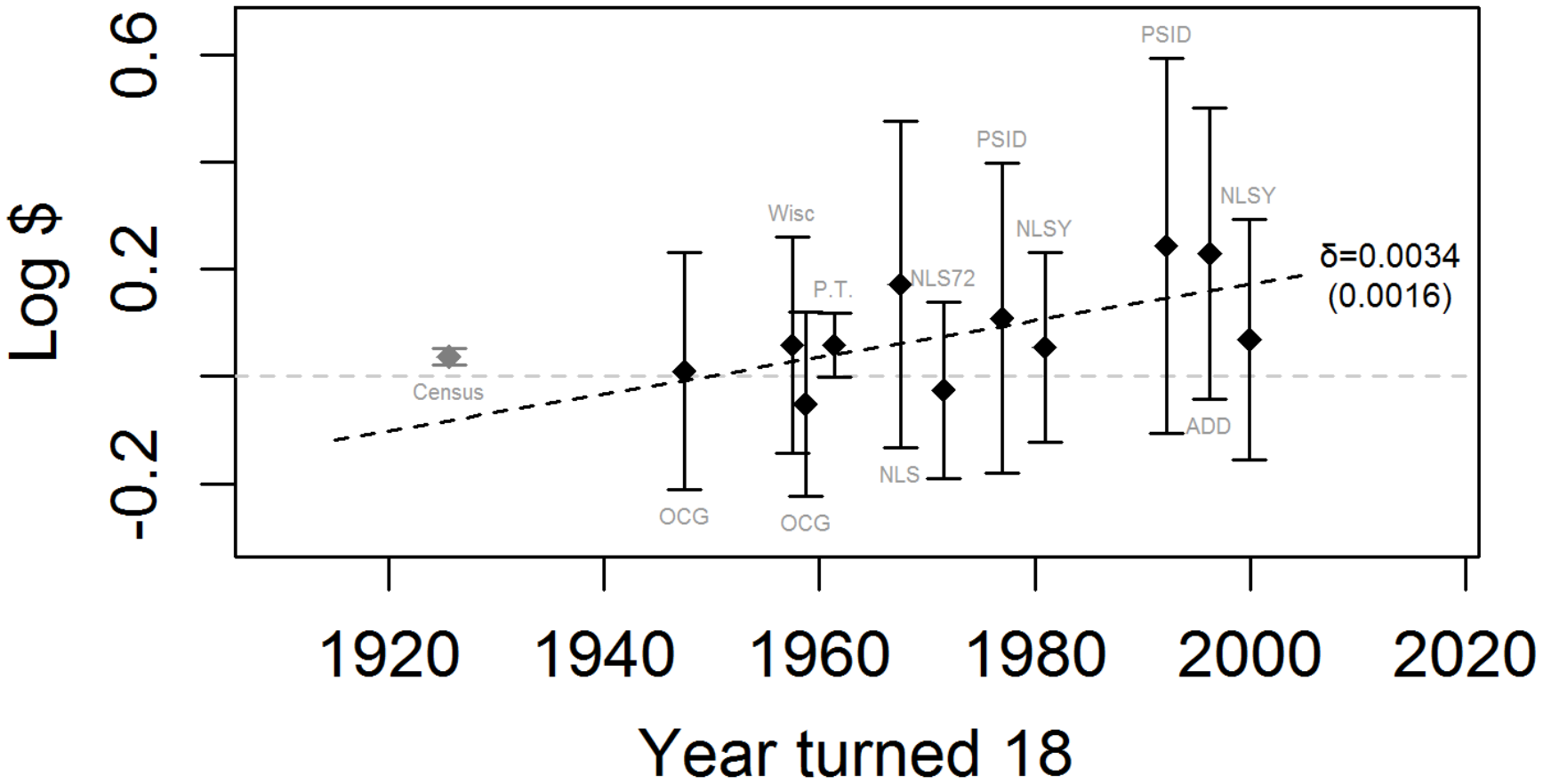
The Regressivity of U.S. Higher Education

Average Return to College Enrollment **in Log $** ($\beta_t$)



The Regressivity of U.S. Higher Education

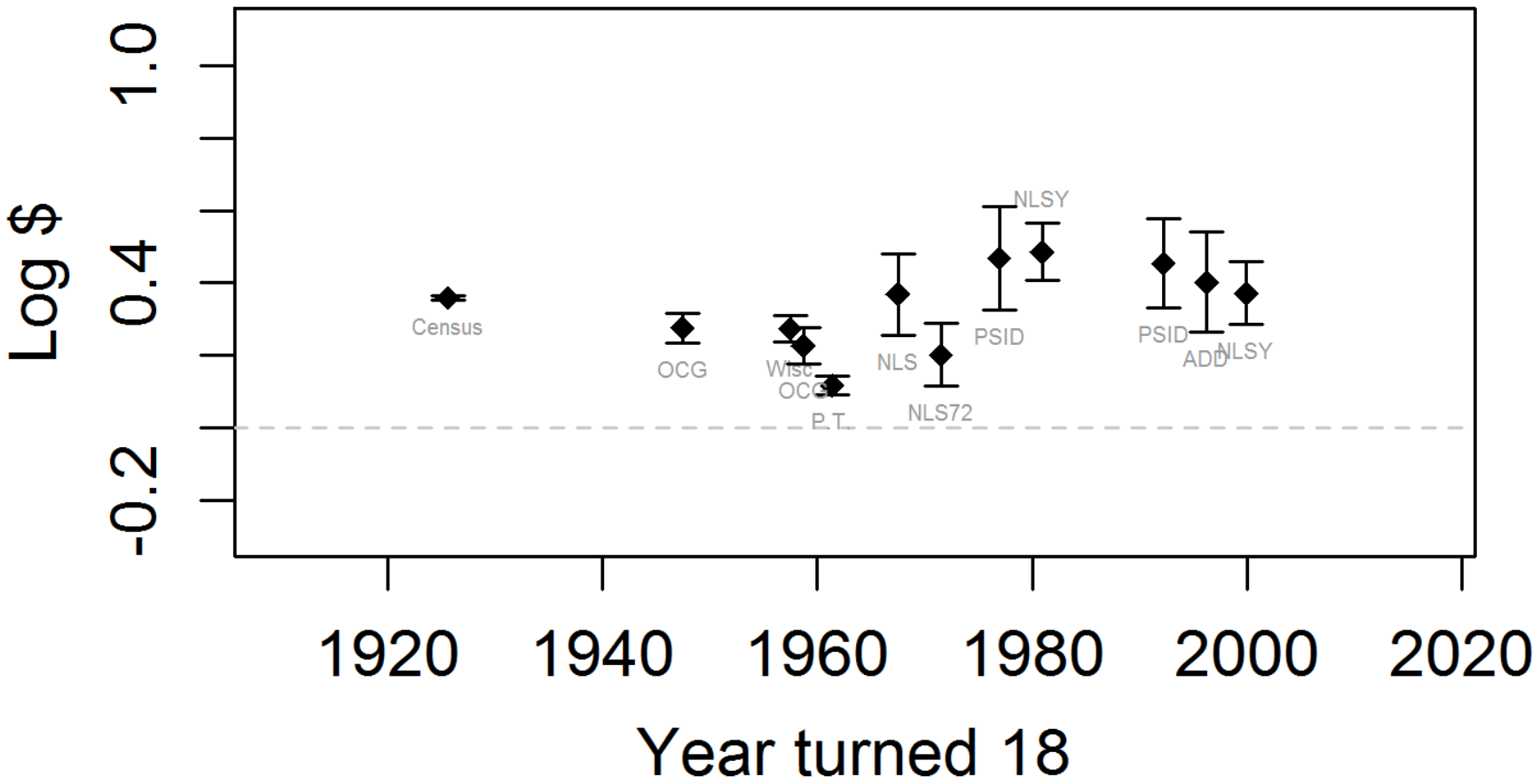
Differential Return to College Enrollment by SES **in Log $** ($\delta_t$)



