Human Capital*

Christopher Campos

Economics of Education, Spring 2026

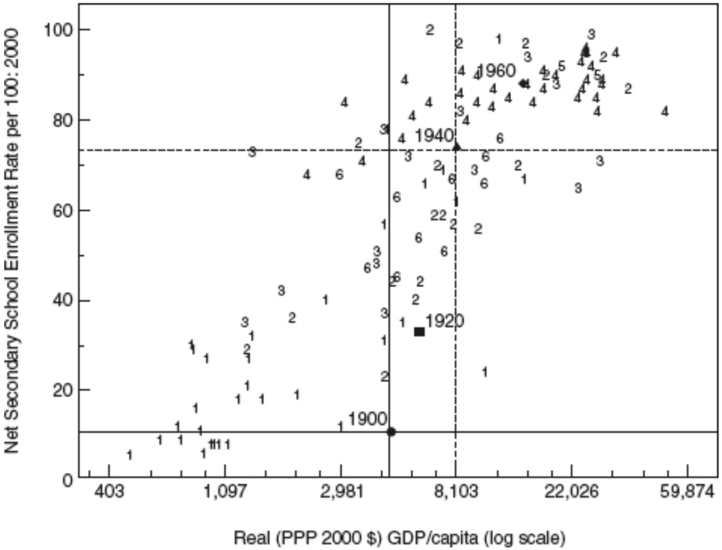
* This slide deck was partly influenced by Chris Walters' labor economics lecture slides.

Motivation

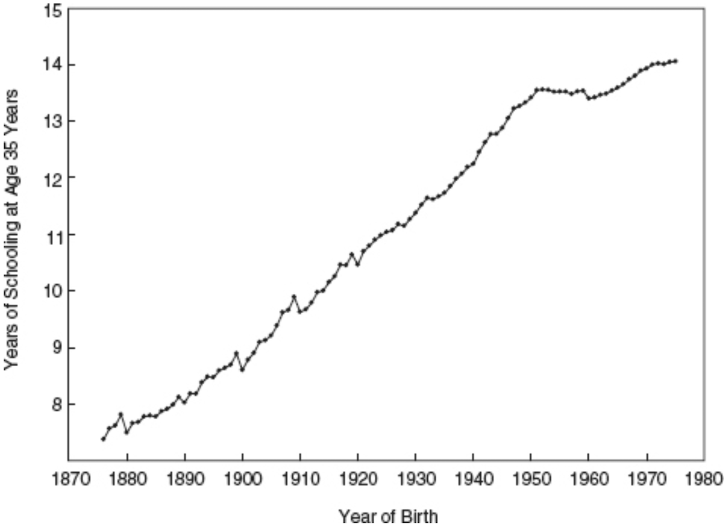
Outline

1. Human Capital Theory
   - → Human capital as an investment
   - → Mincerian earnings function
   - → Ben-Porath model
2. Returns to Schooling
   - → Years of Schooling: Card 2001 and Angrist and Krueger
   - → Marginal students: Zimmerman 2014
   - → College Selectivity: Dale and Krueger
   - → Major Choice: Kirkeboen et al. and Campos et al.
3. Goldin and Katz supply and demand framework for understanding the evolution of returns to schooling

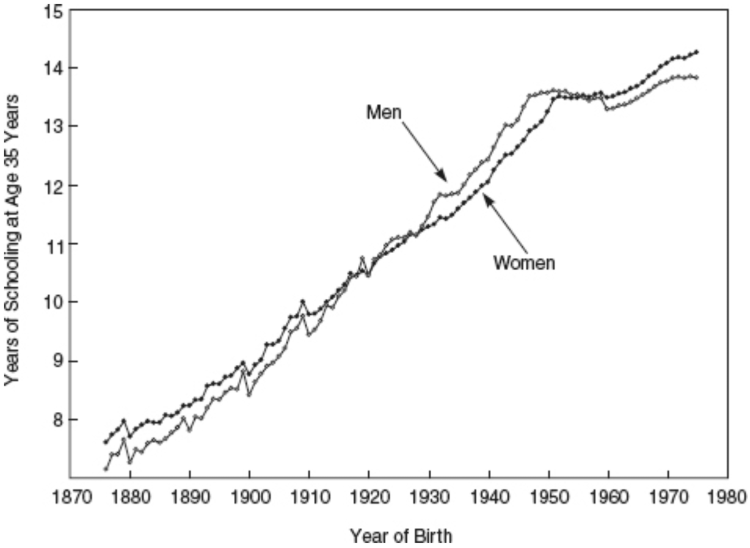
Goldin and Katz: The Human Capital Century



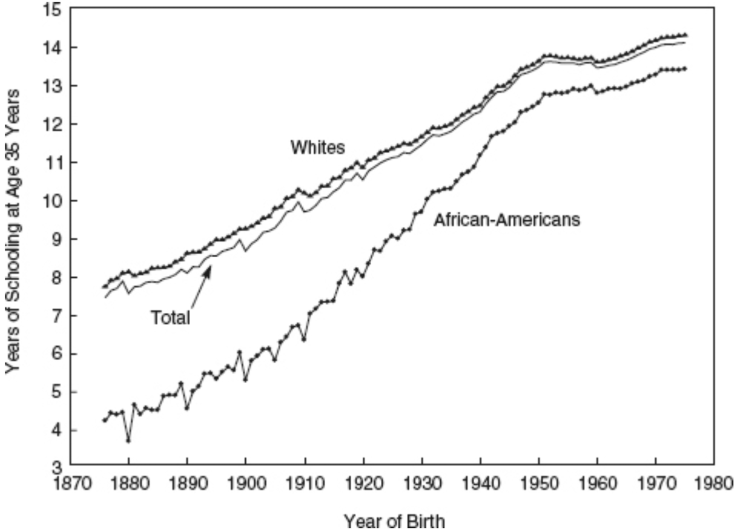
Attainment rose substantially in the United States



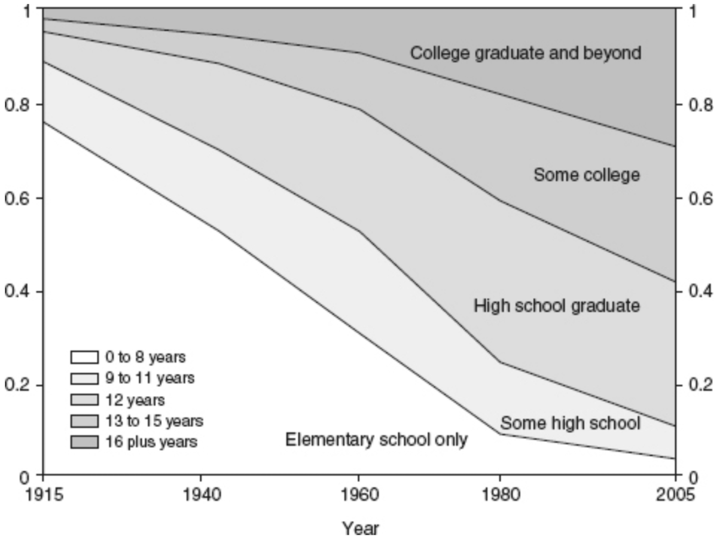
Relative attainment for women has varied across the 20th century



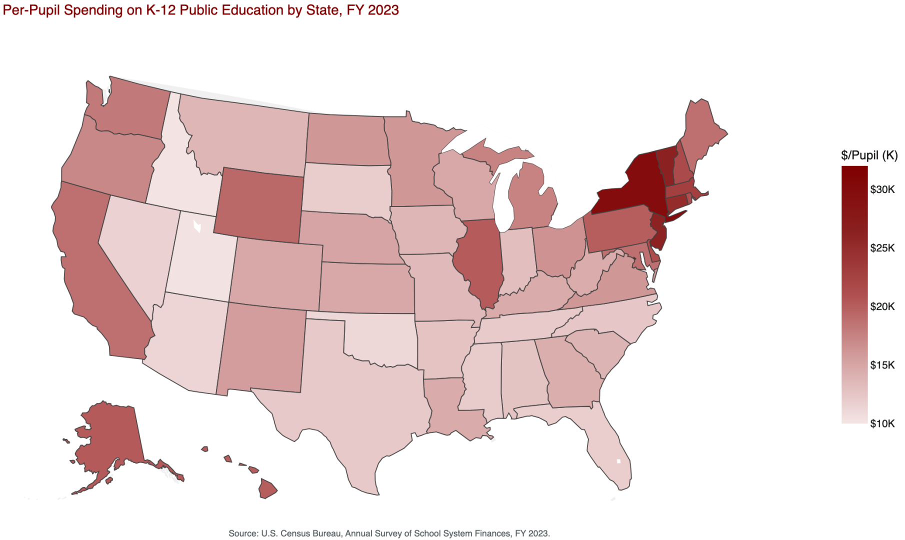
Racial differences in attainment persist but narrow



Higher education takes a more prominent role later in the 20th century



Governments invest a lot in education. Why?



Human Capital

- Important idea in labor economics: marketable skills are a form of *human* capital
- What is human capital?
  - → Roughly speaking, a stock of knowledge, skills, or attributes that contribute to labor market "productivity"
- Productivity has a direct link to earnings/wages, so investments in human capital have a "return"
- How should we think about or measure the return?
- In the first part of lecture we will cover:
  - → Gary Becker proposes an accounting-like framework to explain investment decisions where the *internal rate of return* is key
  - → Jacob Mincer proposes a simple model that relates log earnings to *years of education* and labor market experience
  - → Ben-Porath model predicts age-investment profile that mirrors empirical regularities

What contributes to differences in human capital?

- Innate ability
  - → Luck of the draw ability produces heterogeneity in outcomes, conditional on educational investments
- Years of schooling
  - → Has received the most attention because it is more easily measured
- School Quality
  - → K-12 schooling quality, teacher quality, etc
  - → College selectivity, college major, etc
- Job training
  - → On-the-job training
  - → Reskilling
- Pre-labor market influences
  - → Parental inputs matter a lot

Theory

Intellectual Origins of Human Capital Theory

- **Theodore Schultz (1961):** Popularized the concept in his AEA presidential address. Education and training are not only consumption but *investment* that produce a durable stock of productive capacity. Controversial at the time: equating people with capital struck many as dehumanizing.
- **Gary Becker (1962):** Formalized the investment framework. Key contributions:
  - → Individuals invest up to the point where the marginal return equals the marginal cost of funds/opportunity cost
  - → The distinction between *general* and *specific* human capital determines who pays for and who captures the returns to training
- **Jacob Mincer (1958, 1974):** Operationalized the theory empirically with a functional form linking log earnings to schooling and experience
- Ben-Porath (1967): developed a dynamic lifecycle model of human capital accumulation

Human Capital as an Investment

- Core Tradeoff: An additional year of schooling raises the entire future earnings profile, but delays labor market entry by one year
  - → Marginal benefit: Higher wages in every future period, discounted back to the present
  - → Marginal cost: One year of foregone earnings plus direct costs (tuition, fees, materials)
- The present value of lifetime earnings for an individual choosing s years of schooling:

$$V(s) = \int_{t=s}^{T} e^{-r(t-s)}\, w(t;s)\, dt \; - \; C(s)$$

- Optimal schooling equates the marginal gain in lifetime earnings from one more year to the marginal cost of that year:

$$\frac{dV(s)}{ds} = 0 \quad \Longleftrightarrow \quad \underbrace{\int_{s}^{T} e^{-r(t-s)} \frac{\partial w}{\partial s}\, dt}_{\text{MB: higher future wages}} = \underbrace{w(s;s) + C'(s)}_{\text{MC: foregone earnings + direct costs}}$$

Internal Rate of Return

- Recall the present value of choosing s years of schooling, discounted at rate r:

$$V(s;\, r) = \int_s^T e^{-r(t-s)}\, w(t;s)\, dt \;-\; C(s)$$

- The **internal rate of return** ρ is the discount rate at which an individual is indifferent between s years of schooling and no schooling:

$$V(s;\, \rho) \;=\; V(0;\, \rho)$$

- Decision rule: invest if $\rho > r$
  - → When $\rho > r$, the earnings gains from schooling more than compensate for the costs at market rates $\Rightarrow V(s;\, r) > V(0;\, r)$
- Why might ρ differ across individuals?
  - → **Ability:** Higher-ability individuals may learn more per year of schooling – higher $w(t;s) - w(t;0)$ – raising ρ
  - → **Credit constraints:** Individuals who cannot borrow face a higher effective r, so they underinvest even when ρ is high
  - → **Direct costs:** Subsidies and public schooling lower $C(s)$, raising the NPV at any given r

General vs. Specific Human Capital

Who pays for training depends on whether the skills are transferable:

- General human capital: Skills valued by many employers (literacy, numeracy, programming)
  - → Workers can leave and take the skills with them
  - → Competitive firms will not pay for training they cannot capture ⇒ *workers* bear the cost (via lower wages during training or paying tuition)
  - → Workers also capture the full return via higher post-training wages
- Specific human capital: Skills valuable only at the current firm (proprietary software, internal processes)
  - → Workers cannot take the skills elsewhere, so the outside option is unchanged
  - → Both firm and worker share costs and returns ⇒ wages rise above the outside option but below the worker's marginal product
  - → Creates a *surplus* that makes separation costly for both sides

Jacob Mincer

- Post-World War II Focus on Education: Economists increasingly recognized that formal schooling and on-the-job training could be treated as a form of investment
- Mincer was part of a broader movement at Columbia and the NBER (with Gary Becker and Theodore W. Schultz) that pioneered the concept of *human capital*
- In the 1950s and 1960s, most wage analyses were *descriptive* and lacked a clear theoretical framework linking education, experience, and earnings in a single statistical model
- Mincer developed a semi-log functional form for earnings, making it easier to interpret the *returns* to schooling and the contribution of experience
- See Lemiuex 2003 and Rosen 1992 for more details

Mincer's Earnings Regression

- Mincerian regressions are a reduced-form representation for log earnings in terms of years of schooling (and experience)

$$\ln Y = \alpha + \rho s + \gamma_1 X - \gamma_2 X^2 + e$$

- To arrive at this, Mincer (1958) uses the principle of compensating differences to explain why persons with different schooling levels receive different earnings in the labor market

- Assumptions:
  - → Individuals are identical
  - → Perfect credit markets
  - → Occupations differ in the amount of required schooling
  - → No uncertainty about future earnings

- $Y(s)$ represents annual earnings of an individual with s years of education assumed to be constant over the lifetime
- r is an exogenous interest rate

Mincer's Earnings Regression

- The present value of earnings associated with schooling s is

$$V(s) = \int_s^T Y(s)e^{-rt}dt = \frac{Y(s)}{r}\left(e^{-rs} - e^{-rT}\right)$$

- In equilibrium, individuals are indifferent between schooling choices with allocations driven by demand conditions

- Equating earnings streams across earning levels and taking logs

$$V(s) = V(0) \implies \ln Y(s) = \ln Y(0) + rs + \ln\left(\frac{1 - e^{-rT}}{1 - e^{-r(T-s)}}\right)$$

  with the latter term going to zero as $T \to \infty$

- Model implies that individuals with more education earn more

- When T is large, the percentage increase in lifetime earnings associated with an additional year of schooling must equal r
  - → Because the internal rate of return equates earnings streams across investments, when T is large, r also equals the IRR

- Notice that experience is not part of this equation

Mincer (1974)

- Mincer's second derivation is an accounting argument: track cumulative investments over the lifecycle
- Unlike the 1958 model, this generates the experience-earnings profile from on-the-job investment by assuming a linearly declining investment rate
- Let P_t be *potential* earnings at age t
- Individuals invest in training with cost $C_t = k_t P_t$ (a fraction k_t of potential earnings)
- Let ρ_t be the average return to investments made at age t
- Potential earnings at age t are

$$P_t = P_{t-1}(1 + k_{t-1}\rho_{t-1}) = \prod_{j=0}^{t-1}(1 + \rho_j k_j)P_0$$

- Formal school is defined as years where $k_t = 1$ and assume those investments yield a rate of return ρ_s
- Assume post-schooling investments have rate of return ρ_0
- We can write

$$\ln P_t = \ln P_0 + s\ln(1 + \rho_s) + \sum_{j=s}^{t-1}\ln(1 + \rho_0 k_j)$$

Mincer (1974)

- Assuming ρ_s and ρ_0 are small

$$\ln P_t \approx \ln P_0 + \rho_s s + \rho_0 \sum_{j=s}^{t-1} k_j$$

- Assume a linearly declining rate of post-school investment $\kappa(1 - x/T)$ where $x = t - s \geq 0$ is the amount of work experience as of age t

- We can then write potential earnings in terms of schooling s and experience x

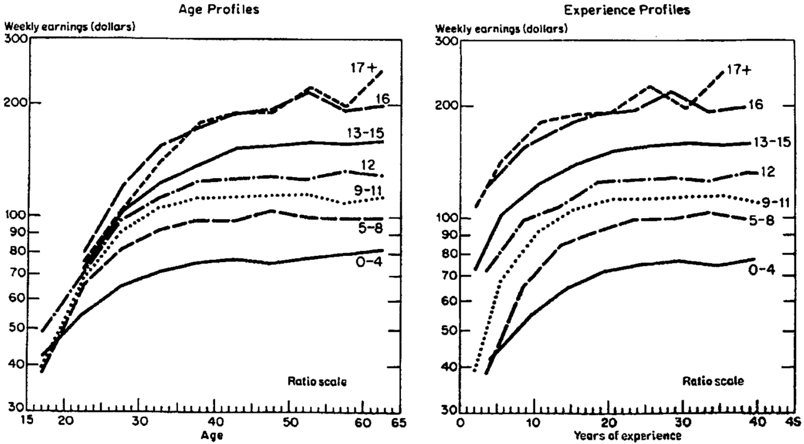
$$\ln P_{x+s} \approx \ln P_0 + \rho_s s + \left(\rho_0 \kappa + \frac{\rho\kappa}{2T}\right) x - \frac{\rho_0 \kappa}{2T} x^2$$

- Observed earnings are potential earnings net of investment costs, producing the Mincerian relationship between earnings, schooling, and experience

$$\begin{aligned} \ln Y(s,x) &= \ln P_{x+s} - \kappa\left(1 - \frac{x}{T}\right) \\ &= \underbrace{\ln P_0 - \kappa}_{\alpha} + \rho_s s + \underbrace{\left(\rho_0 \kappa + \frac{\rho_0 \kappa}{2T} - \frac{\kappa}{T}\right)}_{\gamma_1} x - \underbrace{\frac{\rho_0 \kappa}{2T}}_{\gamma_2} x^2 \end{aligned}$$

Empirical Evidence: Mincer 1974

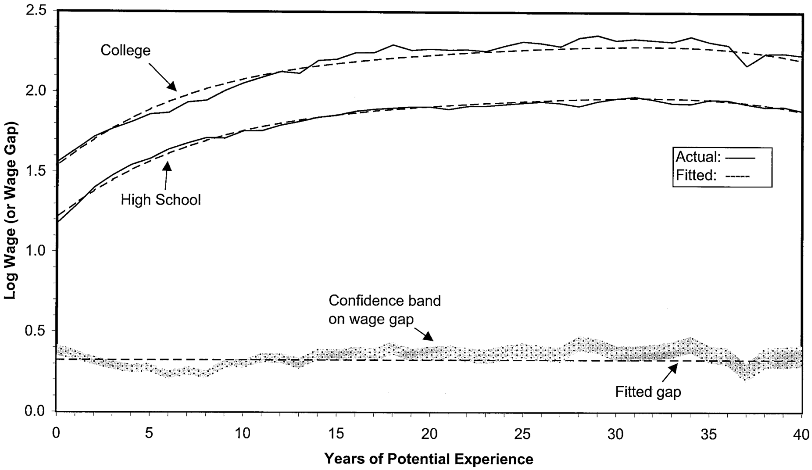
AGE AND EXPERIENCE PROFILES OF RELATIVE WEEKLY EARNINGS OF WHITE, NONFARM MEN, 1959



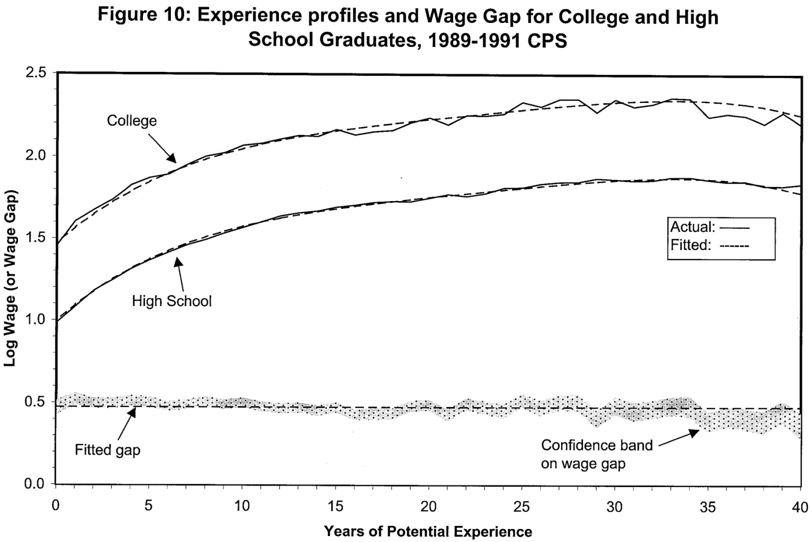
NOTE: Figures on curves indicate years of schooling completed.
SOURCE: 1/1,000 sample of U.S. Census, 1960.