Back

The Regressivity of U.S. Higher Education

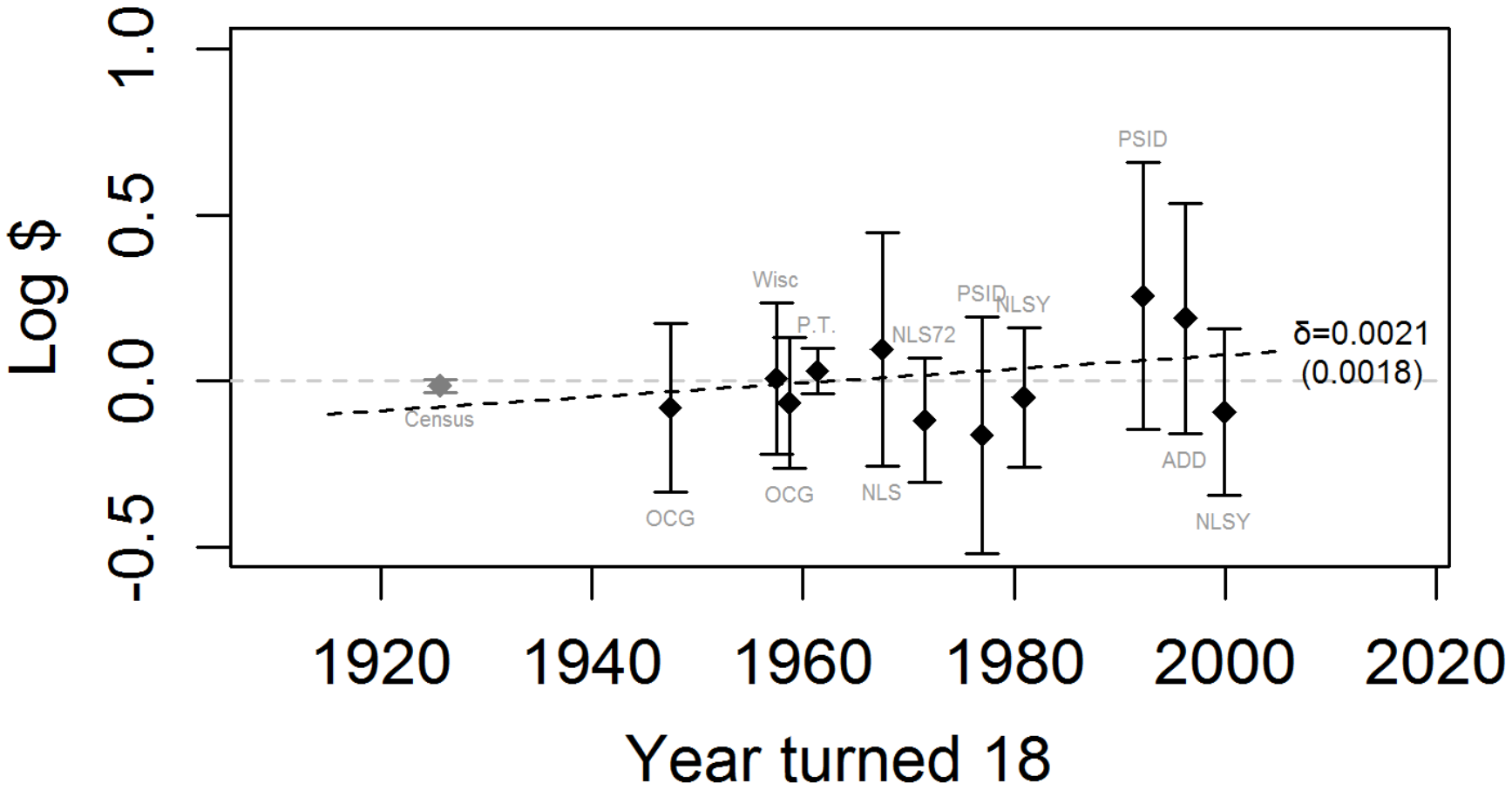
Average Return to College **Attainment in Log $** ($\beta_t$)