Empirical Evidence: 1979-1981 CPS

Figure 9: Experience profiles and Wage Gap for College and High School Graduates, 1979-1981 CPS

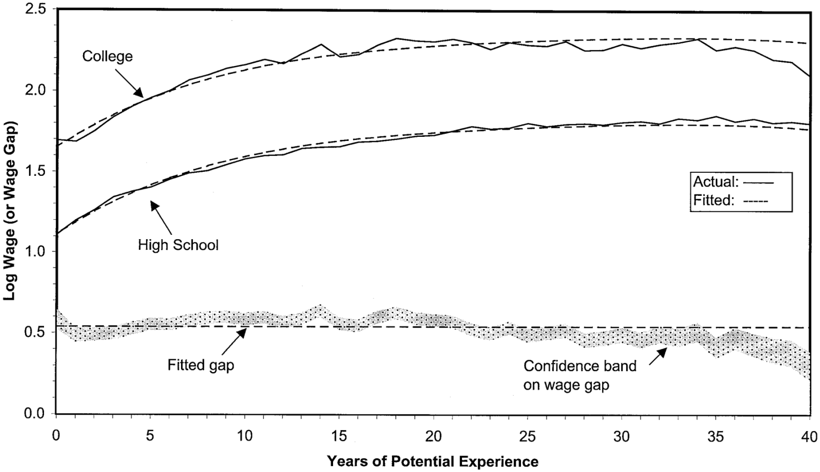


Empirical Evidence: 1989-1991 CPS

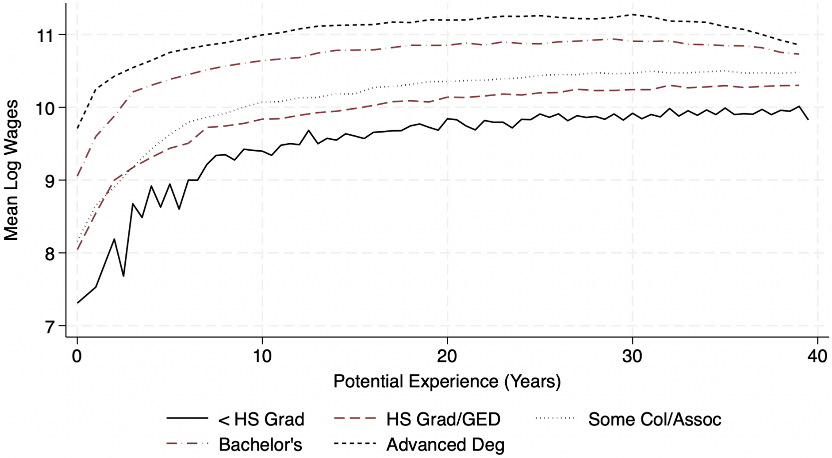


Empirical Evidence: 1999-2001 CPS

Figure 11: Experience profiles and Wage Gap for College and High School Graduates, 1999-2001 CPS



Empirical Evidence: 2019 ACS



Source: 2019 ACS

Takeaways

- The Mincer equation $\ln Y = \alpha + \rho_s s + \gamma_1 x - \gamma_2 x^2$ emerges from two very different arguments:
  1. **1958 (Equilibrium):** Compensating differences across occupations $\Rightarrow$ log-linear schooling premium. *No experience term.*
  2. **1974 (Accounting):** Cumulative human capital investments over the lifecycle $\Rightarrow$ log-linear schooling premium *plus* concave experience profile
- The 1974 model generates the experience profile by *assuming* a linearly declining investment rate $k(x) = \kappa(1 - x/T)$
- Overall, the Mincer model does a remarkably good job explaining lifecycle earnings/wage patterns
  - → Log-Wage difference between high school and college graduates has ranged between 0.45-0.68 log points over the past 45 years
- But why should investment decline with age? The 1974 Mincer model takes this as given

Ben-Porath Model

- **Historical Background:**
  - → Introduced by Yoram Ben-Porath (1967) to formalize life-cycle human capital investment
  - → Builds on Becker's and Mincer's earlier insights about returns to education and training
  - → Ben-Porath (1967): individuals *optimally* reduce investment over the lifecycle because the horizon over which they can recoup returns shrinks
- **Key Question:**
  - → How do individuals optimally allocate time and resources to accumulate human capital?
  - → How does this decision evolve over the working life?
- **Central Trade-Off:**
  - → Investing in human capital (via time/effort) raises future productivity and wages
  - → But it also entails opportunity costs: less time for work (foregone current earnings)

Environment and Timing

- Continuous time $t \in [0, T]$ with fixed end-of-life (or retirement) T
- Human capital stock K_t depreciates at rate δ
- Competitive rental price per unit of human capital services: a_0
- Perfect capital markets: can borrow/lend at constant interest rate r

Earning Capacity vs. Observed Earnings

- **Earning capacity:** what you could earn if you devote all services to market work

$$Y_t = a_0 K_t$$

- But some human capital services are diverted to producing more human capital
- Observed earnings are lower than Y_t when you invest
  - → This is the mechanical source of the early-career earnings dip

Human Capital Production Technology

- Flow output of new human capital:

$$Q_t = \beta_0 (s_t K_t)^{\beta_1} D_t^{\beta_2}, \qquad \beta_1, \beta_2 > 0, \quad \beta_1 + \beta_2 < 1$$

- Inputs:
  - → $s_t \in [0, 1]$: fraction of human-capital services allocated to producing human capital
  - → $D_t \geq 0$: purchased inputs (tuition, materials, etc.), price P_d
- Accumulation with depreciation:

$$\dot{K}_t = Q_t - \delta K_t$$

Cost of Producing Q_t

- Investing requires both:
  - → **Foregone earnings:** divert services away from market work
  - → **Direct costs:** purchase inputs D_t
- Ben-Porath defines investment outlays:

$$I_t = a_0 s_t K_t + P_d D_t$$

- Goal: for a target Q_t, find the *minimum* I_t (a cost function)

The Dynamic Objective

- The individual chooses an entire life-cycle investment path
- Present discounted value of disposable earnings:

$$W_t = \int_t^T e^{-r(v-t)} \left[a_0 K(v) - I(v)\right] \, dv$$

- Why do $K(v)$ and $I(v)$ appear?
  - → $K(v)$ is future earning capacity generated by past investment
  - → $I(v)$ is future investment cost along the chosen path
- So the problem is dynamic: current choices affect future stocks, future costs, and future earnings

Model Roadmap

$$\boxed{K_t \longrightarrow (s_t, D_t) \longrightarrow Q_t,\ I_t,\ \dot{K}_t \longrightarrow \{K(v), I(v)\}_{v\in[t,T]} \longrightarrow W_t}$$

- K_t is the inherited state
- (s_t, D_t) are the controls
- These determine current production Q_t, current cost I_t, and state transition $\dot{K}_t = Q_t - \delta K_t$
- Through the law of motion, current choices determine the entire future path entering W_t

Deriving the Investment Cost Function: Setup

- Let $x_t \equiv s_t K_t$ denote "own" input in efficiency units
- For a given target Q_t, choose inputs to minimize cost:

$$\min_{x_t, D_t} \; a_0 x_t + P_d D_t \quad \text{s.t.} \quad Q_t = \beta_0 x_t^{\beta_1} D_t^{\beta_2}$$

- Bang-for-buck condition (within age t): equalize marginal product per dollar:

$$\frac{\partial Q_t / \partial x_t}{a_0} = \frac{\partial Q_t / \partial D_t}{P_d}$$

- Some algebra yields:

$$\frac{a_0 x_t}{P_d D_t} = \frac{\beta_1}{\beta_2} \qquad \Longrightarrow \qquad a_0 x_t = \frac{\beta_1}{\beta_2} P_d D_t$$

Deriving the Investment Cost Function

- Let $\gamma \equiv \beta_1 + \beta_2$
- Substitute the cost-share rule into the constraint and solve:

$$I_t(Q_t) = \underbrace{\frac{\gamma}{\beta_2} a_0^{\beta_1/\gamma} P_d^{\beta_2/\gamma} \left(\frac{\beta_2}{\beta_1}\right)^{\beta_1/\gamma}}_{\text{constant}} \left(\frac{Q_t}{\beta_0}\right)^{1/\gamma}$$

- Define marginal cost:

$$MC_t(Q_t) \equiv \frac{\partial I_t(Q_t)}{\partial Q_t}$$

- Differentiating $I_t(Q_t) = C\left(\frac{Q_t}{\beta_0}\right)^{1/\gamma}$ gives:

$$MC_t(Q_t) = \frac{C}{\gamma} \beta_0^{-1/\gamma} Q_t^{\frac{1}{\gamma}-1}$$

- Because $\gamma < 1$, the exponent is positive $\Rightarrow MC$ slopes upward

The Value of an Extra Unit of Human Capital

- What is one more unit of K_t worth at age t?
- It yields rental earnings a_0 each instant, but depreciates at rate δ
- Present value of that extra unit:

$$P_t = a_0 \int_t^T e^{-(r+\delta)(v-t)}\, dv = \frac{a_0}{r+\delta}\left(1 - e^{-(r+\delta)(T-t)}\right)$$

- Interpretation:
  - → P_t is the shadow value or demand price of human capital
  - → it declines with age because the remaining payoff horizon shrinks

Solving the Dynamic Problem

- Rewrite the dynamic problem as:

$$\max_{\{Q_v\}_{v\in[t,T]}} \int_t^T e^{-r(v-t)} \left[a_0 K(v) - I(v)\right] dv \qquad \text{s.t.} \qquad \dot{K}_v = Q_v - \delta K_v$$

- We can do this because for any target Q_t, the individual chooses (s_t, D_t) to minimize cost. Therefore, after the within-period problem is solved, investment cost can be written as $I_t = I(Q_t)$ allowing us to rewrite the dynamic problem as above

- A small increase dQ_t has two effects:
  - → Current cost: raises investment outlays by $MC_t\, dQ_t$
  - → Future benefit: raises the stock of human capital, worth $P_t\, dQ_t$

- Therefore

$$dW_t = (P_t - MC_t)\, dQ_t$$

- **At an interior optimum** (phase ii in the paper), $dW_t = 0$, so

$$MC_t(Q_t) = P_t$$

Optimal Investment and the Life-Cycle Pattern

- In the interior phase, optimal production solves

$$MC_t(Q_t) = P_t$$

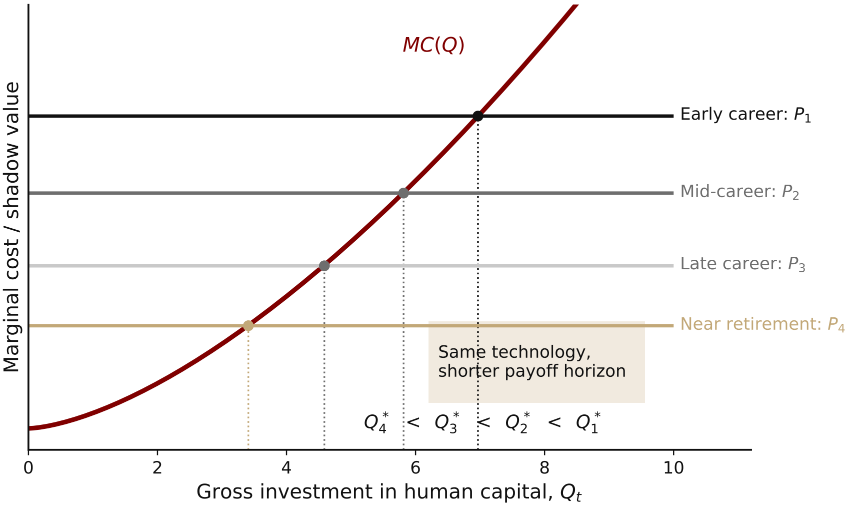
- Intuition:
  - $\rightarrow$ $MC_t(Q_t)$ is pinned down by technology and input prices
  - $\rightarrow$ P_t declines with age as $T - t$ shrinks
  - $\rightarrow$ therefore optimal Q_t declines with age

- This yields the classic Ben-Porath implication:

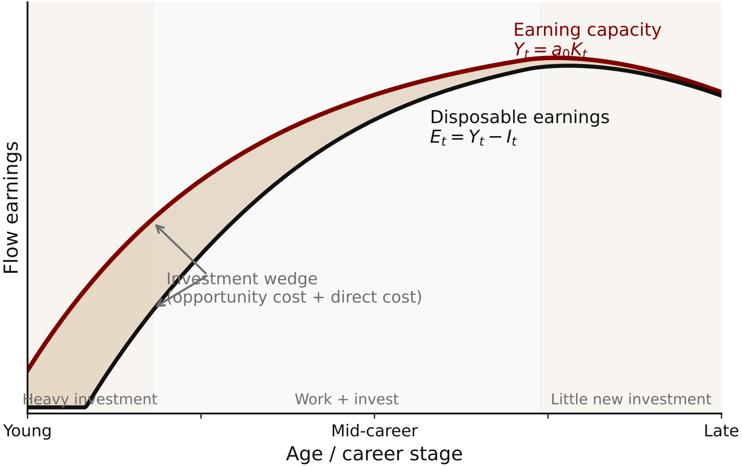
high investment when young $\rightarrow$ declining investment with age

- This is a microfoundation for declining investment rates we saw in Mincer (1974)

Phase (ii): declining shadow value implies declining optimal investment



Earning capacity exceeds disposable earnings while investment is positive

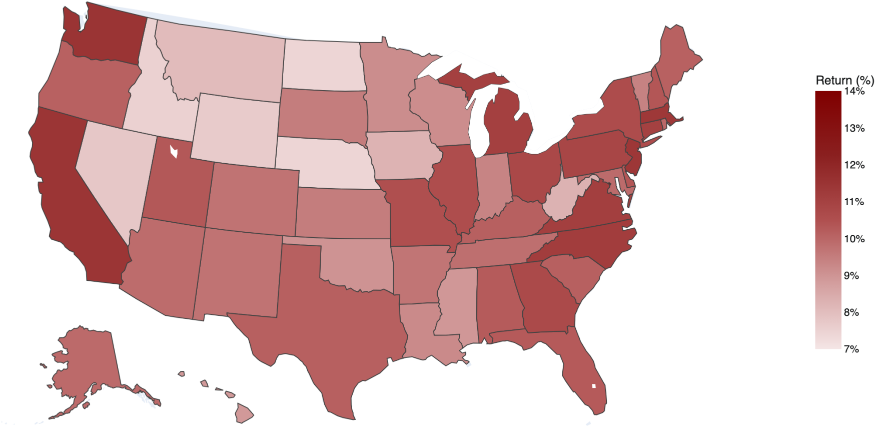


Ben-Porath Model: Implications

- Key predictions:
  - → Human capital investment is most intensive early in life
  - → Investment continues throughout working life but declines over time
  - → Earnings profile is increasing and concave
- These theoretical predictions align well with observed data
- But is the concave experience-earnings profile driven by Ben-Porath investment logic or by passive age/experience effects?

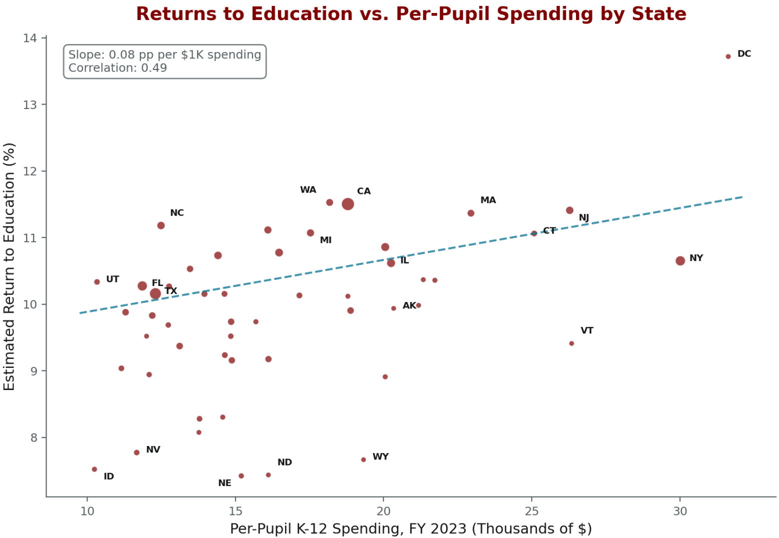
Mincerian returns across states

Estimated Returns to Education by State (Mincer Regression, ACS 2023)



Source: ACS 2023 1-Year PUMS. Sample: full-time employed wage/salary workers aged 25-64. Mincer regression: $\log(wage) = \alpha + \rho \cdot educ + \gamma_1 \cdot exp + \gamma_2 \cdot exp^2$.

Positive association between spending and estimates Mincerian returns. Causal?



Returns to Schooling

Returns to Schooling

- The human capital paradigm motivates empirical studies estimating the “returns” to schooling
- We will start with studies that estimate “returns” to years of schooling
  - → Card 2001 elegantly outlines a framework to interpret the numerous studies
  - → Angrist and Krueger 1999 is a seminal study
  - → Zimmerman 2014 estimates marginal returns to college
- But as we have noted, residual variation in earnings conditional on years of schooling can be explained by a host of factors. Motivates studying
  - → College selectivity: Dale and Krueger 2002, 2014 and Chetty, Deming, and Friedman 2023
  - → Major Choice: Kirkeboen et al. 2016 and Campos et al. 2025

The Card Model

Assume that an individual chooses S to maximize

$$U(S, y) = \log y(S) - h(s)$$

- $h(s)$ is an increasing convex function
- Note that if we assume $U(S, y)$ is the discounted present value (DPV) objective function

$$V(s) = \int_S^\infty y(S)\exp(-rt)dt = y(S)\exp(-rS)/r \implies \log(V(s)) = a + \log y(S) - rS \implies h(s) = rS$$

  → $h(S)$ could be strictly convex if the marginal cost of schooling rises faster than the foregone earnings, either due to tastes or credit constraints

- The marginal rate of substitution between income and schooling is

$$y(S)h'(S)$$

  → In the DPV model, $MRS = ry(S)$ since opportunity costs of Sth year of schooling are just foregone earnings

  → In general, if $h'(S)$ is increasing in S, the MRS rises faster than $y(S)$

Optimal Schooling

- The optimal schooling choice satisfies

$$h'(S) = \frac{y'(S)}{y(S)}$$

- Heterogeneity in schooling decisions due to
  - → $h(S)$: Differences in costs or tastes for schooling
  - → $y'(S)/y(S)$: Differences in marginal return to schooling
- Bias through $y(S)$
  - → Selection on levels due to differences in $y(S)$ across the population
  - → Selection on gains due to differences in $y'(S)$ across the population

Modeling heterogeneity in marginal benefits and costs

- A simple specification of benefits and costs is:

$$\frac{y'(S)}{y(S)} = b_i - k_1 S$$

$$h'(S) = r_i + k_2 S$$

→ (b_i, r_i) are random variables with means $\bar{b}$ and $\bar{r}$
→ k_2 and k_1 are non-negative constants
→ b_i is an individual's idiosyncratic component in the return to schooling
→ r_i represents the heterogeneous marginal cost of schooling

- The simple model implies that the optimal schooling level is

$$S_i^* = \frac{b_i - r_i}{k_1 + k_2} \equiv \frac{b_i - r_i}{k}$$

Heterogeneity in marginal returns to schooling

- Individual i's marginal return to schooling is

$$\beta_i = b_i - k_1 S_i^* = b_i \left(1 - \frac{k_1}{k}\right) + \frac{r_i k_1}{k}$$

- This implies there is a *distribution* of marginal returns unless

  1. $r_i = \bar{r}$ and $k_2 = 0$: no heterogeneity in marginal costs
  2. $b_i = \bar{b}$ and $k_1 = 0$: no heterogeneity in marginal benefits

- This is a partial equilibrium model

  → Marginal returns evolve in response to market conditions

  → More on this later in the lecture

Implications for “returns” to schooling estimates

- The linear marginal cost assumption implies a model for log earnings of the form

$$\log y_i = \alpha_i + b_i S_i - \frac{1}{2} k_1 S_i^2$$

  → A generalization of Mincer (1974) because heterogeneity in the intercept and the slope

- Let’s now think what all this implies for OLS estimates of $\bar{b}$.

$$\log y_i = a_0 + \bar{b} S_i - \frac{1}{2} k_1 S_i^2 + \underbrace{a_i}_{Ability} + \underbrace{(b_i - \bar{b}) S_i}_{Idiosyncratic\ Return}$$

- One can show that

$$b_{OLS} = \underbrace{\bar{b} - k_1 \bar{S}}_{Avg.\ Return} + \underbrace{k \left(\frac{\sigma_{ba} - \sigma_{ra}}{\sigma_b^2 + \sigma_r^2 - 2\sigma_{br}} \right)}_{Ability\ Bias} + \underbrace{k \left(\frac{\sigma_b^2 - \sigma_{br}}{\sigma_b^2 + \sigma_r^2 - 2\sigma_{br}} \right)}_{Gain\ Bias}$$

Implications for “returns” to schooling estimates

$$b_{OLS} = \underbrace{\bar{b} - k_1 \bar{S}}_{Avg.\ Return} + \underbrace{k\left(\frac{\sigma_{ba} - \sigma_{ra}}{\sigma_b^2 + \sigma_r^2 - 2\sigma_{br}}\right)}_{\lambda_0} + \underbrace{k\left(\frac{\sigma_b^2 - \sigma_{br}}{\sigma_b^2 + \sigma_r^2 - 2\sigma_{br}}\right)}_{\psi_0} \bar{S}$$

- In general, the bias in an OLS regression depends on the correlation between idiosyncratic benefits, ability, and idiosyncratic marginal costs
- No heterogeneity in b_i and linear returns to S implies $b_{OLS} = b + \lambda_0$ where bias is due to correlation between marginal cost of schooling r_i and ability a_i
- Typical estimates of b_{OLS} are 7-8 percent, with higher estimates found in IV and twin studies. Why?

Angrist and Krueger 1991

- Angrist and Krueger (1991) seek to estimate the causal return to education
- Instrumental variables strategy motivated by interaction between compulsory schooling and age-at-entry laws
  - → Students can typically drop out on the day they turn 16
  - → Birth date cutoff for starting age: Usually start in the fall of the calendar year in which they turn six
- Generates differences in mean attainment by date of birth
- Viewed from today's vantage point, who are the compliers here?

Example

Birth date	School start date	Earliest dropout date	Completed schooling
December 31, 1929	September 1, 1936	December 31, 1945	9.33 years
January 2, 1930	September 1, 1937	January 2, 1946	8.33 years

Quarter of Birth Predicts Years of Schooling

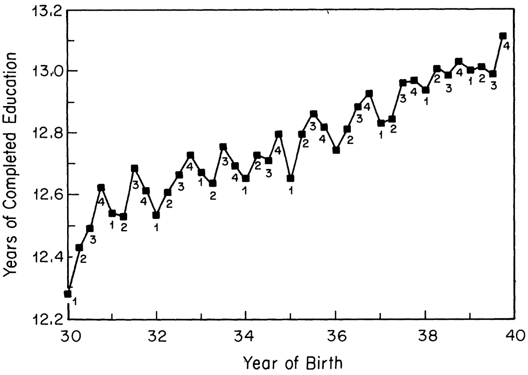


FIGURE I
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.

Quarter of Birth Predicts Lower Levels of Schooling

TABLE I
THE EFFECT OF QUARTER OF BIRTH ON VARIOUS EDUCATIONAL OUTCOME VARIABLES

Outcome variable	Birth cohort	Mean	Quarter-of-birth effect[a] I	II	III	F-test[b] [P-value]
Total years of education	1930–1939	12.79	−0.124	−0.086	−0.015	24.9
			(0.017)	(0.017)	(0.016)	[0.0001]
	1940–1949	13.56	−0.085	−0.035	−0.017	18.6
			(0.012)	(0.012)	(0.011)	[0.0001]
High school graduate	1930–1939	0.77	−0.019	−0.020	−0.004	46.4
			(0.002)	(0.002)	(0.002)	[0.0001]
	1940–1949	0.86	−0.015	−0.012	−0.002	54.4
			(0.001)	(0.001)	(0.001)	[0.0001]
Years of educ. for high school graduates	1930–1939	13.99	−0.004	0.051	0.012	5.9
			(0.014)	(0.014)	(0.014)	[0.0006]
	1940–1949	14.28	0.005	0.043	−0.003	7.8
			(0.011)	(0.011)	(0.010)	[0.0017]
College graduate	1930–1939	0.24	−0.005	0.003	0.002	5.0
			(0.002)	(0.002)	(0.002)	[0.0021]
	1940–1949	0.30	−0.003	0.004	0.000	5.0
			(0.002)	(0.002)	(0.002)	[0.0018]
Completed master's degree	1930–1939	0.09	−0.001	0.002	−0.001	1.7
			(0.001)	(0.001)	(0.001)	[0.1599]
	1940–1949	0.11	0.000	0.004	0.001	3.9
			(0.001)	(0.001)	(0.001)	[0.0091]
Completed doctoral degree	1930–1939	0.03	0.002	0.003	0.000	2.9
			(0.001)	(0.001)	(0.001)	[0.0332]
	1940–1949	0.04	−0.002	0.001	−0.001	4.3
			(0.001)	(0.001)	(0.001)	[0.0050]

Quarter of Birth Predicts Earnings

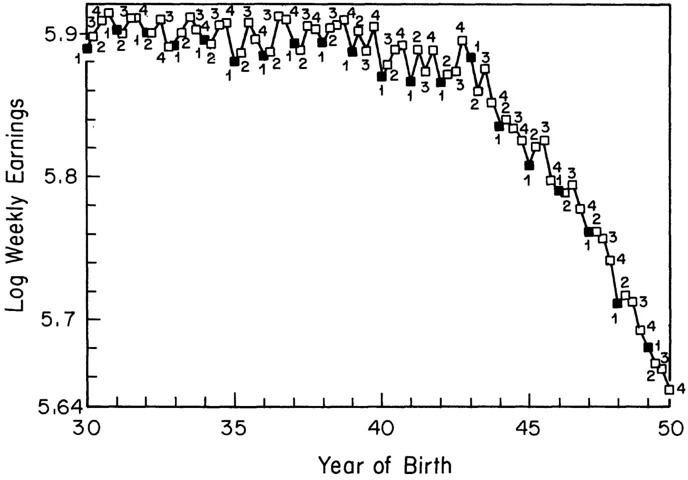


FIGURE V
Mean Log Weekly Wage, by Quarter of Birth
All Men Born 1930–1949; 1980 Census

OLS and TSLS Produce Similar Results

TABLE V

OLS AND TSLS ESTIMATES OF THE RETURN TO EDUCATION FOR MEN BORN 1930–1939: 1980 CENSUS[a]

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0711	0.0891	0.0711	0.0760	0.0632	0.0806	0.0632	0.0600
	(0.0003)	(0.0161)	(0.0003)	(0.0290)	(0.0003)	(0.0164)	(0.0003)	(0.0299)
Race (1 = black)	—	—	—	—	−0.2575	−0.2302	−0.2575	−0.2626
					(0.0040)	(0.0261)	(0.0040)	(0.0458)
SMSA (1 = center city)	—	—	—	—	0.1763	0.1581	0.1763	0.1797
					(0.0029)	(0.0174)	(0.0029)	(0.0305)
Married (1 = married)	—	—	—	—	0.2479	0.2440	0.2479	0.2486
					(0.0032)	(0.0049)	(0.0032)	(0.0073)
9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8 Region-of-residence dummies	No	No	No	No	Yes	Yes	Yes	Yes
Age	—	—	−0.0772	−0.0801	—	—	−0.0760	−0.0741
			(0.0621)	(0.0645)			(0.0604)	(0.0626)
Age-squared	—	—	0.0008	0.0008	—	—	0.0008	0.0007
			(0.0007)	(0.0007)			(0.0007)	(0.0007)
χ^2 [dof]	—	25.4 [29]	—	23.1 [27]	—	22.5 [29]	—	19.6 [27]

a. Standard errors are in parentheses. Sample size is 329,509. Instruments are a full set of quarter-of-birth times year-of-birth interactions. The sample consists of males born in the United States. The sample is drawn from the 5 percent sample of the 1980 Census. The dependent variable is the log of weekly earnings. Age and age-squared are measured in quarters of years. Each equation also includes an intercept.

OLS and TSLS estimates across a range of studies

TABLE II

OLS and IV Estimates of the Return to Education with Instruments Based on Features of the School System

Author	Sample and Instrument		Schooling Coefficients OLS	IV
1. Angrist and Krueger (1991)	1970 and 1980 Census Data, Men. Instruments are quarter of birth interacted with year of birth. Controls include quadratic in age and indicators for race, marital status, urban residence.	1920–29 cohort in 1970	0.070 (0.000)	0.101 (0.033)
		1930–39 cohort in 1980	0.063 (0.000)	0.060 (0.030)
		1940–49 cohort in 1980	0.052 (0.000)	0.078 (0.030)
2. Staiger and Stock (1997)	1980 Census, Men. Instruments are quarter of birth interacted with state and year of birth. Controls are same as in Angrist and Krueger, plus indicators for state of birth. LIML estimates.	1930–39 cohort in 1980	0.063 (0.000)	0.098 (0.015)
		1940–49 cohort in 1980	0.052 (0.000)	0.088 (0.018)
3. Kane and Rouse (1993)	NLS Class of 1972, Women. Instruments are tuition at 2 and 4-year state colleges and distance to nearest college. Controls include race, part-time status, experience. Note: Schooling measured in units of college credit equivalents.	Models without test score or parental education	0.080 (0.005)	0.091 (0.033)
		Models with test scores and parental education	0.063 (0.005)	0.094 (0.042)
4. Card (1995b)	NLS Young Men (1966 Cohort) Instrument is an indicator for a nearby 4-year college in 1966, or the interaction of this with parental education. Controls include race, experience (treated as endogenous), region, and parental education	Models that use college proximity as instrument (1976 earnings)	0.073 (0.004)	0.132 (0.049)
		Models that use college proximity × family background as instrument	—	0.097 (0.048)

OLS and TSLS estimates across a range of studies

5. Conneely and Uusitalo (1997)	Finnish men who served in the army in 1982, and were working full time in civilian jobs in 1994. Administrative earnings and education data. Instrument is living in university town in 1980. Controls include quadratic in experience and parental education and earnings.	Models that exclude parental education and earnings	0.085 (0.001)	0.110 (0.024)
		Models that include parental education and earnings	0.083 (0.001)	0.098 (0.035)
6. Harmon and Walker (1995)	British Family Expenditure Survey 1978–86 (men). Instruments are indicators for changes in the minimum school leaving age in 1947 and 1973. Controls include quadratic in age, survey year, and region.		0.061 (0.001)	0.153 (0.015)
7. Ichino and Winter-Ebmer (1998)	Austria: 1983 Census, men born before 1946. Germany: 1986 GSOEP for adult men. Instrument is indicator for 1930–35 cohort. (Second German IV also uses dummy for father's veteran status). Controls include age, unemployment rate at age 14, and father's education (Germany only). Education measure is dummy for high school or more.	Austrian Men	0.518 (0.015)	0.947 (0.343)
		German Men	0.289 (0.031)	0.590/0.708 (0.844) (0.279)
8. Lemieux and Card (1998)	Canadian Census, 1971 and 1981: French-speaking men in Quebec and English-speaking in Ontario. Instrument is dummy for Ontario men age 19–22 in 1946. Controls include full set of experience dummies and Quebec-specific cubic experience profile.	1971 Census:	0.070 (0.002)	0.164 (0.053)
		1981 Census:	0.062 (0.001)	0.076 (0.022)
9. Meghir and Palme (1999)	Swedish Level of Living Survey (SLLS) data for men born 1945–55, with earnings in 1991, and Individual Statistics (IS) sample of men born in 1948 and 1953, with earnings in 1993. Instrument is dummy for attending "reformed" school system at age 13. Other controls include cohort, father's education, and county dummies. Models for IS data also include test scores at age 13.	SLLS Data (Years of education)	0.028 (0.007)	0.036 (0.021)
		IS Data (Dummy for 1–2 years of college relative to minimum schooling)	0.222 (0.020)	0.245 (0.082)

Zimmerman (2014): Returns for Marginal Students

- Much of the literature on returns to schooling focuses on students on the margin of dropout in mid-20th century
- Perhaps more policy-relevant: Return for current marginal college students
- Zimmerman (2014) uses a combined SAT/GPA cutoff to estimate earnings effects at Florida International University
- Results suggest earnings on the order of 11 percent per year of attendance

First Stage Estimates

Table 4: Effects on academic outcomes

Dep. Var.	Main	Controls	BW=0.5	BW=0.15	Loc. Lin.
A. Admissions and attendance					
Admitted to FIU	0.234***	0.233***	0.246***	0.282***	0.205***
	(0.021)	(0.018)	(0.022)	(0.023)	(0.016)
Attend FIU	0.104***	0.105***	0.112***	0.0980**	0.088**
	(0.025)	(0.026)	(0.029)	(0.040)	(0.027)
Attend SUS	0.119***	0.118***	0.126***	0.125**	0.104***
	(0.021)	(0.023)	(0.025)	(0.037)	(0.023)
Years SUS	0.457**	0.463**	0.492**	0.495**	0.420*
	(0.089)	(0.094)	(0.097)	(0.114)	(0.103)
SUS FTE terms	0.644*	0.643*	0.698*	0.650*	0.622
	(0.179)	(0.192)	(0.190)	(0.185)	(0.207)
B. SUS Graduation					
Within 4 years	-0.007	-0.008	-0.008	-0.009	-0.005
	(0.007)	(0.007)	(0.007)	(0.009)	(0.008)
Within 5 years	0.002	0.001	0.008	-0.002	0.007
	(0.018)	(0.019)	(0.018)	(0.021)	(0.021)
Within 6 years	0.057	0.057	0.056	0.044	0.069
	(0.022)	(0.022)	(0.026)	(0.022)	(0.024)
C. Other academic outcomes					
Years CC	-0.172*	-0.171*	-0.222**	-0.199**	-0.164*
	(0.053)	(0.051)	(0.067)	(0.055)	(0.061)
CC FTE terms	-0.338***	-0.327**	-0.394**	-0.412**	-0.300**

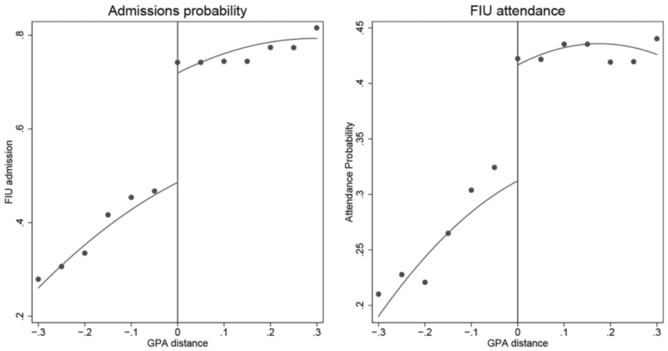
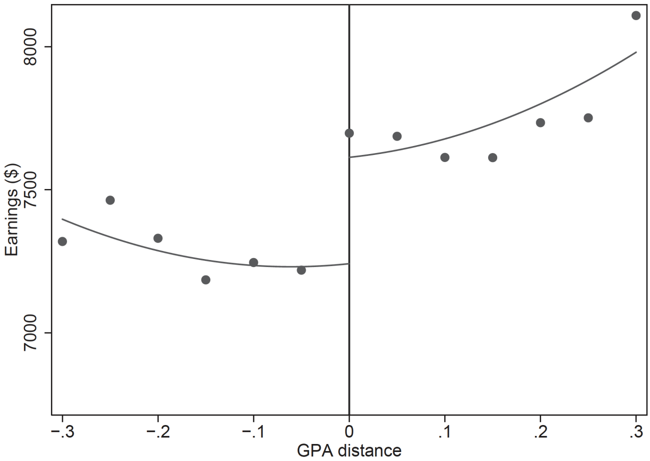


Fig. 4.—Admissions and FIU attendance. Lines are fitted values based on the main specification. Dots, shown every .05 grade points, are rolling averages of values within .05 grade points on either side that have the same value of the threshold-crossing dummy.