Back

The Regressivity of U.S. Higher Education

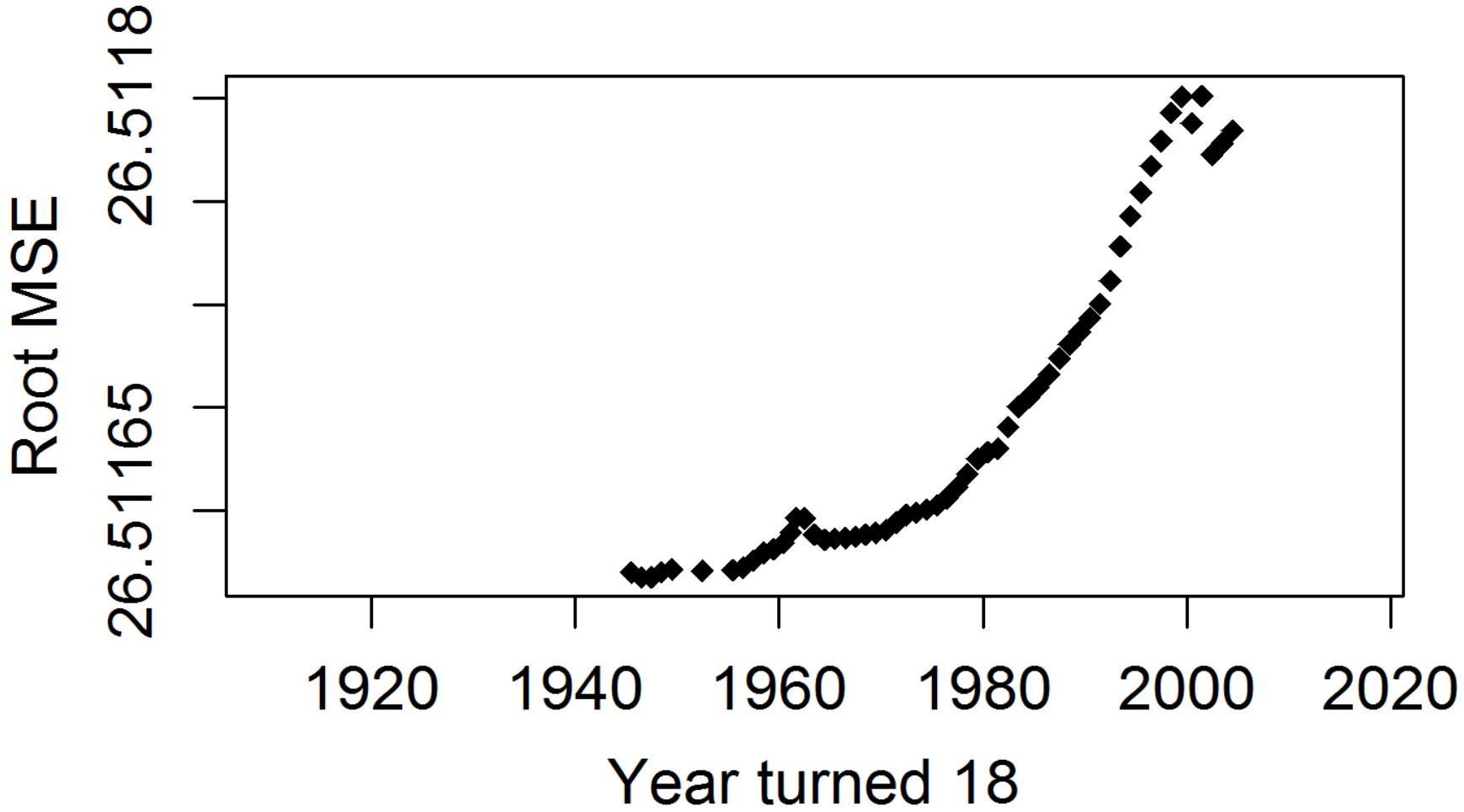
Differential Return to College **Attainment** by SES **in Log $** ($\delta_t$)



Back

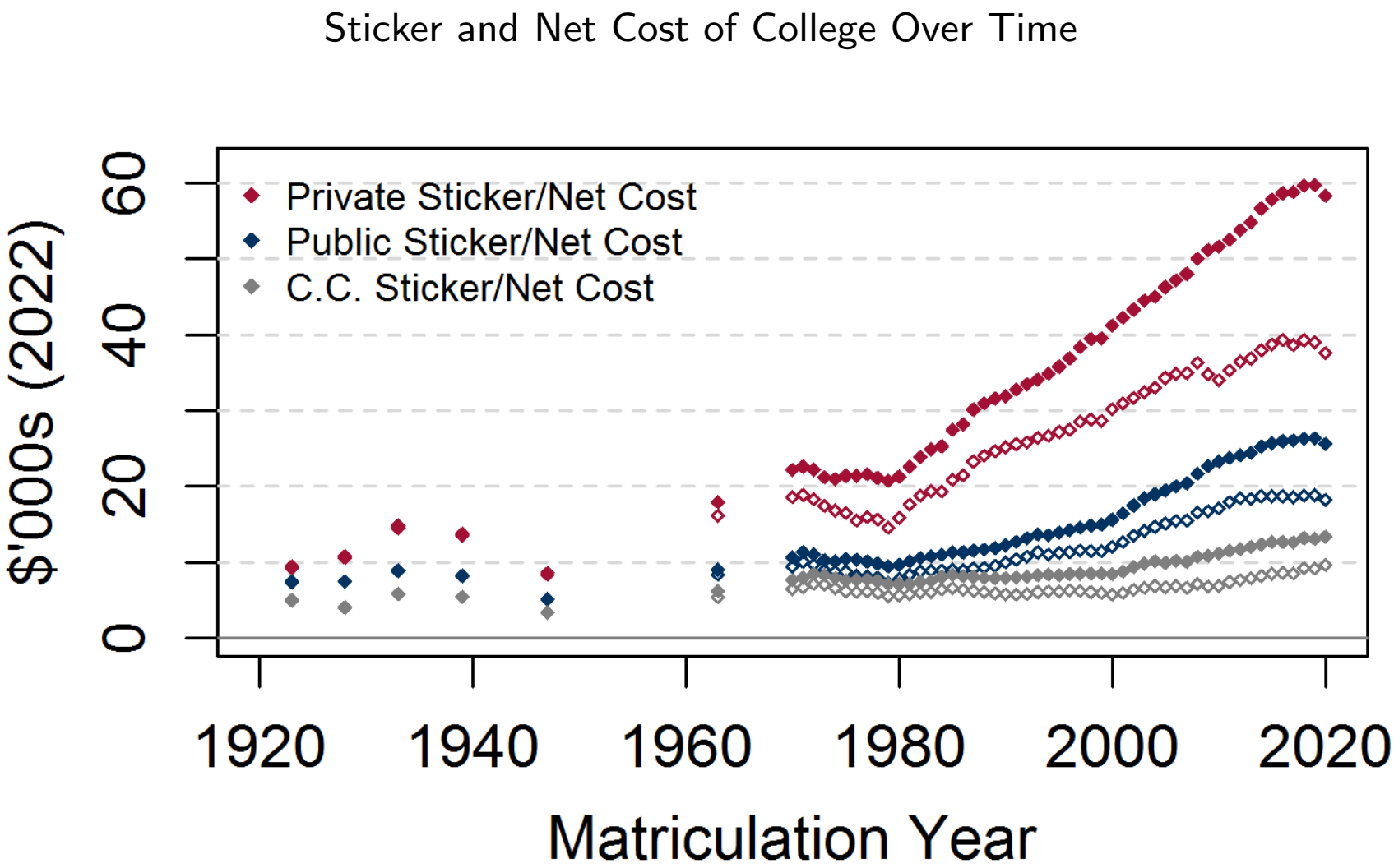
The Regressivity of U.S. Higher Education