Earnings Effects

Figure 8: Quarterly earnings by distance from GPA cutoff



RF and TSLS

Table 5: Earnings effects 8 to 14 years after high school completion

	Main	Controls	BW=0.5	BW=0.15	Loc. Lin.
Reduced form estimates					
Above cutoff	372*	366**	409**	479**	410**
	(141)	(130)	(154)	(198)	(147)
Instrumental variables estimates					
FIU admission	1593*	1575**	1665**	1700**	2001*
	(604)	(584)	(645)	(621)	(696)
Years of SUS attendance	815**	792**	833**	966***	977**
	(276)	(262)	(271)	(305)	(306)
BA degree	6547*	6442*	7366*	10769	5958**
	(2496)	(2411)	(2998)	(5726)	(2024)
N	6542	6542	9659	3294	6542

Schooling Quality

- So far, we have not said much about heterogeneity in schooling quality
  - → Heterogeneity in returns were idiosyncratic but not necessarily due to different options
- Zimmerman findings speak to students who are on the margin of college vs no college
- Many college students are not on that margin. Instead, deciding between
  - → Different colleges
  - → Different majors
- What are the causal impacts of college quality and major quality?
  - → Dale and Krueger: College selectivity
  - → Kirkeboen et al. and Campos et al.: College major

Overview: Dale and Krueger (QJE 2002, JHR 2014)

- Dale and Krueger investigate the returns to attending more selective colleges
- Research design: compare students who applied to, were admitted by, but chose to attend different colleges
- Key idea: Application choices reflect student self-assessment, while admissions reflect school assessments
- For example, a student accepted by Harvard but choosing BU may provide a better counterfactual for Harvard students than a random BU student
- Core approach: compare outcomes of students making different enrollment choices from the same set of offers

Notation and Setup

- Let $j \in \{1, \ldots, J\}$ index colleges and i index students
- C_{ij} indicates student i applied to college j, A_{ij} indicates they were admitted
- Collect application and admission vectors:

$$C_i = (C_{i1}, \ldots, C_{iJ}), \quad A_i = (A_{i1}, \ldots, A_{iJ})$$

- $D_i \in \{1, \ldots, J\}$ is the college attended
- Let Y_{ij} be the potential earnings if i attends j. Observed earnings:

$$Y_i = \sum_j \mathbf{1}\{D_i = j\} Y_{ij}$$

- DK's first identification assumption:

$$(Y_{i1}, \ldots, Y_{iJ}) \perp\!\!\!\perp D_i \mid (C_i, A_i)$$

Interpretation of Assumption

- College choice is conditionally ignorable given (C_i, A_i)
- Implies conditional independence of potential outcomes and choices:

$$\mathbb{E}[Y_i|D_i = j, C_i = c, A_i = a] - \mathbb{E}[Y_i|D_i = k, C_i = c, A_i = a] = \mathbb{E}[Y_{ij} - Y_{ik}|C_i = c, A_i = a]$$

- Within application/admission portfolios, differences in observed outcomes yield average treatment effects
- Limitation: most students won't have overlapping portfolios. DK compare average SAT score levels across groups
- DK argue (C_i, A_i) summarizes key unobservables like private info about student ability or school reputation

Second Approach: Self-Revelation via Portfolio Quality

- Alternative strategy: control for the average SAT score of schools to which a student applied
- This treats portfolio strength as a sufficient statistic for selectivity and ability
- Referred to by DK as a *self-revelation* model: assumes all selection is captured by application behavior
- No adjustment for information from admission or final choice
- Stronger assumption than the first, but simplifies estimation and improves precision
- Still a selection-on-observables strategy

Data: College and Beyond (C&B)

- Data source: College and Beyond (C&B) survey
  - → Combines administrative records and surveys for students from 34 colleges (1951, 1976, 1989)
  - → Institutions are more selective than average
  - → Includes SAT scores, application and admission data, transcripts
- DK (2002): use self-reported earnings in 1995 for 1976 cohort (mid/late 30s)
- DK (2014): match C&B with SSA earnings through 2007:
  - → Enables longer-run earnings analysis
  - → Higher data quality
  - → Also includes 1989 cohort

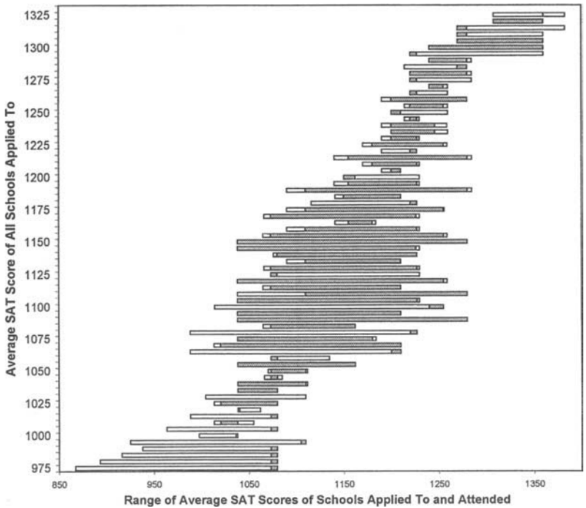


FIGURE I

Range of Schools Applied to and Attended by Most Common Sets of Matched Applicants

Each bar represents the range of the average SAT scores of the schools that a given set of matched applicants applied to; the shaded area represents the range of schools that students in each set attended. Only matched sets that represent

TABLE III
LOG EARNINGS REGRESSIONS USING COLLEGE AND BEYOND SURVEY, SAMPLE OF MALE AND FEMALE FULL-TIME WORKERS

	Model					
	Basic model: no selection controls		Matched-applicant model	Alternative matched-applicant models		
	Full sample	Restricted sample	Similar school-SAT matches*	Exact school-SAT matches**	*Barron's* matches***	Self-revelation model
Variable	1	2	3	4	5	6
School-average SAT score/100	0.076	0.082	−0.016	−0.106	0.004	−0.001
	(0.016)	(0.014)	(0.022)	(0.036)	(0.016)	(0.018)
Predicted log(parental income)	0.187	0.190	0.163	0.232	0.154	0.161
	(0.024)	(0.033)	(0.033)	(0.079)	(0.028)	(0.025)
Own SAT score/100	0.018	0.006	−0.011	0.003	−0.005	0.009
	(0.006)	(0.007)	(0.007)	(0.014)	(0.005)	(0.006)
Female	−0.403	−0.410	−0.395	−0.476	−0.400	−0.396
	(0.015)	(0.018)	(0.024)	(0.049)	(0.017)	(0.014)
Black	−0.023	−0.026	−0.057	−0.028	−0.057	−0.034
	(0.035)	(0.053)	(0.053)	(0.049)	(0.039)	(0.035)
Hispanic	0.015	0.070	0.020	−0.248	0.036	0.007
	(0.052)	(0.076)	(0.099)	(0.206)	(0.066)	(0.053)
Asian	0.173	0.245	0.241	0.368	0.163	0.155
	(0.036)	(0.054)	(0.064)	(0.141)	(0.049)	(0.037)
Other/missing race	−0.188	−0.048	0.060	−0.072	−0.050	−0.192
	(0.119)	(0.143)	(0.180)	(0.083)	(0.134)	(0.116)
High school top 10 percent	0.061	0.091	0.079	0.091	0.079	0.063
	(0.018)	(0.022)	(0.026)	(0.032)	(0.024)	(0.019)
High school rank missing	0.001	0.040	0.016	0.029	0.025	−0.009
	(0.024)	(0.026)	(0.038)	(0.066)	(0.027)	(0.022)
Athlete	0.102	0.088	0.104	0.169	0.093	0.094
	(0.025)	(0.030)	(0.039)	(0.096)	(0.033)	(0.024)
Average SAT score/100 of schools applied to						0.090
						(0.013)
One additional application						0.064
						(0.011)
Two additional applications						0.074
						(0.022)
Three additional applications						0.112
						(0.028)
Four additional applications						0.085
						(0.027)
Adjusted R^2	0.107	0.110	0.112	0.142	0.106	0.113
N	14,238	6,335	6,335	2,330	9,202	14,238

TABLE IV
The Effect of School-Average SAT Score on Earnings in Models that Use Alternative Selection Controls, C&B Sample of Male and Female Full-Time Workers

Type of selection control	Parameter estimates: School-average SAT score	Parameter estimates: Selection control	N
(1) None (basic model)	0.076 (0.016)	—	14,238
(2) Average SAT score/100 of schools applied to (self-revelation model)	−0.001 (0.018)	0.090 (0.013)	14,238
(3) Average SAT score/100 of schools accepted by	−0.001 (0.021)	0.084 (0.017)	14,238
(4) Highest SAT score/100 of schools accepted by	−0.007 (0.018)	0.091 (0.021)	14,238
(5) Highest SAT score/100 of all schools applied to	0.010 (0.015)	0.075 (0.013)	14,238
(6) Highest SAT score/100 of schools applied to but not attended	0.042 (0.013)	0.051 (0.006)	9,358
(7) Average SAT score/100 of schools rejected by	0.052 (0.015)	0.072 (0.012)	3,805
(8) Highest SAT score/100 of schools accepted by not attended	0.039 (0.014)	0.049 (0.010)	8,257

TABLE V

LINEAR REGRESSIONS PREDICTING WHETHER STUDENT ATTENDED MOST SELECTIVE COLLEGE FOR C&B SAMPLE OF STUDENTS ADMITTED TO MORE THAN ONE SCHOOL

	Parameter estimates	
	Matched-applicant model*	Self-revelation model
Predicted log (parental income)	−0.024	−0.037
	(0.026)	(0.030)
Own SAT score/100	0.020	0.021
	(0.005)	(0.007)
Female	0.034	0.033
	(0.014)	(0.028)
Black	0.056	−0.005
	(0.026)	(0.037)
Hispanic	−0.019	0.042
	(0.064)	(0.074)
Asian	0.019	0.074
	(0.026)	(0.050)
Other/missing race	−0.095	0.010
	(0.093)	(0.081)
High school top 10 percent	−0.014	−0.020
	(0.021)	(0.028)
High school rank missing	−0.035	−0.040
	(0.036)	(0.058)
Athlete	0.056	0.059
	(0.023)	(0.045)
Average SAT score/100 of schools applied to		−0.122
		(0.040)
One additional application		0.149
		(0.037)
Two additional applications		0.076
		(0.033)
Three additional applications		0.020
		(0.038)
N	5536	8257

Table 3

Comparing Parameter Estimates of the Effect of College Average SAT Score on Earnings Using C&B and SSA Data, 1976 Cohort

	C&B sample[a]		Merged C&B and SSA sample[b]									
	Log 1995 C&B earnings		Log 1995 C&B earnings		Log 1995 SSA earnings (topcoded)		Log 1995 SSA earnings (not topcoded)		Log (median of 1993 to 1997 earnings), SSA data		Log (median of 1993 to 1997 earnings), SSA data	
	1	2	3	4	5	6	7	8	9	10	11	12
	Basic	Self – revelation	Basic	Self - revelation	Basic	Self – revelation	Basic	Self - revelation	Basic	Self – revelation	Basic	Self – revelation
Parameter estimate for school SAT/100	0.076	-0.001	0.068	-0.007	0.048	-0.021	0.058	-0.015	0.059	-0.025	0.061	-0.023
	(.008)	(.012)	(.007)	(.012)	(.009)	(.014)	(.009)	(.015)	(.008)	(.012)	(.007)	(.012)
	{.016}	{.018}	{.014}	{.018}	{.016}	{.018}	{.017}	{.016}	{.012}	{.013}	{.013}	{.014}
N	14,238		10,886		10,886		10,886		11,932		12,075	
Sample restriction	Full-time workers (according to C&B survey)		Full-time workers (according to C&B survey)		Full-time workers (according to C&B survey)		Full-time workers (according to C&B survey)		Median earnings greater than zero (SSA data)		Median earnings greater than $13,822 in 2007 dollars (SSA data)	

Table 6

Effect of College Characteristics on Earnings, 1976 Cohort of Men and Women

	College characteristic: Log net tuition		College characteristic: Barron's index	
	Basic	Self-revelation	Basic	Self-revelation
Effect on log (median of 1983 through 1987 annual earnings)				
Parameter estimate for	0.014	-0.007	0.010	0.001
school quality measure	(.010)	(.013)	(.005)	(.013)
N = 11,984	{.024}	{.027}	{.012}	{.015}
Effect on log (median of 1988 through 1992 annual earnings)				
Parameter estimate for	0.092	0.012	0.055	0.020
school quality measure	(.012)	(.016)	(.006)	(.017)
N = 12,407	{.028}	{.028}	{.011}	{.015}
Effect on log (median of 1993 through 1997 annual earnings)				
Parameter estimate for	0.124	0.013	0.071	0.017
school quality measure	(.015)	(.019)	(.007)	(.010)
N = 12,075	{.030}	{.038}	{.009}	{.015}
Effect on log (median of 1998 through 2002 annual earnings)				
Parameter estimate for	0.140	0.017	0.077	0.014
school quality measure	(.012)	(.017)	(.008)	(.012)
N = 12,064	{.026}	{.034}	{.008}	{.019}
Effect on log (median of 2003 through 2007 annual earnings)				
Parameter estimate for	0.143	0.026	0.080	0.023
school quality measure	(.018)	(.023)	(.009)	(.012)
N = 11,894	{.032}	{.039}	{.010}	{.017}

Table 7

Effect of College Characteristics on 2007 Earnings, 1989 Cohort of Men and Women

	College characteristic					
	School SAT score/100		Log net tuition		Barron's index	
	Basic	Self-revelation	Basic	Self-revelation	Basic	Self-revelation
Parameter estimate for	0.056	-0.008	-0.011	-0.108	0.069	-0.002
effect of quality measure	(.014)	(.019)	(.025)	(.028)	(.017)	(.022)
on log 2007 earnings	{.031}	{.034}	{.062}	{.070}	{.038}	{.042}
Sample size	6,479		6,479		6,479	

Table 8

Effect of School Characteristics on 2007 Earnings (Black and Hispanic Students Only, 1989 Cohort)

	School SAT score/100		Log net tuition		Barron's index	
Dependent variable	Basic	Self-revelation	Basic	Self-revelation	Basic	Self-revelation
All black and Hispanic students						
Parameter estimate for	0.067	0.076	0.173	0.138	0.063	0.049
effect of quality measure	(.019)	(.032)	(.056)	(.071)	(.022)	(.036)
on log 2007 earnings	{.028}	{.042}	{.076}	{ .092}	{.033}	{.046}
Sample size	1,508		1,508		1,508	
All black and Hispanic students, excluding historically black colleges and universities						
Parameter estimate for	0.122	0.120	0.187	0.116	0.158	0.143
effect of quality measure	(.030)	(.042)	(.064)	(.079)	(.040)	(.053)
on log 2007 earnings	{.035}	{.056}	{.081}	{.101}	{.038}	{.051}
Sample size	995		995		995	

Dale and Krueger takeaways

- DK compare students who choose more- vs. less-selective colleges from the same set of admissions offers
- One concern: students with higher unobserved ability may systematically choose more selective schools, biasing estimates upward
- Conversely, frugal students may opt for cheaper schools and pursue higher-paying jobs, potentially biasing results downward
- Striking result: once DK control for observable characteristics of students' application portfolios, the estimated advantage of selective colleges largely disappears

College Majors

- Heterogeneity in pecuniary returns across majors rivals the college wage premium
  (Altonji et al., 2012, 2017; Hastings et al. 2013; Kirkeboen et al. 2016)
- Earnings inequality within education groups has risen substantially
- Men and women tend to choose different college majors and that has implications for the gender earnings gap
  (Sloane et al., 2019; Aguirre et al, 2022; Patnaik et al. 2021; Ahimbisibwe et al. 2024)
- College majors represent multiple unordered treatments, complicating treatment effect estimation without strong assumptions
- We will discuss:
  - → Kirkeboen et al. 2016
  - → Campos et al. 2025

Kirkebøen, Leuven, and Mogstad (QJE 2016)

- Kirkebøen, Leuven, and Mogstad (QJE 2016) study field of study returns in Norway
- Substantive questions:
  - → What are the earnings returns to different fields?
  - → Do individuals sort across fields based on comparative advantage?
- Methodological question: With multiple treatments and instruments, how can we estimate treatment effects?

Institutional Context: Norwegian College Admissions

- Norway uses a centralized college admissions system
- Students apply to field-institution pairs (e.g., teaching at Univ. of Oslo)
- Applicants can rank up to 15 choices
- Scores are based on high school GPA; students are ranked accordingly
- Admissions are sequential: top student gets first available option, and so on
- Cutoff scores serve as instruments for field-specific treatment

Potential Outcomes Framework (3-choice case)

- Let $Y_i(0), Y_i(1), Y_i(2)$ be potential outcomes
- Instrument: $Z_i \in \{0, 1, 2\}$ (field offered)
- Potential treatment: $D_i(0), D_i(1), D_i(2)$ (field chosen given Z_i)
- Define indicators:

$$D_{ij}(z) = \mathbf{1}\{D_i(z) = j\}, \quad Z_{ij} = \mathbf{1}\{Z_i = j\}$$

Observed Field and Outcome

- Field choices:

$$D_{i1} = D_{i1}(0) + [D_{i1}(1) - D_{i1}(0)]Z_{i1} + [D_{i1}(2) - D_{i1}(0)]Z_{i2}$$
$$D_{i2} = D_{i2}(0) + [D_{i2}(1) - D_{i2}(0)]Z_{i1} + [D_{i2}(2) - D_{i2}(0)]Z_{i2}$$

- Outcome:

$$Y_i = Y_i(0) + [Y_i(1) - Y_i(0)]D_{i1} + [Y_i(2) - Y_i(0)]D_{i2}$$

- More compact:

$$D_i = D_i(Z_i), \quad Y_i = Y_i(D_i(Z_i))$$

Assumptions

- Exogeneity of instrument:

$$(Y_i(0), Y_i(1), Y_i(2), D_i(0), D_i(1), D_i(2)) \perp Z_i$$

- Monotonicity:

$$D_{i1}(1) \geq D_{i1}(0), \quad D_{i1}(1) \geq D_{i1}(2) \quad \forall i$$
$$D_{i2}(2) \geq D_{i2}(0), \quad D_{i2}(2) \geq D_{i2}(1) \quad \forall i$$

- Interpretation: raising Z_i increases probability of choosing the corresponding field.