Root Mean Squared Error from Regressivity Models with Kink Points

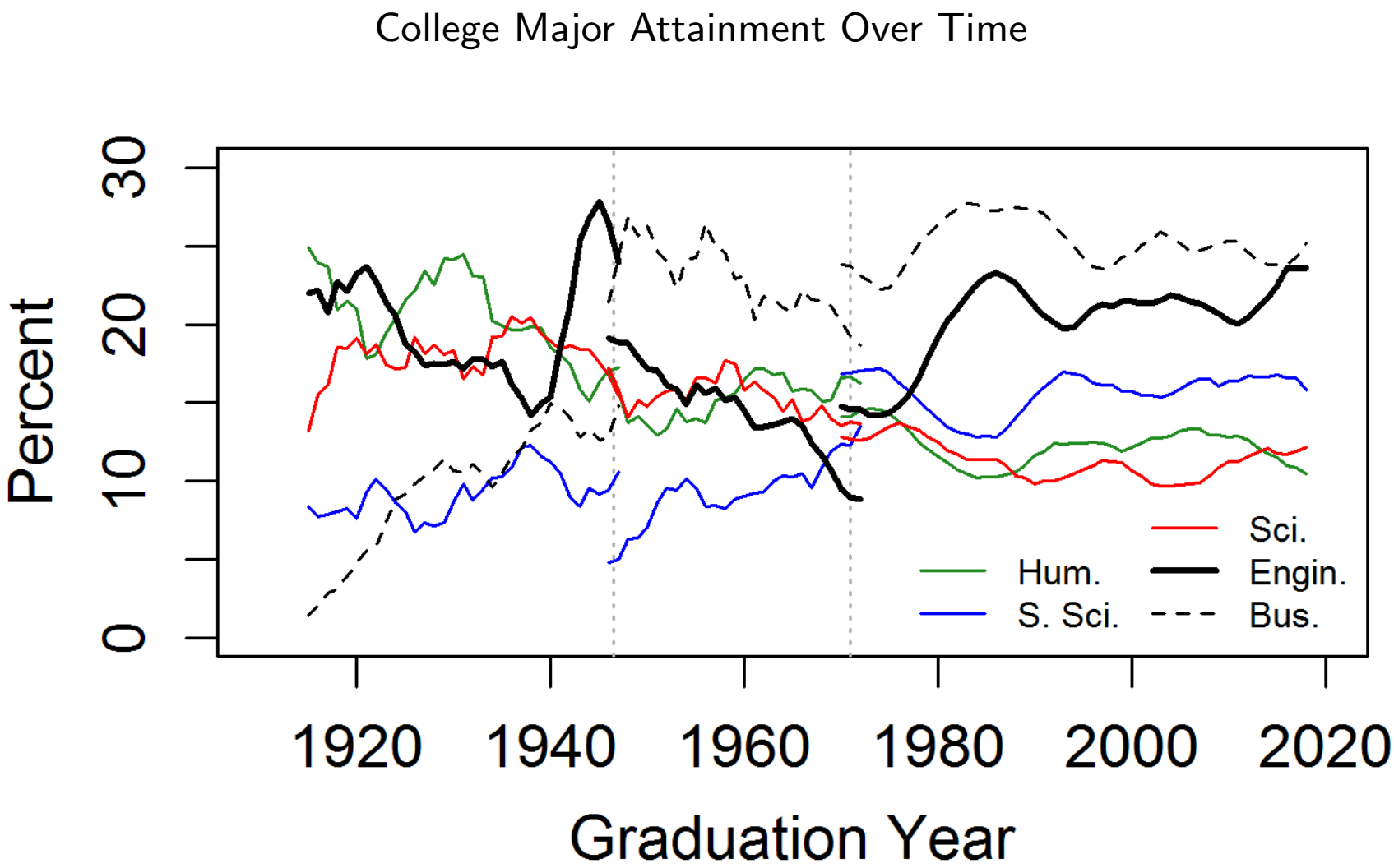


Back

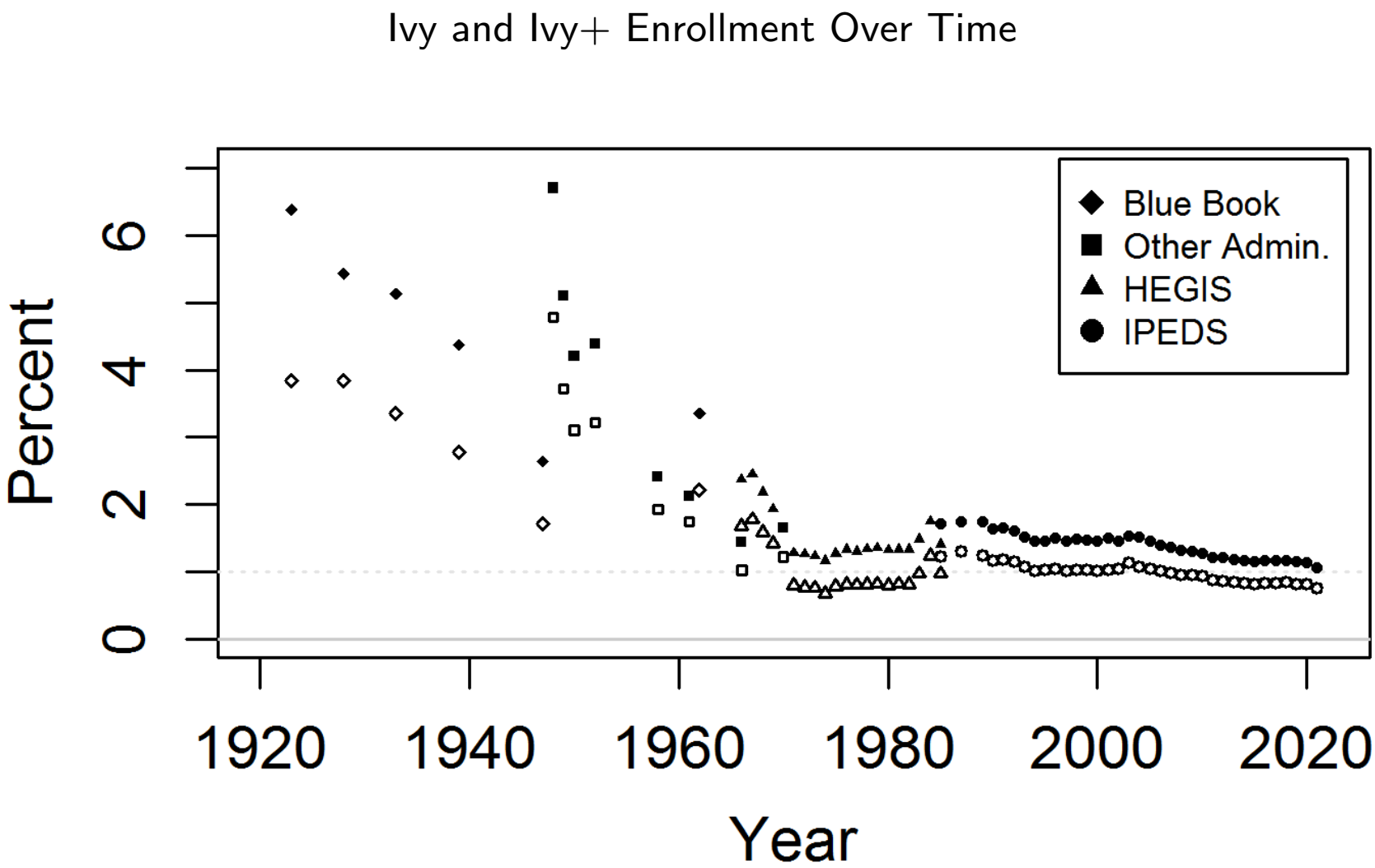
High-Level Trends in US Higher Education



Explanation I: Selection



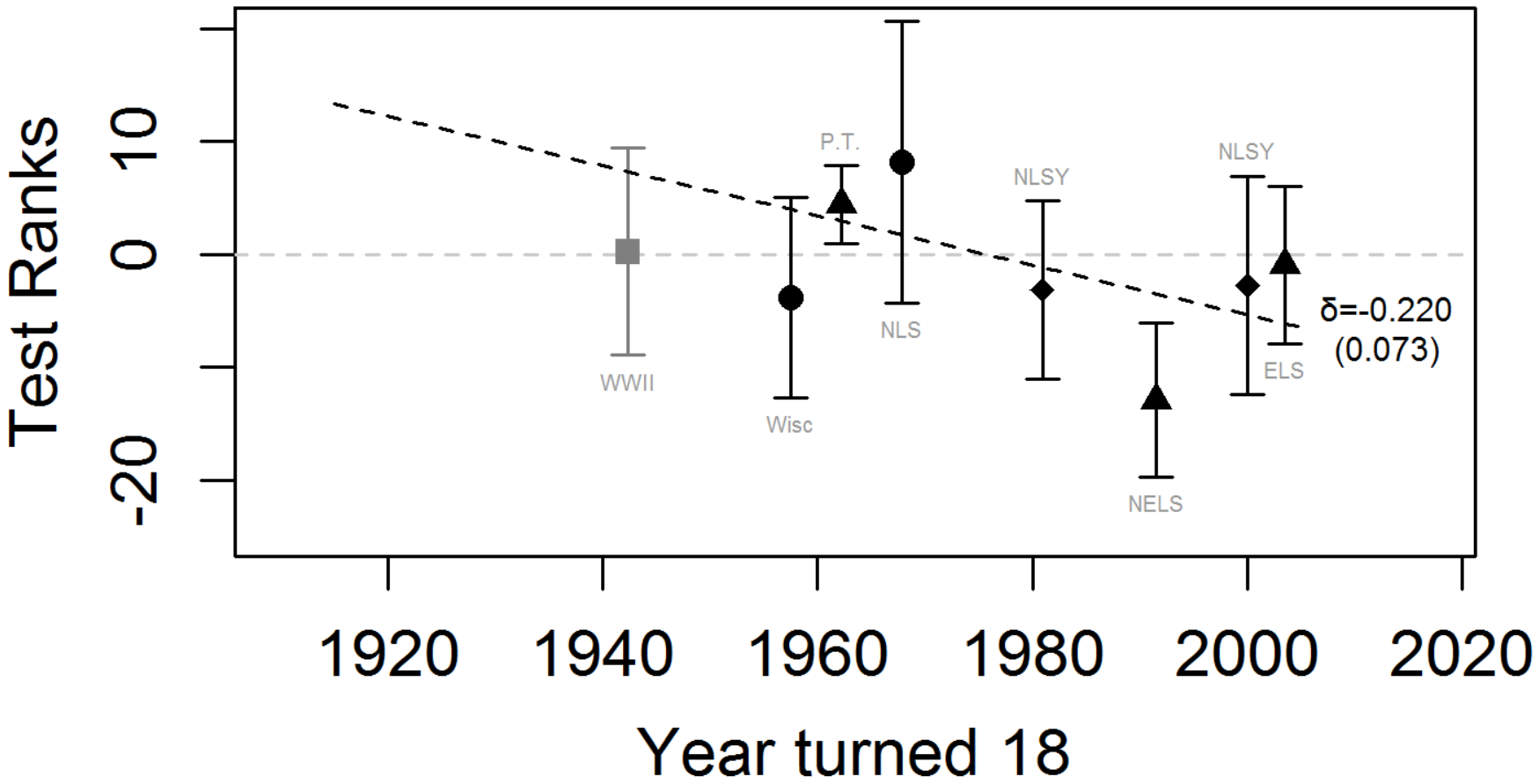
Explanation I: Selection



Back

Explanation I: Selection

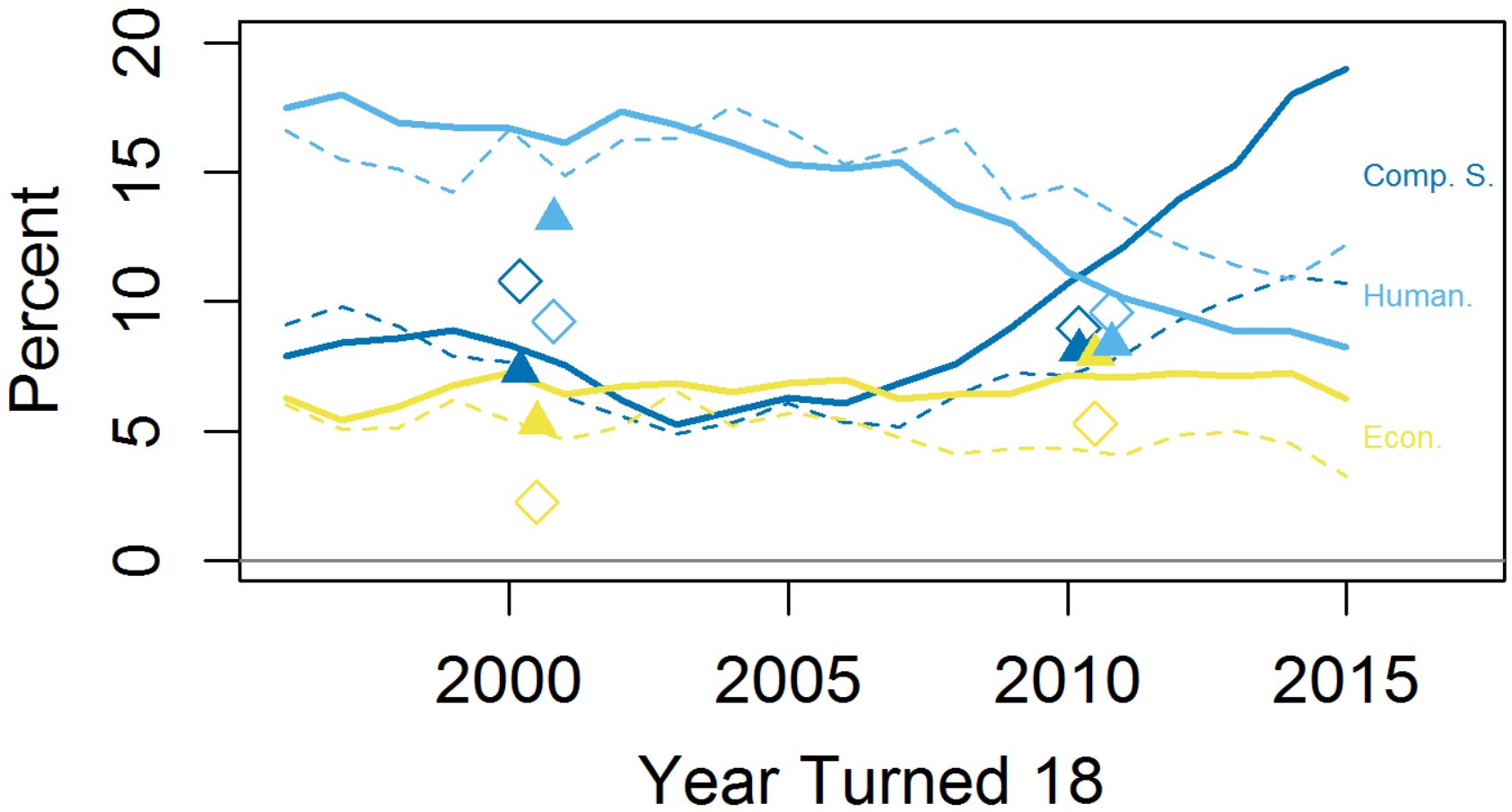
Pre-College Human Capital-Based Selection into College Attainment