Treatment Effects and Estimation

- The goal is to estimate

$$\Delta_i(j) = Y_i(j) - Y_i(0)$$

- Estimating equation:

$$Y_i = \beta_0 + \beta_1 D_{i1} + \beta_2 D_{i2} + \varepsilon_i$$

- What do we learn from OLS and IV applied to this model?

OLS Estimates

$$\beta_1^{OLS} = \mathbb{E}[\Delta_i(1) \mid D_i = 1] + \mathbb{E}[Y_i(0) \mid D_i = 1] - \mathbb{E}[Y_i(0) \mid D_i = 0]$$
$$\beta_2^{OLS} = \mathbb{E}[\Delta_i(2) \mid D_i = 2] + \mathbb{E}[Y_i(0) \mid D_i = 2] - \mathbb{E}[Y_i(0) \mid D_i = 0]$$

- OLS = TOT + selection bias

What Happens with IV?

- General IV with many treatments/instruments is hard to interpret
- Suppose we focus on one instrument at a time:

$$Y_i = \alpha_1 + \gamma_1 D_{i1} + u_i$$

- If LATE assumptions hold, then:

$$\gamma_1 = \mathbb{E}[Y_i(1) - Y_i(D_i(0)) \mid D_i(1) = 1, D_i(0) \neq 1]$$

which is a weighted average of sub-LATES:

$$\begin{aligned}\gamma_1 &= \omega_1 \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i(1) = 1, D_i(0) = 0] \\ &\quad + (1 - \omega_1)\mathbb{E}[Y_i(1) - Y_i(2) \mid D_i(1) = 1, D_i(0) = 2]\end{aligned}$$

- IV estimates are a weighted average of effects that draw in students from different counterfactuals
  - → $C_{12} = \{D_i(1) = 1,\ D_i(0) = 2\}$: compliers drawn into 1 with fallback 2
  - → $C_{10} = \{D_i(1) = 1,\ D_i(0) = 0\}$: compliers drawn into 1 with fallback 0
- Weight identified as:

$$1 - \omega_1 = \frac{P(C_{12})}{P(C_{10}) + P(C_{12})} = \frac{\mathbb{E}[D_{i2} \mid Z_i = 0] - \mathbb{E}[D_{i2} \mid Z_i = 1]}{\mathbb{E}[D_{i1} \mid Z_i = 1] - \mathbb{E}[D_{i1} \mid Z_i = 0]}$$

Underdetermined System

$$\gamma_1 = \omega_1 \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i(1) = 1, D_i(0) = 0] \quad + (1 - \omega_1)\mathbb{E}[Y_i(1) - Y_i(2) \mid D_i(1) = 1, D_i(0) = 2]$$
$$\gamma_2 = \omega_2 \mathbb{E}[Y_i(2) - Y_i(0) \mid D_i(2) = 2, D_i(0) = 0] \quad + (1 - \omega_2)\mathbb{E}[Y_i(2) - Y_i(1) \mid D_i(2) = 2, D_i(0) = 1]$$

- Each LATE equation includes two unknowns—additional assumptions needed to identify both

Constant Treatment Effects

- Suppose treatment effects are constant:

$$Y_i(1) - Y_i(0) = \Delta_1, \quad Y_i(2) - Y_i(0) = \Delta_2 \quad \forall i$$

- Then the LATEs become:

$$\gamma_1 = \omega_1 \Delta_1 + (1 - \omega_1)(\Delta_1 - \Delta_2)$$
$$\gamma_2 = \omega_2 \Delta_2 + (1 - \omega_2)(\Delta_2 - \Delta_1)$$

- Two equations in two unknowns — solvable system

Strong assumptions that allow IV to “work”

- Propositions 1 and 2: IV generally yields uninterpretable combinations of treatment effects
- Exceptions:
  1. Constant treatment effects
  2. Restrictive preferences:
$$D_{i2}(1) = D_{i2}(0), \quad D_{i1}(2) = D_{i1}(0)$$
- In (1), system is identified using classical simultaneous equations
- In (2), $\omega_1 = \omega_2 = 1$, so only two unknowns remain
- But restrictive preferences are strong assumptions—imply that only one type of substitution occurs. Switching Z from 0 to 1 can only induce movement from 0 into 1 and similarly for switching Z from 0 to 2

Using Preference Lists: Key Advantage of the Data

- Without strong assumptions, IV cannot identify field-specific causal effects
- Unique feature of Norwegian data: fallback choice is observable because students submit rank-ordered lists
- Since students rank all options, we observe $D_i(0)$, their fallback
- This enables us to condition on what field they would attend if not admitted to their top choice

Restricting the Sample Based on Fallbacks

$$Y_i = \beta_0 + \beta_1 D_{i1} + \beta_2 D_{i2} + \varepsilon_i$$

- Estimate equation using IV in subsample where $D_i(0) = 0$.
- Then:

$$\beta_1^{IV} = \mathbb{E}[\Delta_i(1) \mid D_i(1) = 1, D_i(0) = 0]$$
$$\beta_2^{IV} = \mathbb{E}[\Delta_i(2) \mid D_i(2) = 2, D_i(0) = 0]$$

- Conditional estimation mimics restrictive preference setting by dropping people along substitution margins we do not want

Extending to Many Fields: 2SLS System

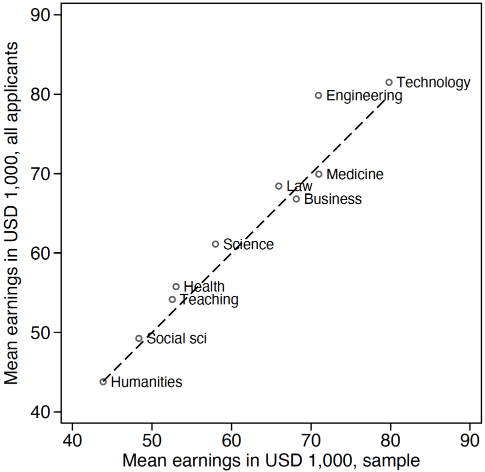
- Condition on fallback field k, then estimate:

$$Y_i = \sum_{j \neq k} (\alpha_{jk} + \beta_{jk} D_{ij}) + X_i' \delta_k + \varepsilon_i$$

$$D_{ij} = \sum_{\ell \neq k} (\lambda^j_{\ell k} + \pi^j_{\ell k} Z_{ij}) + X_i' \theta^j_k + \eta_{ij}$$

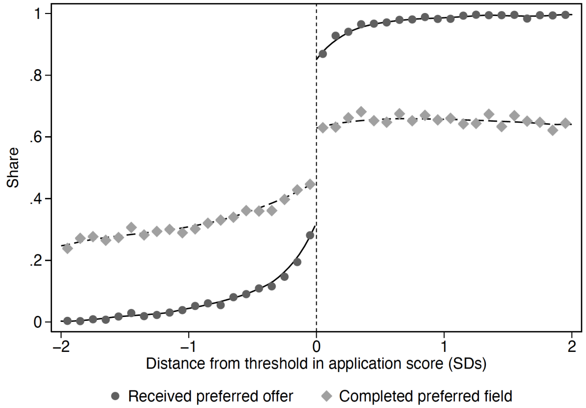
- X_i includes running variable (distance to cutoff)
- Theoretically:

$$\beta_{jk} = \mathbb{E}[Y_i(j) - Y_i(k) \mid D_i(j) = j, D_i(k) = k, j \text{ ranked above } k]$$



Note: This figure reports mean earnings by field for our sample of applicants and for all applicants. Earnings are measured eight years after application. The measures of earnings are regression adjusted for year of application.

Figure 2. Mean earnings by field: Sample and all applicants



Note: This figure shows the sample fraction that is offered or complete the preferred field by application score. We pool all admission cutoffs and normalize the data so that zero on the x-axis represents the admission cutoff to the preferred field. We plot unrestricted means in bins and include estimated local linear regression lines on each side of the cutoff.

Figure 3. Admission thresholds and preferred field offer and completion

TABLE IV

2SLS estimates of the payoffs to field of study (USD 1,000)

	Next Best Alternative (k)								
	Humanities	Soc Science	Teaching	Health	Science	Engineering	Technology	Business	Law
Completed field (j)									
Humanities		21.4*	−4.7	−22.9*	5.0	−38.5**	6.9	−42.2**	−156.3
		(11.0)	(9.8)	(12.1)	(11.9)	(14.7)	(48.3)	(10.6)	(437.3)
Social Science	18.7**		9.8	−10.8	55.5**	−55.4**	−110.4	−28.4**	−76.1
	(6.7)		(11.6)	(13.0)	(21.5)	(20.6)	(103.0)	(10.7)	(86.4)
Teaching	22.3**	31.4**		1.8	23.5**	−33.9**	−35.3	−21.1**	22.8
	(5.0)	(7.9)		(6.6)	(9.5)	(12.5)	(37.1)	(7.1)	(127.9)
Health	18.8**	30.7**	7.7**		28.9**	−27.9**	−43.4**	−17.4**	−55.2
	(6.3)	(7.6)	(2.8)		(7.6)	(10.4)	(20.8)	(4.0)	(97.7)
Science	53.7**	69.6**	38.6**	29.6**		−2.2	16.8	−4.9	148.3
	(18.4)	(22.4)	(14.2)	(11.5)		(14.6)	(18.1)	(10.5)	(276.2)
Engineering	59.8	−5.5	75.2**	0.2	52.4**		−46.0	−13.0	−57.7
	(50.6)	(58.2)	(37.5)	(16.4)	(21.0)		(43.9)	(23.7)	(166.6)
Technology	41.9**	58.7**	22.1*	32.5**	68.1**	−5.6		7.0	−53.1
	(10.8)	(10.1)	(12.4)	(10.1)	(9.6)	(12.0)		(9.5)	(147.5)

TABLE IV

(CONTINUED)

	Next Best Alternative (k)								
	Humanities	Soc Science	Teaching	Health	Science	Engineering	Technology	Business	Law
Business	48.1**	61.9**	31.0**	30.2**	58.0**	−3.4	28.5*		3.5
	(11.3)	(12.0)	(8.8)	(10.9)	(10.5)	(12.6)	(15.6)		(83.0)
Law	46.3**	55.6**	36.6**	21.5*	40.1**	−27.5	−15.6	−1.4	
	(7.2)	(8.3)	(11.6)	(11.5)	(9.7)	(18.3)	(18.0)	(8.7)	
Medicine	83.3**	79.4**	62.6**	45.6**	81.3**	21.1	40.1**	23.3**	14.8
	(9.8)	(10.7)	(9.0)	(7.0)	(9.7)	(20.7)	(11.7)	(8.8)	(83.6)
Female	−7.0**	−6.3**	−10.3*	−5.6**	−5.3**	−5.1**	−4.1**	−7.0**	−10.6
	(1.1)	(1.6)	(1.3)	(0.9)	(1.3)	(1.0)	(1.6)	(3.5)	(6.9)
Application score	−0.6	4.3**	4.0**	1.6**	−0.7	1.1*	−0.1	0.1	13.8
	(0.8)	(1.6)	(0.9)	(0.6)	(0.7)	(0.6)	(1.3)	(2.8)	(14.6)
Average y^k	30.0	23.4	46.2	51.8	27.3	87.9	78.4	75.6	105.8
Observations	8,391	11,030	10,987	3,269	6,422	3,085	1,245	4,403	1,251

Notes. From 2SLS estimation of equations (14)–(15), we obtain a matrix of the payoffs to field j as compared to k for those who prefer j and have k as next-best field. Each cell is a 2SLS estimate (with standard errors in parenthesis) of the payoff to a given pair of preferred field and next-best field. The rows represent completed fields and the columns represent next-best fields. The row labeled average y^k reports the weighted average of the levels of potential earnings for compliers in the given next-best field. The final row reports the number of observations for every next-best field. Stars indicate statistical significance, at the *10% level and **5% level.

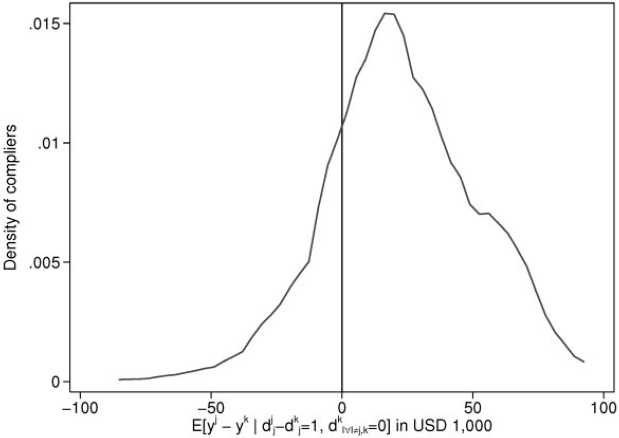


FIGURE VIII

Distribution of Estimated Payoffs ($1,000) to Field of Study

This figure graphs the complier-weighted distribution of estimates in Table IV. Three outliers with <0.001 of compliers are excluded.

Roy Model Interpretation

- β_{jk} and $-\beta_{kj}$ both measure treatment effects across the j vs. k margin
- Under Roy model:

$$\beta_{jk} > -\beta_{kj}$$

- Roy models imply higher returns for those with stronger preferences for a field
- Question: What does Roy-style selection imply about potential asymmetries in estimated effects?

Predictions from Roy Model

$$\beta_{jk} = \mathbb{E}[Y_i(j) - Y_i(k) \mid D_i(j) = j, D_i(k) = k, j \text{ ranked above } k]$$

- Roy model predicts $\beta_{jk} > -\beta_{kj}$
- Treatment effects larger for individuals with stronger revealed preferences
- Example: Going from k (humanities) to j (engineering) helps those who prefer j over k, but by less among those who prefer k over j

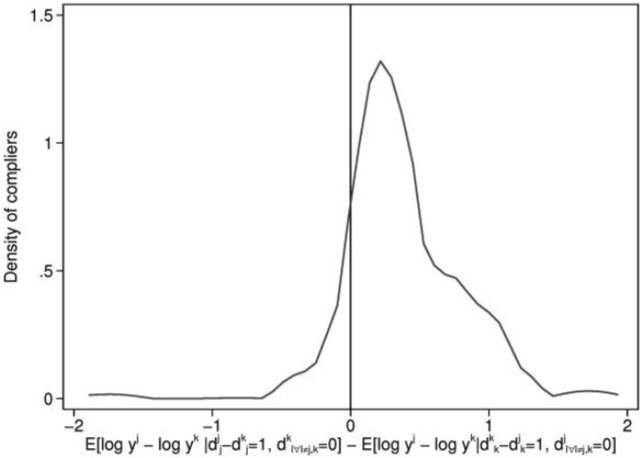
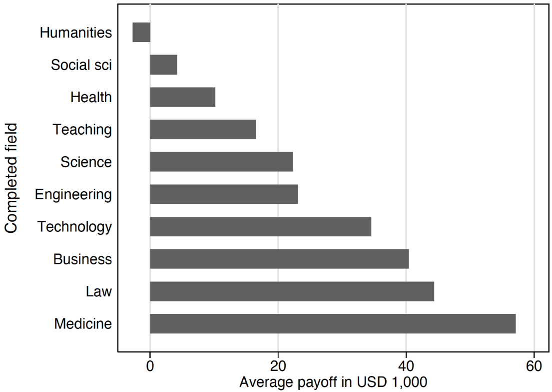


FIGURE XII

Testable Implication of Sorting Based on Comparative Advantage



Note: This figure graphs the weighted averages of payoffs to different completed fields by next-best field. The payoffs come from Table 5. For each field, the weights sum to one and reflect the proportion of compliers by next-best field.

Figure 9. Average estimated payoffs (1,000 USD) by completed field

Takeaways

- IV is hard to interpret with multiple endogenous variables and instruments
- Knowing preferences clarifies counterfactuals and enables identification
- Returns to field of study are large—comparable to returns to college
- Comparative advantage plays a key role:
  - $\rightarrow$ Gain from going from k to j when k is fallback is *smaller* than when j is preferred

Major Choice and Gender Gaps

- Heterogeneity in pecuniary returns across majors rivals the college wage premium
  (Altonji et al., 2012, 2017; Hastings et al. 2013; Kirkeboen et al. 2016)
- Men and women tend to choose different college majors and that has implications for the gender earnings gap
  (Sloane et al., 2019; Aguirre et al, 2022; Patnaik et al. 2021; Ahimbisibwe et al. 2024)
  - → Differences in ability do not explain these gaps
  - → Differences in beliefs about ability also do not entirely explain gaps

Major Choice and Gender Gaps

- Heterogeneity in pecuniary returns across majors rivals the college wage premium
  (Altonji et al., 2012, 2017; Hastings et al. 2013; Kirkeboen et al. 2016)
- Men and women tend to choose different college majors and that has implications for the gender earnings gap
  (Sloane et al., 2019; Aguirre et al, 2022; Patnaik et al. 2021; Ahimbisibwe et al. 2024)
  - → Differences in ability do not explain these gaps
  - → Differences in beliefs about ability also do not entirely explain gaps

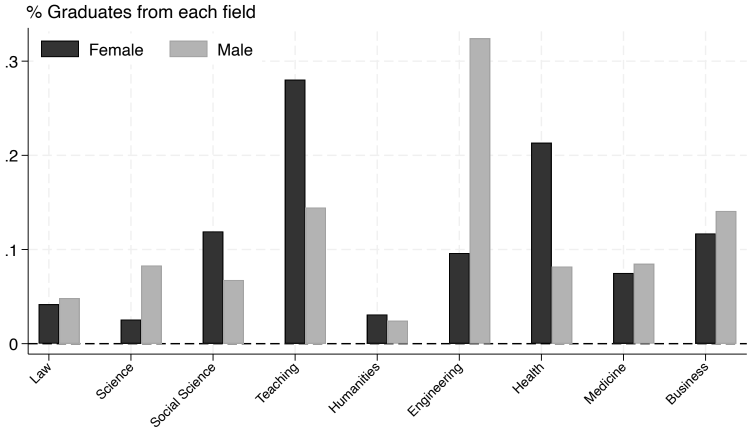
Open Question: Why do men and women tend to choose different majors?

→ Do college majors impact pecuniary *and* non-pecuniary outcomes?