Back

Explanation II: College Majors

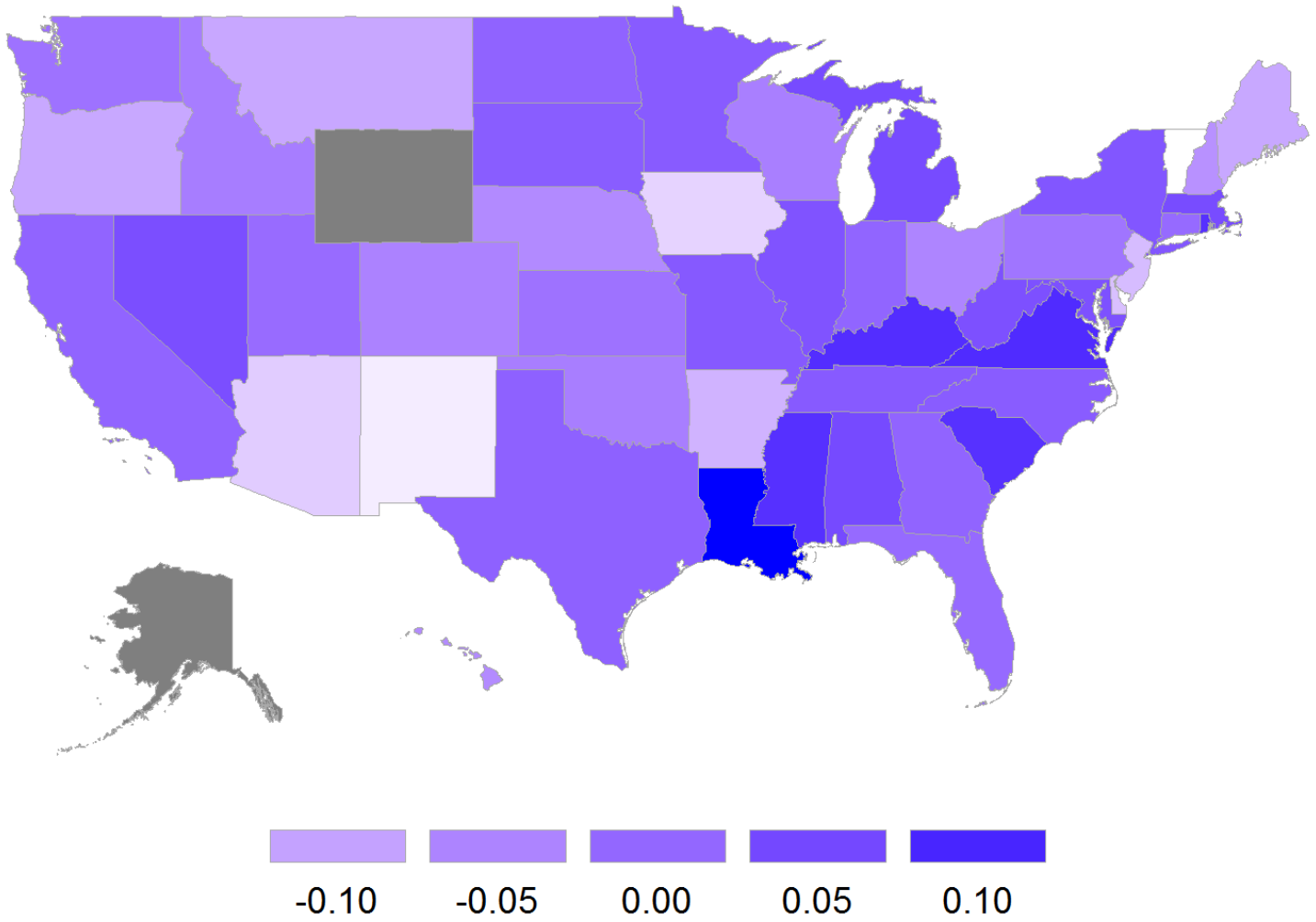
Annual Share of Declared College Majors by Parental Income Tercile



Back

Explanation III: Institutions

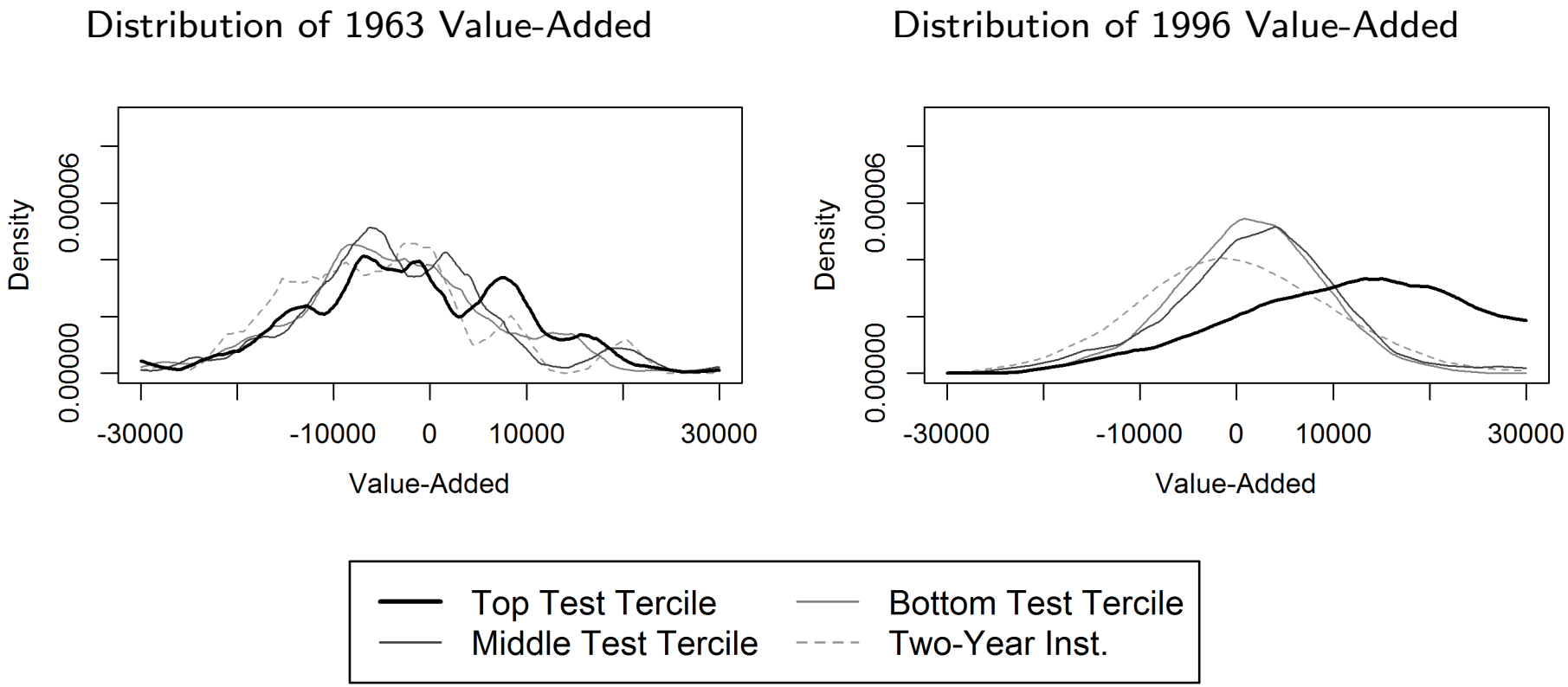
Geographic Distribution of Mid-Century Institutional Value-Added



Back

Explanation III: Institutions

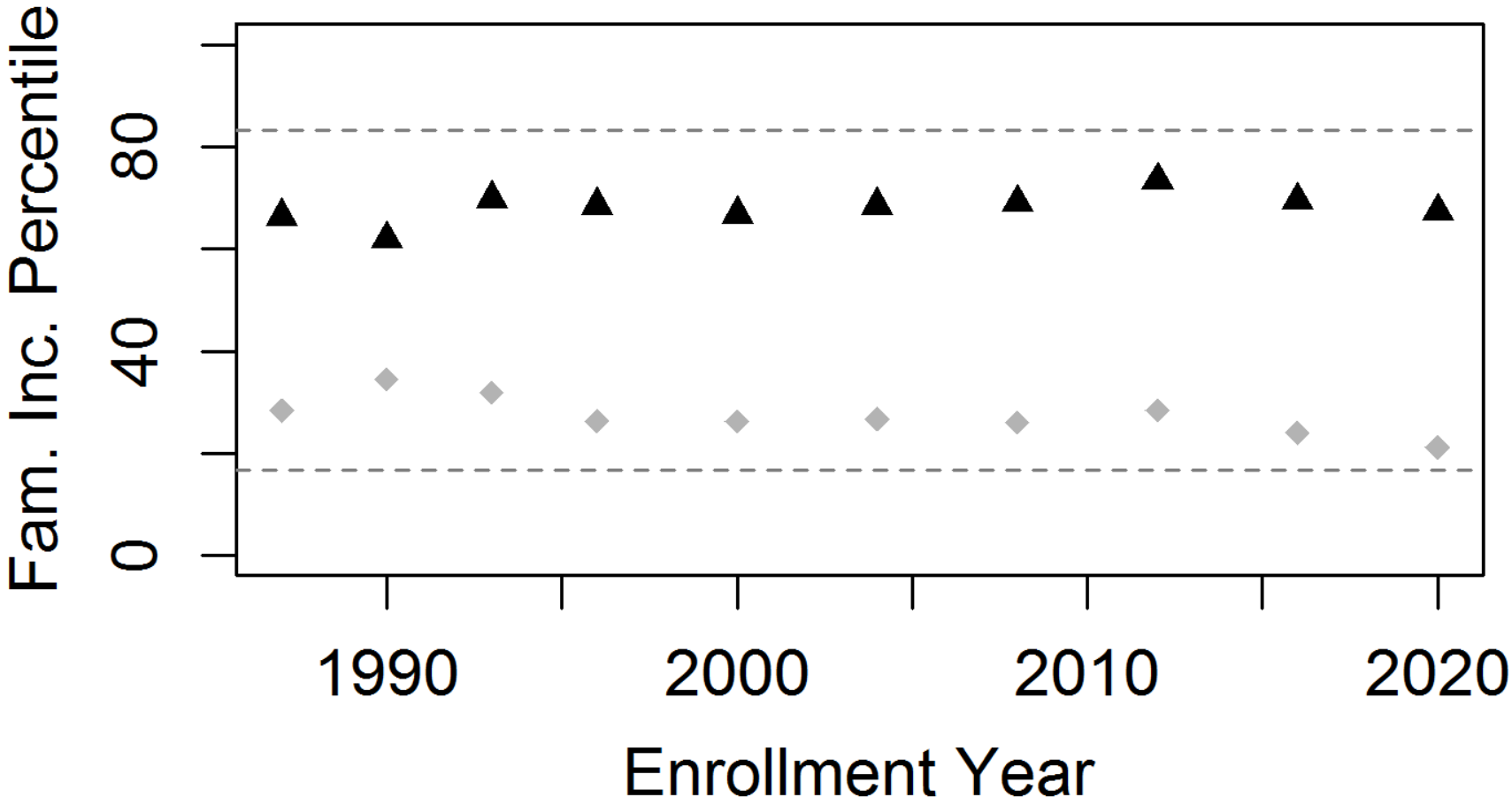
Institutional Value-Added **by testing tercile** in the 1960s and 1990s