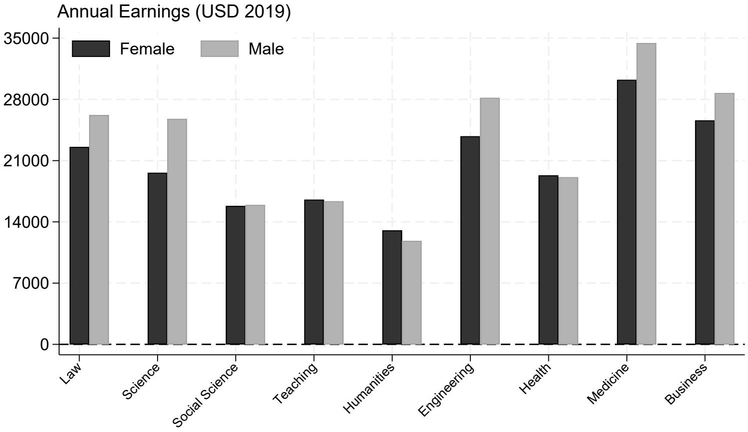
→ Are there gender differences in preferences for pecuniary and non-pecuniary impacts?

Methodological: Can we combine insights from DK and KLM to estimate returns?

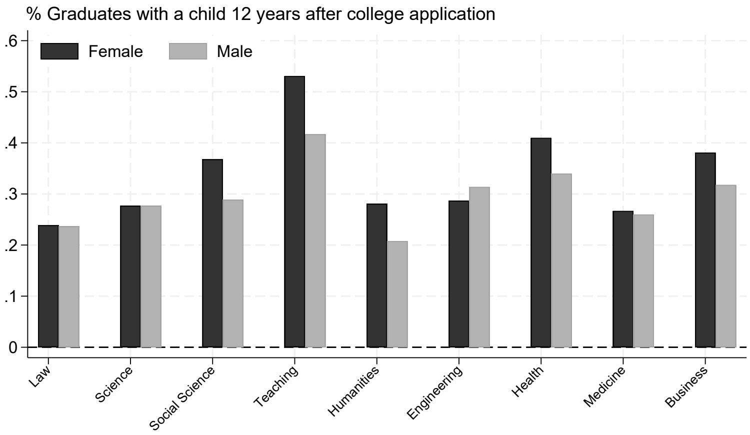
Motivation: In Chile, women prefer teaching, while men prefer STEM



Motivation: In Chile, men sort into more lucrative fields



Motivation: In Chile, women sort into fields with higher fertility rates



Vertical and Horizontal Returns

Environment

- Index the population of college applicants by $i \in \mathcal{I}$ and majors by $j \in \mathcal{J}$
- Applicants submit a rank-ordered list, $R_i \in \mathcal{R}$, containing a partial ordering of majors, and also report an entrance exam score, $S_i \in \mathcal{S}$
- A student type is defined as $\theta_i = (R_i, S_i) \in \Theta$
- A centralized assignment generates a single offer for each student, $Z_i \in \mathcal{J}$

Vertical and Horizontal Returns

Environment

- Index the population of college applicants by $i \in \mathcal{I}$ and majors by $j \in \mathcal{J}$
- Applicants submit a rank-ordered list, $R_i \in \mathcal{R}$, containing a partial ordering of majors, and also report an entrance exam score, $S_i \in \mathcal{S}$
- A student type is defined as $\theta_i = (R_i, S_i) \in \Theta$
- A centralized assignment generates a single offer for each student, $Z_i \in \mathcal{J}$

Potential Outcomes:

The potential outcome of student i completing major j is

$$Y_{ij} = Y_{i0} + \mu_{ij} \qquad (1)$$

- $a_i \equiv Y_{i0}$: summarizes student i's ability (or their outcome in the case they do not complete college)
- μ_{ij}: idiosyncratic return of i from completing major j

Vertical and Horizontal Returns

Assumption 1

Student types proxy for student ability,

$$a_i = \sum_{\theta \in \Theta} \phi_\theta 1[\theta_i = \theta] + u_i,$$

with $E[u_i|M_i, \theta_i] = E[u_i|\theta_i]$.

- This Assumption is useful for identification and similar to Dale and Krueger 2002 and Mountjoy and Hickman 2021
- Some relationship to Abduklkadiroglu et al. 2022 that isolates random variation embedded in systems of centralized assignment

Vertical and Horizontal Returns

Assumption 2

Let $Z_i \in \mathcal{J}$ correspond to a student's single offered major and let D_i be an indicator for compliance with the assignment. Let F_i correspond to their enrolled or completed major,

$$F_i = D_i Z_i + (1 - D_i) F_i',$$

where F_i' is a major student i completes in case they do not comply. Compliance, D_i is as good as random conditional on type θ:

$$D_i \perp\!\!\!\perp \mu_{ij} \mid \theta \quad \text{and} \quad D_i \perp\!\!\!\perp u_i \mid \theta.$$

- This assumption rules out a particular type of sorting on gains analogous to Hoxby 2009 critique of Dale and Krueger
- Assumption is similar to Angrist et al. 2024 requiring students to comply with their randomized offer to estimate risk-controlled value-added models
- Assumption more plausible in settings with a single offer and centralized assignment

Vertical and Horizontal Returns

Compare mean outcomes of major j graduates to non-graduates ($j = 0$):

$$E[Y_i|M_i = j, \theta_i] - E[Y_i|M_i = 0, \theta_i] = \underbrace{E[\mu_{ij}|M_i = j, \theta_i]}_{ATT_j(\theta)} + \underbrace{E[a_i|M_i = k, \theta_i] - E[a_i|M_i = 0, \theta_i]}_{\text{Selection bias = 0, by Assumption 1}}. \quad (1)$$

Vertical and Horizontal Returns

Compare mean outcomes of major j graduates to non-graduates ($j = 0$):

$$E[Y_i|M_i = j, \theta_i] - E[Y_i|M_i = 0, \theta_i] = \underbrace{E[\mu_{ij}|M_i = j, \theta_i]}_{ATT_j(\theta)} + \underbrace{E[a_i|M_i = k, \theta_i] - E[a_i|M_i = 0, \theta_i]}_{\text{Selection bias = 0, by Assumption 1}}. \quad (1)$$

We can write $ATT_j(\theta)$ as:

$$ATT_j(\theta) = \underbrace{E[\ \mu_{ij} \mid \theta\ ]}_{ATE_j(\theta)} + \underbrace{(E[\mu_{ij} \mid M_i = j, \theta] - E[\mu_{ij}|\theta])}_{\text{Sorting on Gains into Major conditional on type = 0, by Assumption 2}}$$

Vertical and Horizontal Returns

Compare mean outcomes of major j graduates to non-graduates ($j = 0$):

$$E[Y_i|M_i = j, \theta_i] - E[Y_i|M_i = 0, \theta_i] = \underbrace{E[\mu_{ij}|M_i = j, \theta_i]}_{ATT_j(\theta)} + \underbrace{E[a_i|M_i = k, \theta_i] - E[a_i|M_i = 0, \theta_i]}_{\text{Selection bias = 0, by Assumption 1}}. \quad (1)$$

We can write $ATT_j(\theta)$ as:

$$ATT_j(\theta) = \underbrace{E[\,\mu_{ij} \mid \theta\,]}_{ATE_j(\theta)} + \underbrace{(E[\mu_{ij} \mid M_i = j, \theta] - E[\mu_{ij}|\theta])}_{\text{Sorting on Gains into Major conditional on type = 0, by Assumption 2}}$$

Additional TE heterogeneity nested in $ATE_j(\theta)$. A type-specific observation comparison recovers:

$$ATE_j(\theta) = \underbrace{E[\mu_{ij}|\mathcal{I}]}_{ATE_j(\mathcal{I})} + \underbrace{(E[\mu_{ij}|\theta] - E[\mu_{ij}|\mathcal{I}])}_{M_{ij}}, \quad (2)$$

Vertical and Horizontal Returns

Given Assumption 1 and 2, type-specific observational comparisons recover:

$$E[Y_i|M_i = j, \theta_i] - E[Y_i|M_i = 0, \theta_i] = \underbrace{E[\mu_{ij}|\mathcal{I}]}_{ATE_j(\mathcal{I})} + \underbrace{(E[\mu_{ij}|\theta] - E[\mu_{ij}|\mathcal{I}])}_{M_{ij}}, \quad (1)$$

- $ATE_j(\mathcal{I})$: average treatment effect among the pool of *all* college applicants
- M_{ij} : sorting on gains among type θ relative to typical college applicant
- Our approach adopts a mix of empirical assumptions to identify each component and we validate using local random variation:
  - → $ATE_j(\mathcal{I})$: parametric assumptions on potential outcomes allow us to identify heterogeneous mean return that varies by gender
  - → M_{ij} : a choice model along with distributional assumptions on unobserved preference heterogeneity allows us to identify this kind of sorting on gain

Major Choice

We model student i's indirect utility from majoring in field j as:

$$u_{ij} = \sum_{k \in \mathcal{K}} \rho^k_{c(i)} \mu^k_j(X_i) + \lambda_{dc(i)} d_{ij} + \lambda_{pc(i)} p_{c(i)j} + \eta_{ij} \tag{2}$$

- $\mu^k_j(X_i)$: outcome k mean among those with observables X_i; agents have preferences for different mean outcomes through ρ^k_c
- $p_{c(i)j}$: net price for major j among individuals in cell c
- d_{ij} is the distance between student i high-school and the campus of the major j ranked (preference shifter)
- Unobserved preference heterogeneity is captured by $\eta_{ij} \sim$ EV Type 1
- $c(i)$ stand for the *cell* to which student i belongs. A cell is defined by: i) Geographic location, ii) High-school type, iii) PSU score ranges, iv) Gender

Major Choice

$$u_{ij} = \underbrace{\sum_{k \in \mathcal{K}} \rho^k_{c(i)} ATE^k_j(X_i) + \gamma_{pc(i)} p_{ij} + \gamma_{dc(i)} d_{ij}}_{\delta_{jc} := \text{mean utility}} \quad + \quad \eta_{ij}$$

- The empirical indirect utility specification is:

$$u_{ij} = \delta_{jc} + \gamma_{dc(i)} d_{ij} + \eta_{ij}$$

  → Assumption: Rational expectations on average returns

  → Assumption: Similar net price schedule among those in cell c

- Assume truthful reporting and estimate model via maximum likelihood ▸ Estimation Details

Potential Outcomes

$$Y_{ij} = \underbrace{\sum_l \psi_l(\eta_{i\ell} - \bar{\eta}) + \gamma X_i}_{\text{Ability: } a_i} + \underbrace{\underbrace{\alpha_j + \beta_j G_i}_{\text{Mean Return: } ATE_j(G_i)} + \underbrace{\psi_j^*(\eta_{ij} - \bar{\eta}) + e_{ij}}_{\text{Match Effect: } M_{ij}}}_{\text{Idiosyncratic Return: } \mu_{ij}}$$

- Ability depends on
  - → X_i, including entrance exam scores
  - → Preference heterogeneity η_{ij}
    - ∘ ψ_ℓ : captures selection on levels into major ℓ
- Idiosyncratic return μ_{ij} depends on
  - → Mean returns, ATE_j, that depend on gender G_i
  - → Match effects, M_{ij}, governed by preference heterogeneity
    - ∘ ψ_j^* : captures selection into major j on potential *gains* (comp. advantage)

Potential Outcomes

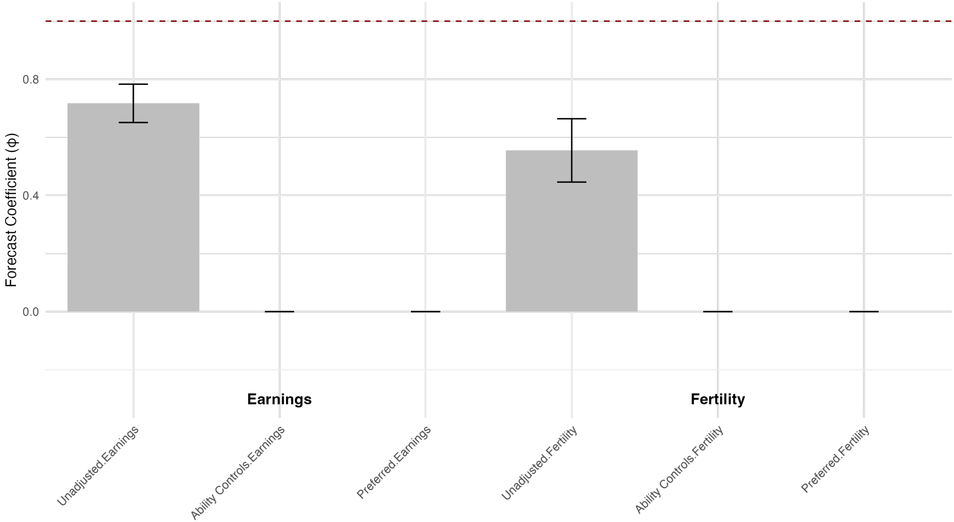
With rank-ordered choice data, R_i, the conditional expectation of an *observed* outcome Y_i is:

$$E[Y_i|R_i, X_i, G_i, j(i) = j] = \alpha_j + \beta_j G_i + \gamma X_i + \sum_{\ell} \psi_\ell \lambda_\ell(X_i, G_i, R_i) + \psi_j^* \lambda_j(X_i, G_i, R_i),$$

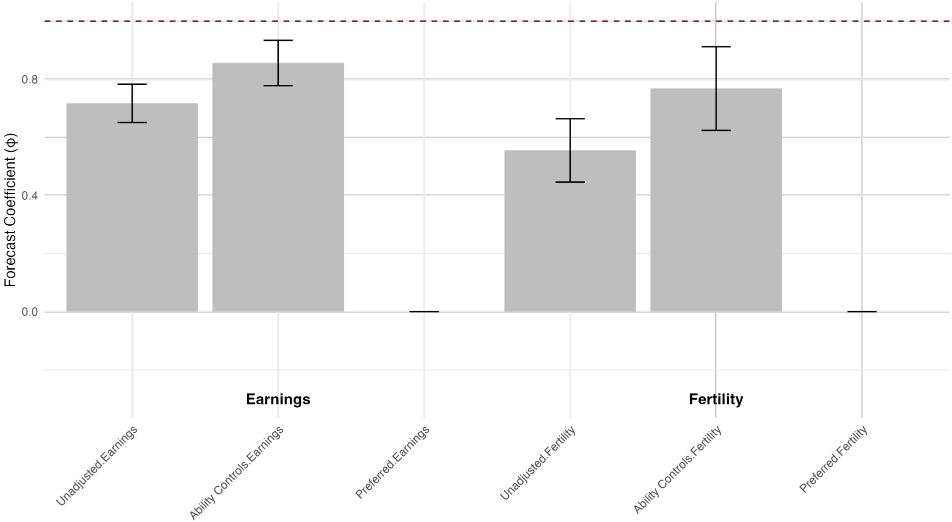
- $\lambda_j(X_i, G_i, R_i) \equiv E[\eta_{ij} - \bar{\eta} \mid X_i, G_i, R_i]$ are control functions derived from rank-ordered choice data
- Empirical match effects depend on observables X_i, gender, and the composition of the rank-ordered list, R_i
- The conditional average treatment effect of major j is

$$\tau_j(G_i, X_i, R_i) = \underbrace{\alpha_j + \beta_j G_i}_{ATE(G_i)} + \underbrace{\psi_j^* \lambda_j(X_i, G_i, R_i)}_{M_{ij}}$$

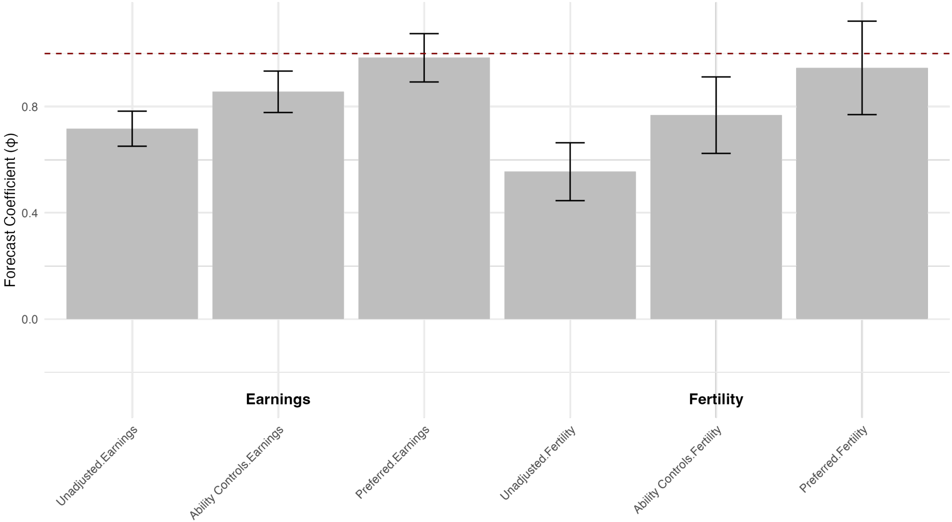
Uncontrolled models are forecast biased



Adding ability controls helps but estimates are still forecast biased

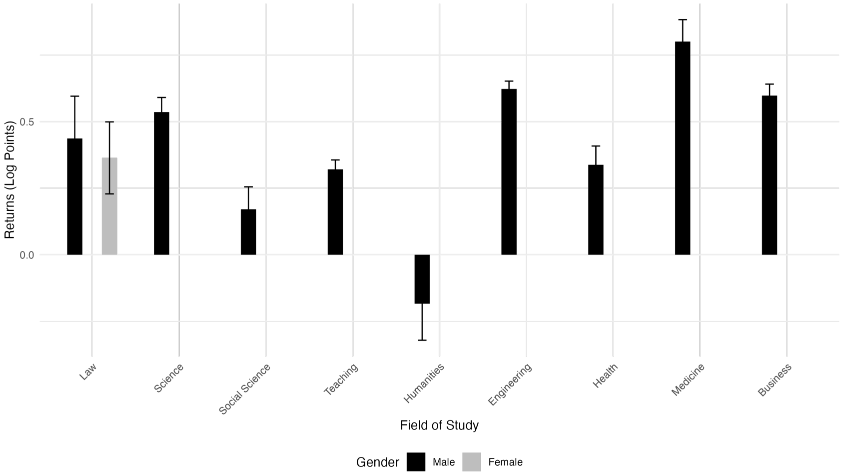


Preferred model produces estimates that are forecast unbiased



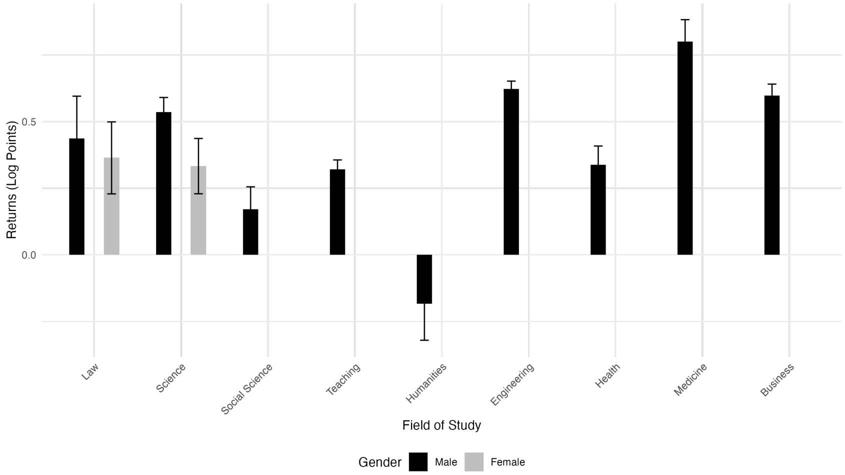
Returns to Majors Vary Substantially ($\mu_F = 0.43$, $\mu_M = 0.51$; $\sigma_F = 0.23$, $\sigma_M = 29$)

Men and women experience a similar return in Law



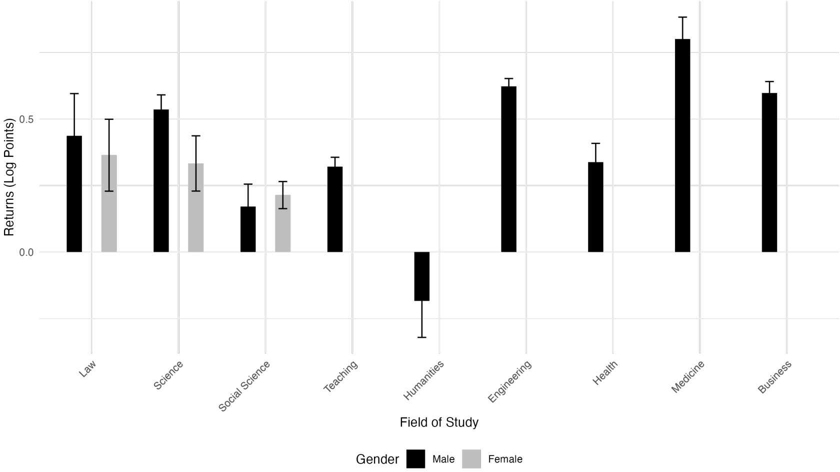
Returns to Majors Vary Substantially ($\mu_F = 0.43$, $\mu_M = 0.51$; $\sigma_F = 0.23$, $\sigma_M = 29$)

Women experience a lower return in Science



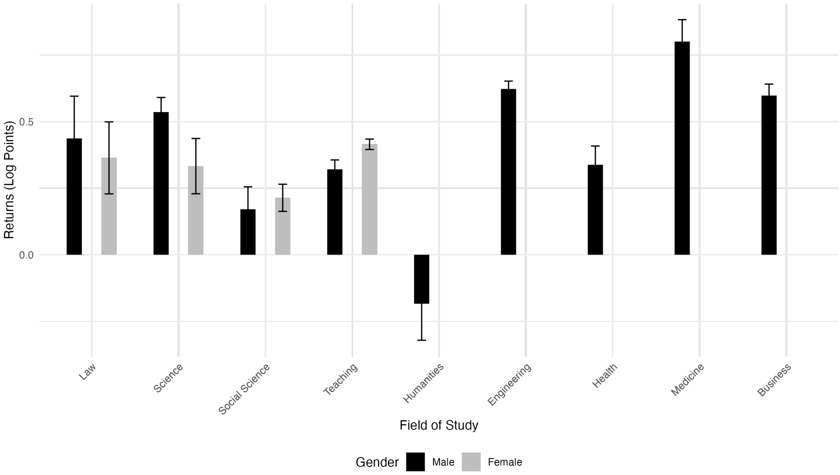
Returns to Majors Vary Substantially ($\mu_F = 0.43$, $\mu_M = 0.51$; $\sigma_F = 0.23$, $\sigma_M = 29$)

Men and women experience a similar return in Social Science



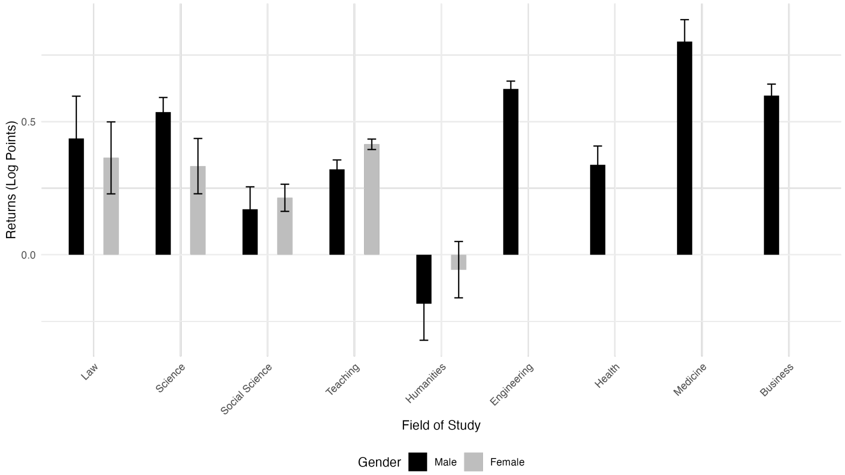
Returns to Majors Vary Substantially ($\mu_F = 0.43$, $\mu_M = 0.51$; $\sigma_F = 0.23$, $\sigma_M = 29$)

Women obtain a larger return in Teaching



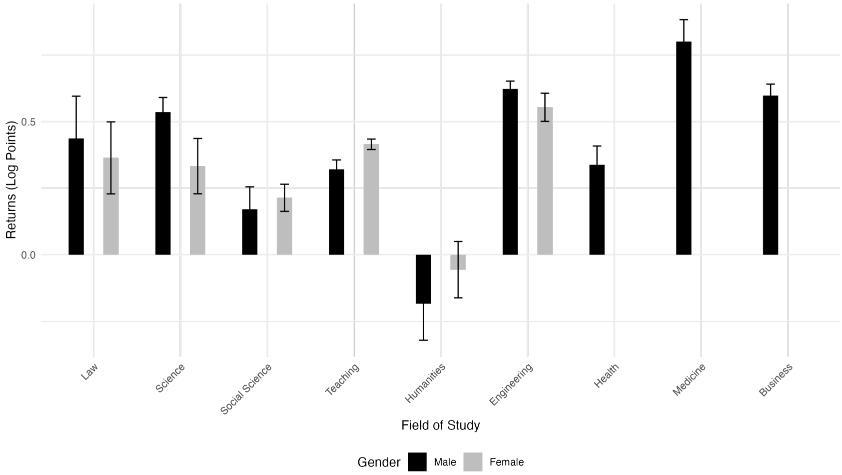
Returns to Majors Vary Substantially ($\mu_F = 0.43$, $\mu_M = 0.51$; $\sigma_F = 0.23$, $\sigma_M = 29$)

Men and women experience a similar return in Humanities



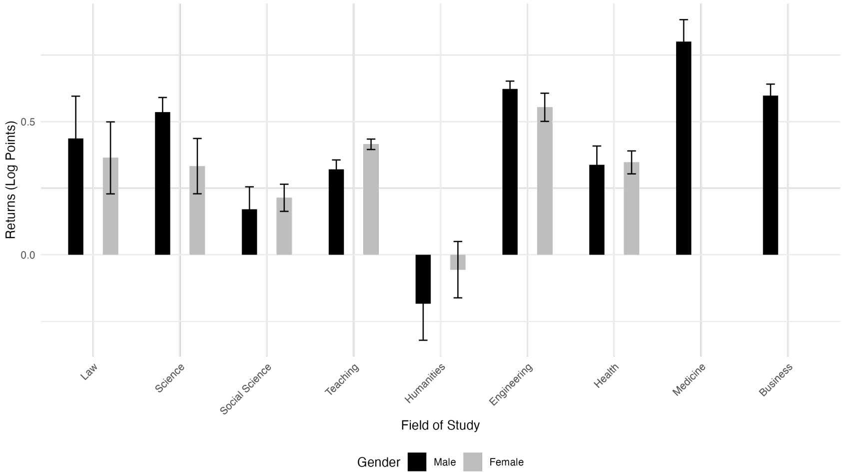
Returns to Majors Vary Substantially ($\mu_F = 0.43$, $\mu_M = 0.51$; $\sigma_F = 0.23$, $\sigma_M = 29$)

Women receive a modestly lower return in Engineering



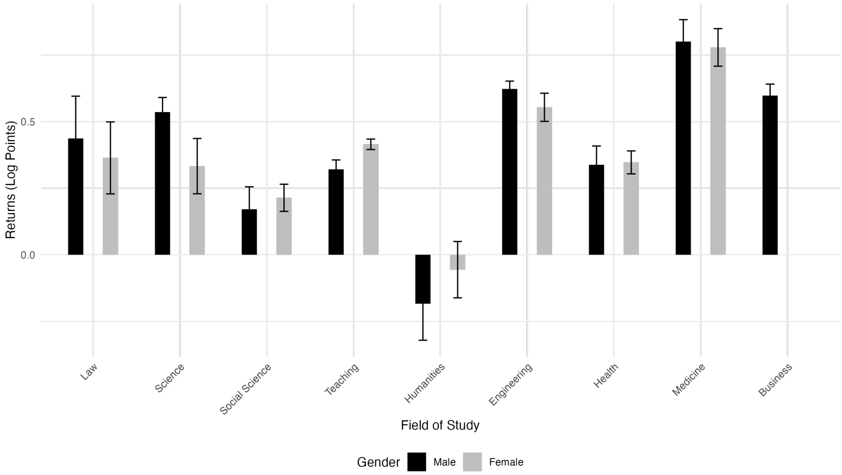
Returns to Majors Vary Substantially ($\mu_F = 0.43$, $\mu_M = 0.51$; $\sigma_F = 0.23$, $\sigma_M = 29$)

Men and women experience a similar return in Health



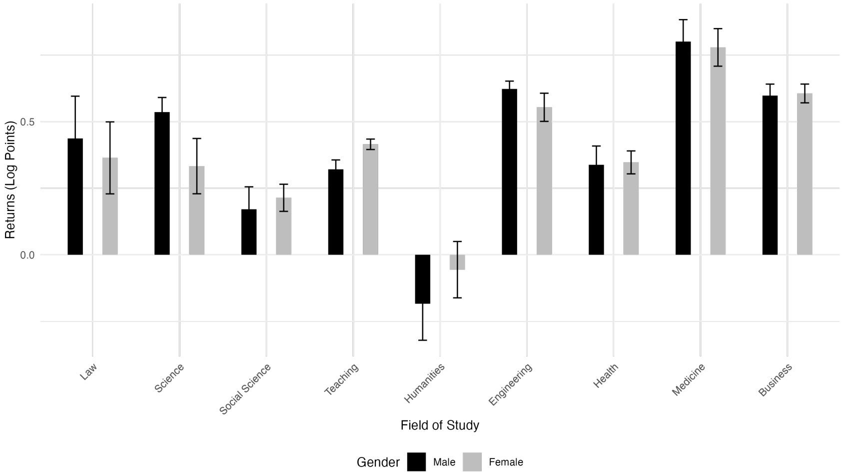
Returns to Majors Vary Substantially ($\mu_F = 0.43$, $\mu_M = 0.51$; $\sigma_F = 0.23$, $\sigma_M = 29$)

Men and women experience a similar return in Medicine



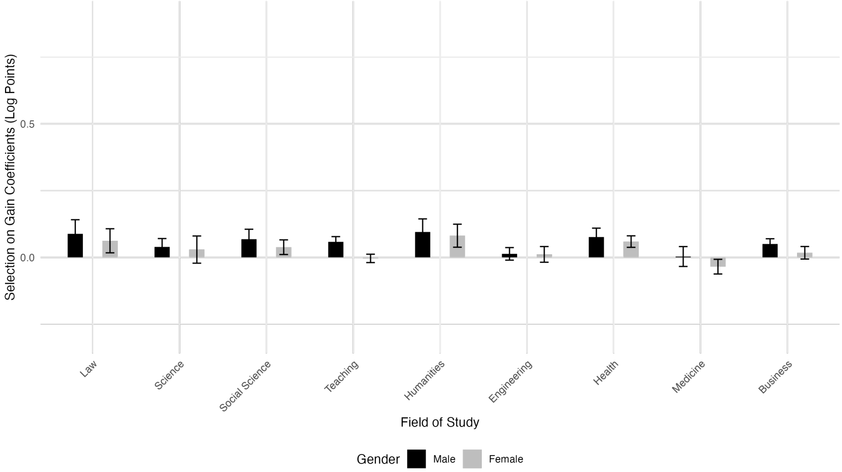
Returns to Majors Vary Substantially ($\mu_F = 0.43$, $\mu_M = 0.51$; $\sigma_F = 0.23$, $\sigma_M = 29$)

Men and women experience a similar return in Business



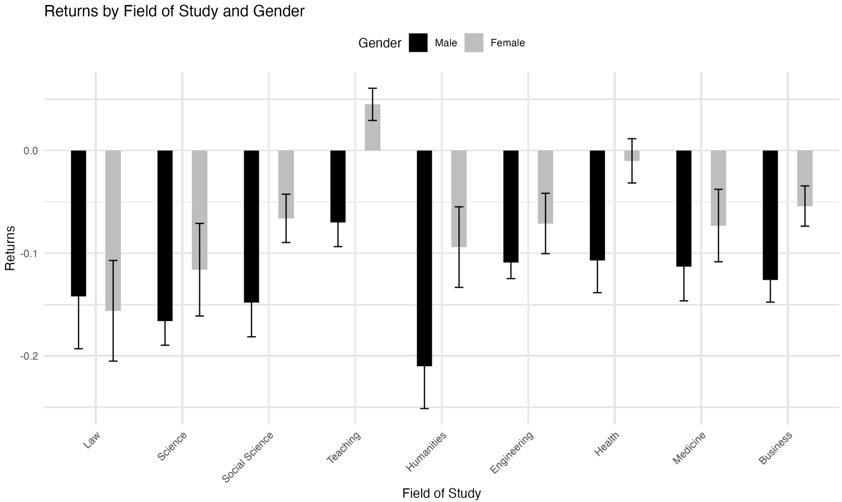
Evidence of positive selection on earnings gains

No evidence of gender differences



Modest variation in fertility impacts ($\mu_F = -0.03$, $\mu_M = -0.11$; $\sigma_F = 0.03$, $\sigma_M = 0.06$)

A near across the board more negative fertility impact for men



The importance of earnings and fertility for major choice

- The connection between majors and earnings follows from canonical human capital models (Becker, 1964)
- Survey evidence suggests that family considerations are determinants of major choice (Wiswall and Zafar 2021)
  - → Are stated preferences related to estimated returns?
  - → Are there gender differences?

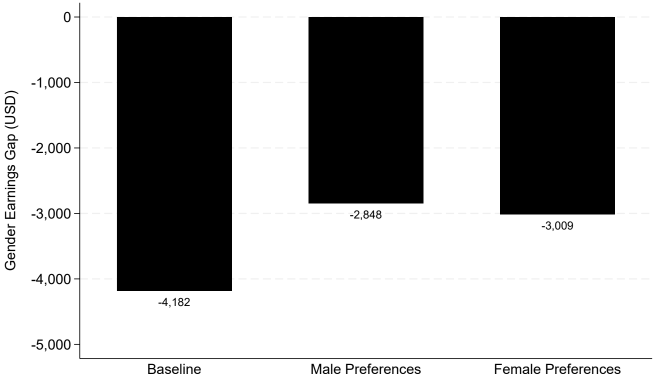
The importance of earnings and fertility for major choice

- The connection between majors and earnings follows from canonical human capital models (Becker, 1964)
- Survey evidence suggests that family considerations are determinants of major choice (Wiswall and Zafar 2021)
  - → Are stated preferences related to estimated returns?
  - → Are there gender differences?
- We can empirically study this by leveraging our estimated mean-utilities and returns at the cell level:

$$\hat{\delta}_{jc} = \rho_G^E A\bar{T}E_{jc}^E + \rho_G^F A\bar{T}E_{jc}^F + \gamma' X_i + e_{jc},$$

  - → $A\bar{T}E_{jc}^E$ and $A\bar{T}E_{jc}^F$ represent the returns of major j graduates for each cell c
- Equipped with these estimates, we can quantify the contribution to the overall gender earnings gap

Differences in preferences for family considerations account for 17-24% of the earnings gap



Policy Counterfactuals

- We take our estimates and preferences as given and estimate counterfactual outcomes under different assignment schemes:

    1. Gender-neutral Engineering seat expansions: 10%, 30%, 50%
    2. Gender quotas in Engineering programs: 10%, 30%, 50%

Policy Counterfactuals

- We take our estimates and preferences as given and estimate counterfactual outcomes under different assignment schemes:

  1. Gender-neutral Engineering seat expansions: 10%, 30%, 50%
  2. Gender quotas in Engineering programs: 10%, 30%, 50%

- Technically and politically feasible to reallocate students in settings with centralized assignments

  → 74% of Chilean citizens support affirmative action policies (Bursztyn et al., 2023)

Why focus on Engineering?

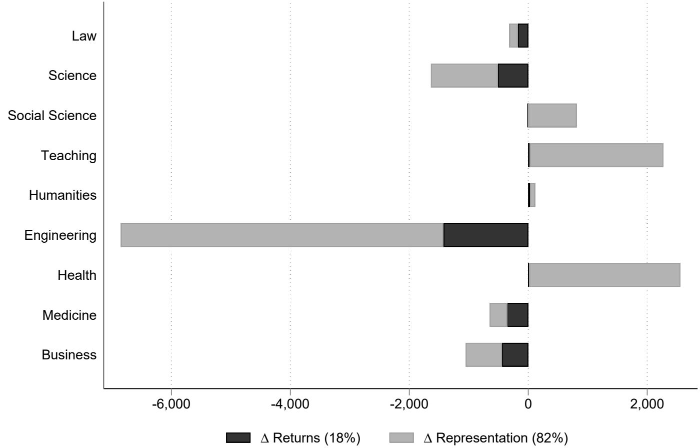
A Simple Decomposition: $\mathbb{E}[\textit{Earnings} \mid \textit{Female}] - \mathbb{E}[\textit{Earnings} \mid \textit{Male}]$

$$\sum_j s_j^F \mu_j^F - \sum_j s_j^M \mu_j^M = \underbrace{\sum_j (s_j^F - s_j^M)\mu_j^F}_{\Delta \text{Representation}} + \underbrace{\sum_j s_j^M (\mu_j^F - \mu_j^M)}_{\Delta \text{Returns}}$$

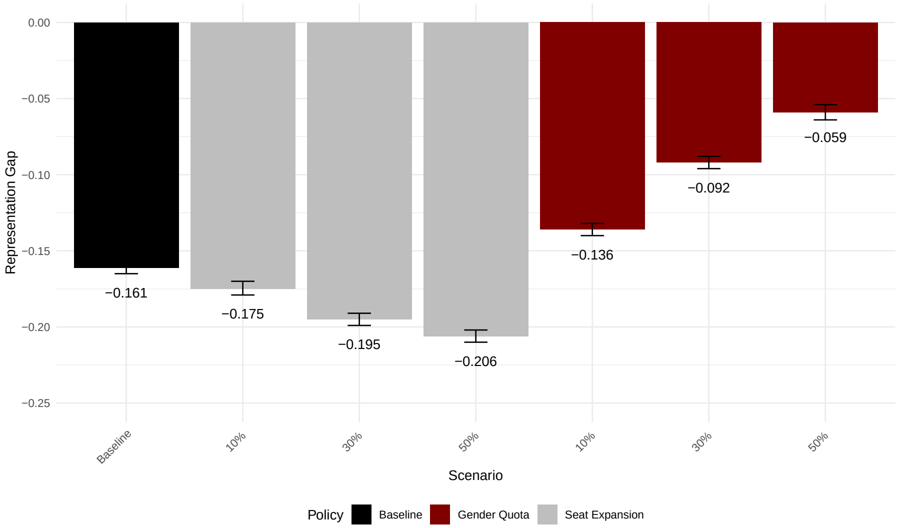
- s_j^F, s_j^M : major j graduation shares for women and men, respectively
- μ_j^F, μ_j^F : major j average earnings for women and men, respectively

Why focus on Engineering?