Back

Explanation III: Institutions

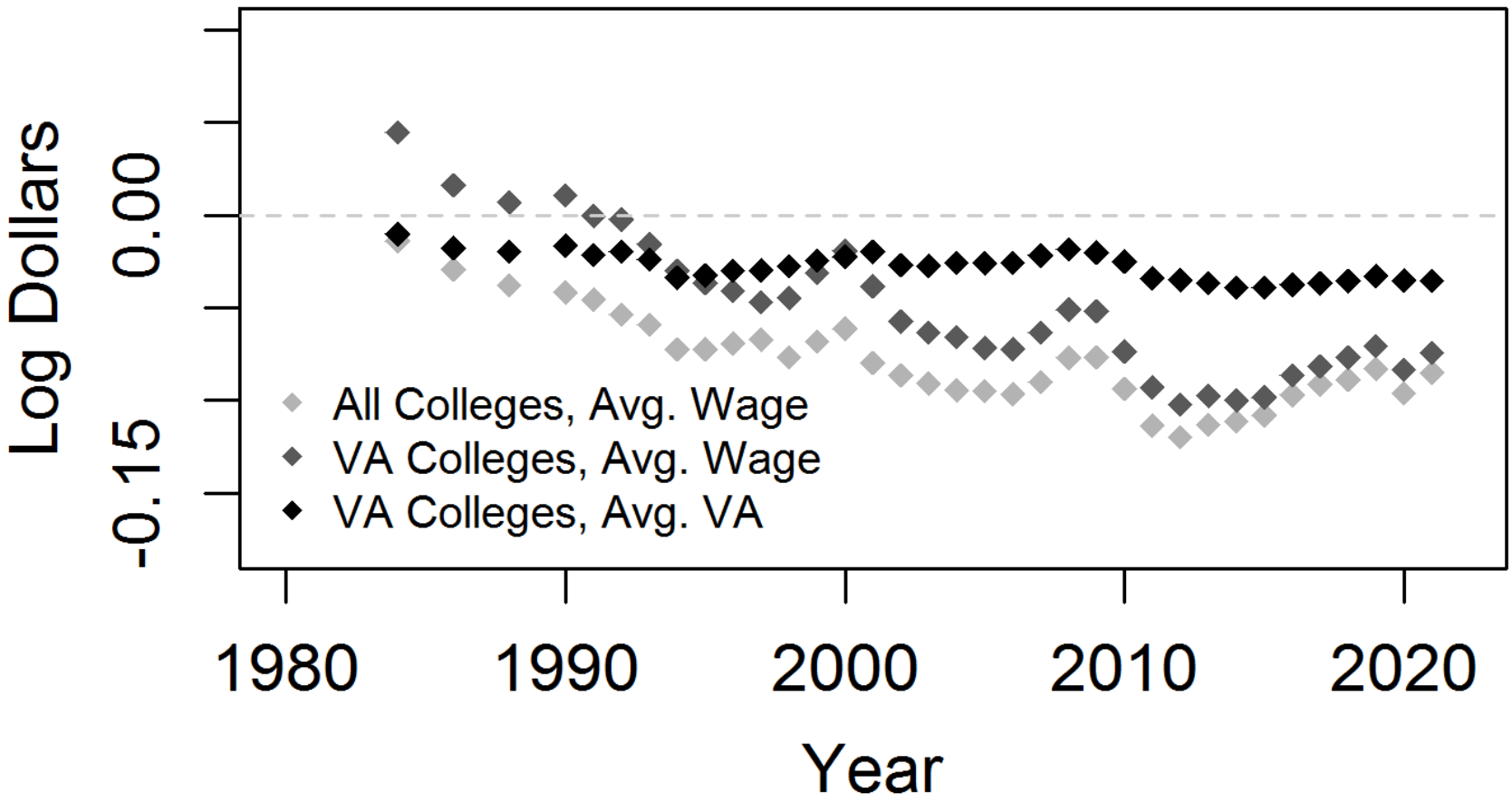
Average Parental Income of Pell and Non-Pell Students (NPSAS)



Back

Explanation III: Institutions

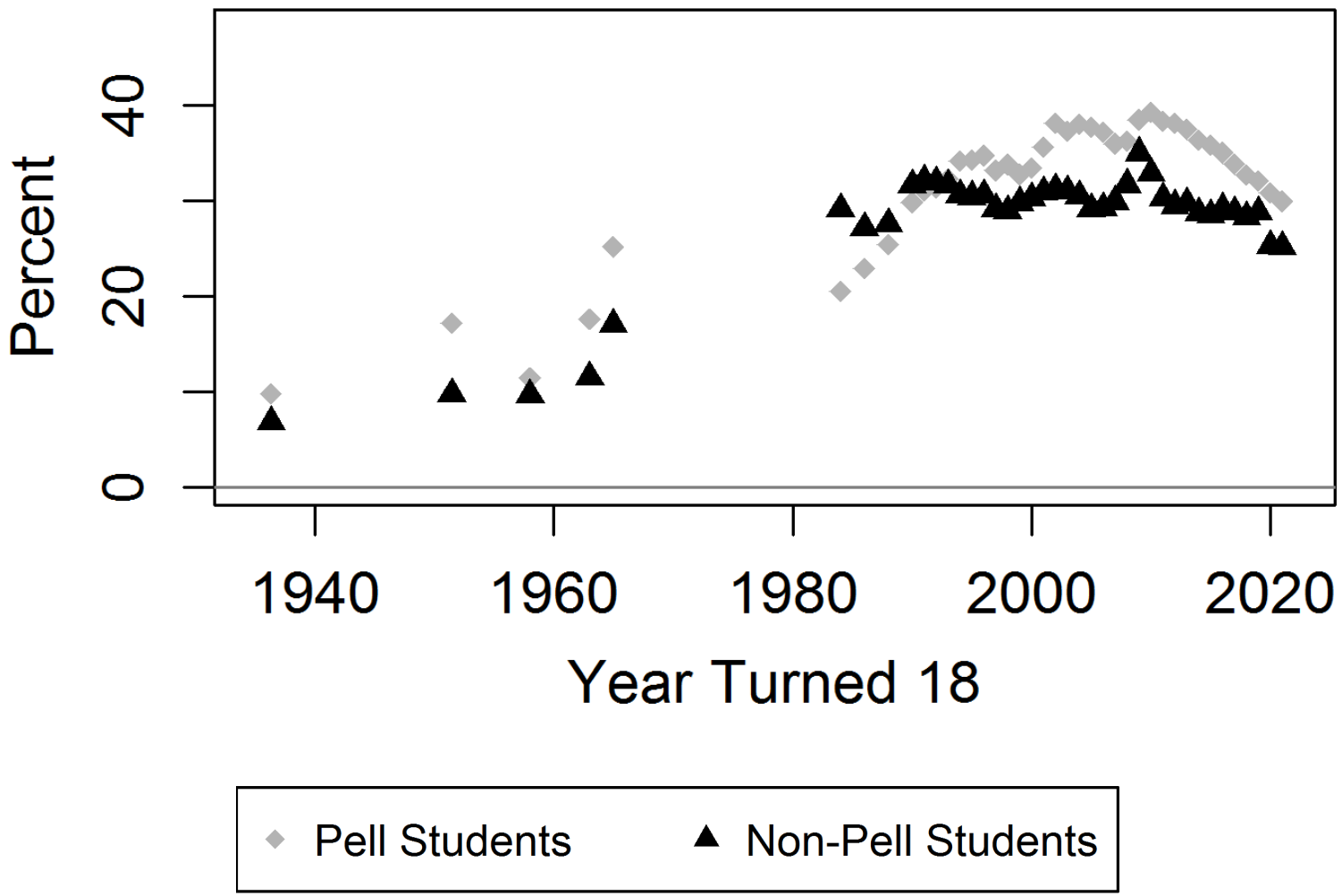
Institutional Enrollment by Pell Eligibility Indexed by Average Wage or Value-Added



Back

Explanation III: Institutions

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



Back