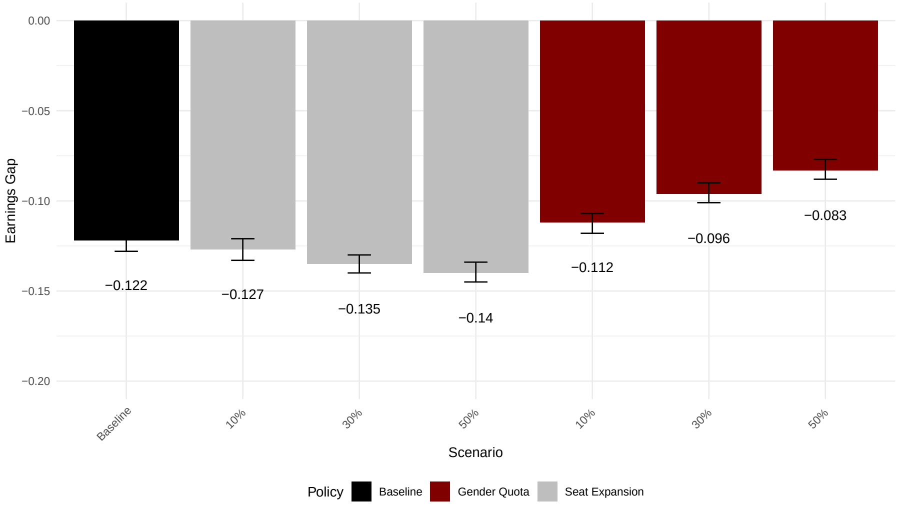
A Simple Decomposition: $\mathbb{E}[Earnings \mid Female] - \mathbb{E}[Earnings \mid Male]$



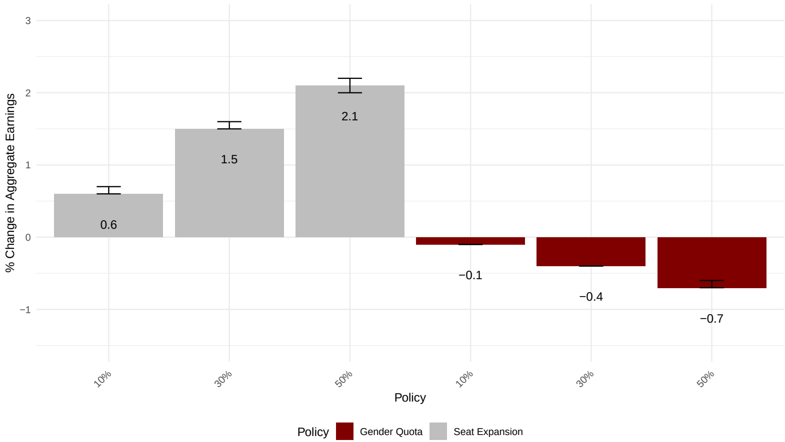
Counterfactual Representation Gaps



Counterfactual Earnings Gap



Limited impacts on efficiency



▸ Why no impact on efficiency?

▸ Who are the women nudged into Engineering?

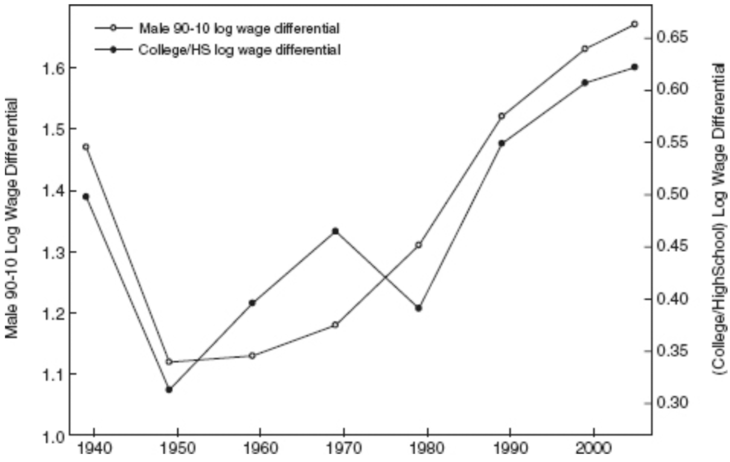
College Wage Premiums and Inequality

Explaining the evolving returns to education over the 20th century

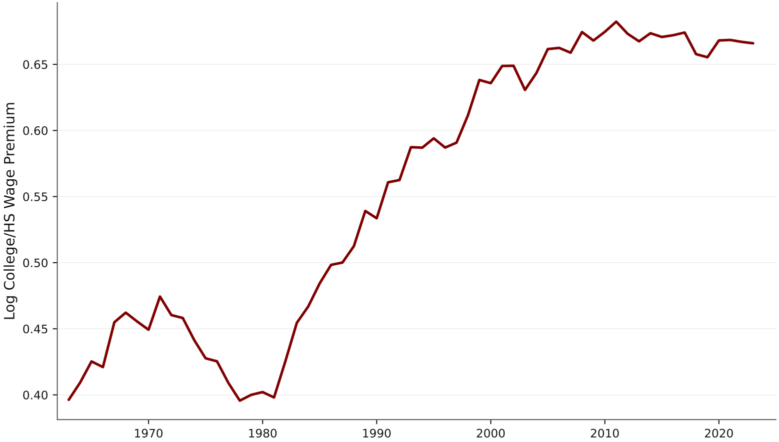
- A simple supply and demand framework adequately captures the evolving nature of the college wage premium across the 20th century
- A few facts motivate:
  1. The college wage premium was large at the turn of the 20th century, decreased in the middle, and began increasing again in the latter years
  2. Income inequality has increased substantially in the past 75 years
  3. The human capital stock has steadily increased throughout the 20th century
  4. Substantial improvements in technology across the 20th century
- The college wage premium is an equilibrium outcome relating relative demand and supply of high versus low skilled workers

Evolution of the college wage premium

Evolution of inequality

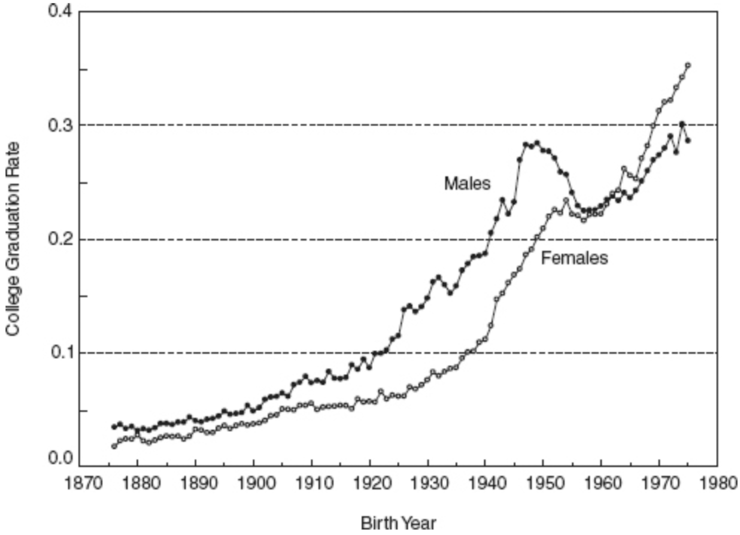


Evolution of college wage premium inequality (more recent years)

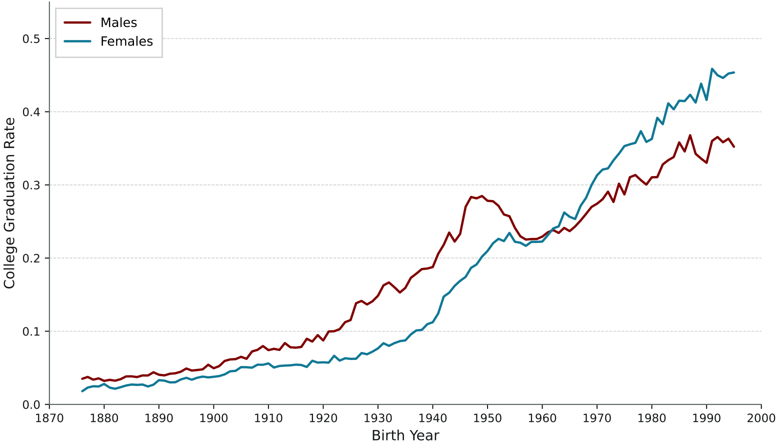


Source: Authors' calculations using March CPS 1964–2024. Composition-adjusted following Autor, Katz & Kearney (2008).

Evolution of college-educated human capital stock



Evolution of college-educated human capital stock (more recent years)



Source: Goldin and Katz (2008) for 1876–1975 birth cohorts; CPS MORG 2005–2018 for 1976–1993; ACS 2023 PUMS for 1994–1995. College graduation measured at approximately age 30.

The Canonical Model

- Two skill groups: high-skilled (H) and low-skilled (L)
- Closed-economy, perfectly competitive labor market
- Labor is supplied *inelastically* by both skill groups
- Workers in each skill group are assumed to be *imperfect substitutes*
- Total output (single composite good):

$$Y = \left(A_L L^{\frac{\sigma-1}{\sigma}} + A_H H^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

→ L and H are efficiency units of low- and high-skilled labor

→ A_L and A_H are factor-augmenting technologies for low-skill (L) and high-skill (H)

→ σ = elasticity of substitution between L and H

Wage Determination

- In a perfectly competitive market, each skill group is paid its marginal product
- Define the skill premium:

$$\omega = \frac{w_H}{w_L}$$

where competitive wages are:

$$w_L = \frac{\partial Y}{\partial L}, \quad w_H = \frac{\partial Y}{\partial H}$$

- First order conditions for each skill group imply:

$$\ln \omega = \ln \left(\frac{A_H}{A_L} \right) \frac{\sigma - 1}{\sigma} - \frac{1}{\sigma} \ln \left(\frac{H}{L} \right)$$

Model Insights

$$\ln \omega = \ln \left(\frac{A_H}{A_L} \right) \frac{\sigma - 1}{\sigma} - \frac{1}{\sigma} \ln \left(\frac{H}{L} \right)$$

- A rise in H/L (relative supply of high-skilled labor) lowers the skill premium
- A rise in (A_H/A_L) (skill-biased technological change) raises the skill premium
- The elasticity of substitution σ governs how strongly changes in supply or technology affect wages
- Technological progress need not benefit all equally:
  - → If $\sigma > 1$, an increase in A_H/A_L raises ω *unless* countered by a sufficiently large increase in H/L
  - → An increase in H/L need not lead to a decrease in ω if there is a coinciding and offsetting change in A_H/A_L
  - → There is an implicit “race” between education and technology

The model does remarkably well

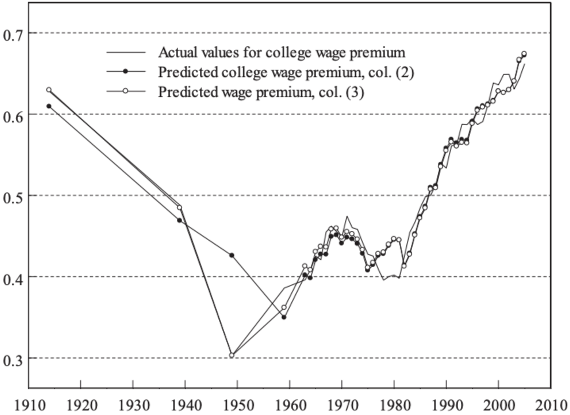


FIG. 2

Actual vs Predicted College wage Premium: 1915 to 2005.

Reproduced from Figure 2 of Goldin, C., Katz, L. F., 2007. The Race Between Education and Technology: The Evolution of US Educational Wage Differentials, 1890 to 2005. National Bureau of Economic Research.

The model does remarkably well

Table 1
Changes in the College Wage Premium and the Supply and Demand for College Educated Workers: 1915 to 2005 (100 × Annual Log Changes)

	Relative Wage	Relative Supply	Relative Demand ($\sigma_{SU} = 1.4$)	Relative Demand ($\sigma_{SU} = 1.64$)	Relative Demand ($\sigma_{SU} = 1.84$)
1915-40	-0.56	3.19	2.41	2.27	2.16
1940-50	-1.86	2.35	-0.25	-0.69	-1.06
1950-60	0.83	2.91	4.08	4.28	4.45
1960-70	0.69	2.55	3.52	3.69	3.83
1970-80	-0.74	4.99	3.95	3.77	3.62
1980-90	1.51	2.53	4.65	5.01	5.32
1990-2000	0.58	2.03	2.84	2.98	3.09
1990-2005	0.50	1.65	2.34	2.46	2.56
1940-60	-0.51	2.63	1.92	1.79	1.69
1960-80	-0.02	3.77	3.74	3.73	3.73
1980-2005	0.90	2.00	3.27	3.48	3.66
1915-2005	-0.02	2.87	2.83	2.83	2.82

Sources: The underlying data are presented in Appendix Table A8.1 and are derived from the 1915 Iowa State Census, 1940 to 2000 Census IPUMS, and 1980 to 2005 CPS MORG samples.

Some caveats

Without further adjustments to post 1992 demand-side adjustments, $\sigma = 1.61$ overpredicts the evolution of the wage premium. Did SBTC change or did σ change?

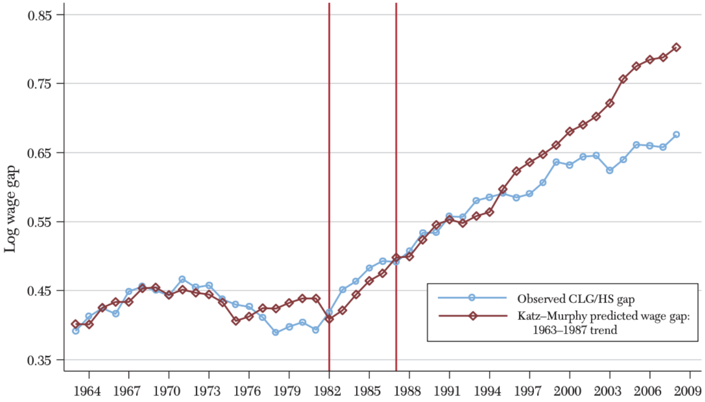
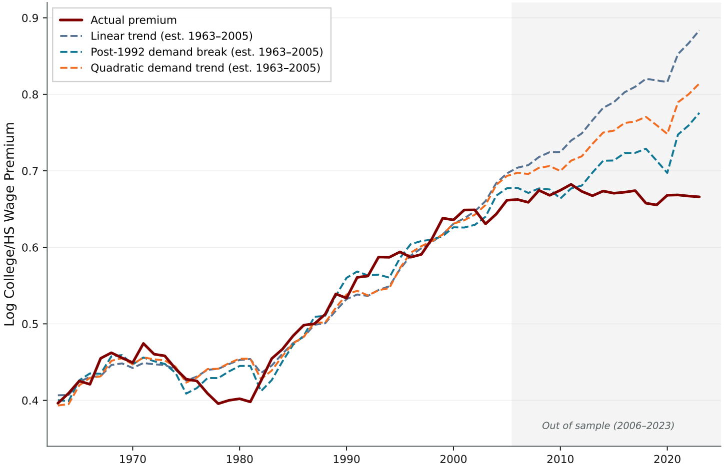


Figure 1. Katz–Murphy Prediction Model for the College–High School Wage Gap

Some caveats

Even the post 1992 demand-side adjustment only does well until the early 2010s



Source: Authors' calculations. Predicted values from Katz-Murphy (1992) supply-demand framework.
All specifications estimated on 1963–2005 and extrapolated out of sample to 2023.

Some caveats

Real wages dropping for some groups and at odds with canonical model predictions

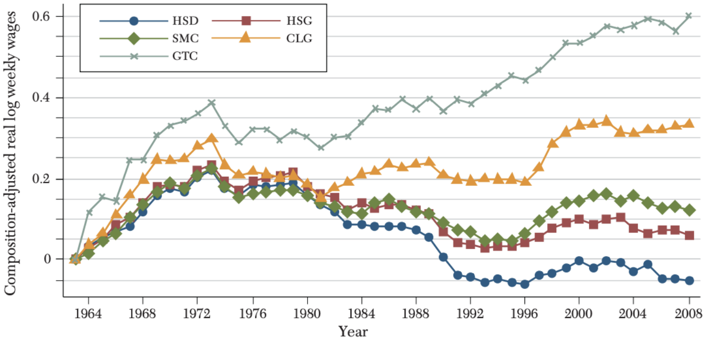


Figure 3. Real, Composition-Adjusted Log Weekly Wages for Full-Time Full-Year Workers 1963–2008 Males

Takeaways

Successes

- With "institutional" adjustments, fits long-term US data surprisingly well
- Explains changing college wage premiums since the early 20th century

Limitations

- Technology only enters as factor-augmenting (A_H, A_L)
- Real wage declines for some groups are hard to explain when $\sigma > 1$
- Does not capture task-based reallocation or direct labor displacement by machines $\rightarrow$ see Acemoglu and Autor 2011
- Ignores multidimensional skills (e.g., social, cognitive, technical) $\rightarrow$ see Deming 2017

Wrapping up the lecture

- We began with an overview of theory that motivates the study of human capital
- To start, there was a single "return" to education, which Mincerian models suggest is roughly 10/
- Schooling returns can be heterogeneous (for reasons unrelated to idiosyncratic differences in ability and costs)
    - → Angrist and Krueger and other earlier IV studies: estimates correspond to compliers shifting into more schooling; likely those who have larger returns and are shifted in when "costs" are lowered
    - → College selectivity and majors introduce heterogeneity in returns
- The college wage premium is not static and constantly responding to market forces, the so-called "race" between technology and education