America's Missing Entrepreneurs

Raj Chetty, Harvard and NBER
Gaia Dossi, LSE
Matthew Smith, US Treasury Department[1]
John Van Reenen, LSE, MIT, and NBER
Owen Zidar, Princeton and NBER
Eric Zwick, Chicago Booth and NBER

Bank of England, November 16, 2023

Preliminary

[1] This research was conducted while the author was an employee at the U.S. Department of the Treasury. The findings, interpretations, and conclusions expressed in this paper are entirely those of the author and do not necessarily reflect the views or the official positions of the U.S. Department of the Treasury. Any taxpayer data used in this research was kept in a secured Treasury or IRS data repository, and all results have been reviewed to ensure that no confidential information is disclosed.

Motivation

- A large share of growth in output, employment is driven by a few new firms (e.g., Apple, Amazon) [Davis Haltiwanger Schuh 1996, Haltiwanger Jarmin Miranda 2013]

- Can we increase economic output by expanding supply of entrepreneurs?

- Big differences in entrepreneurship rates by sex, parental income and race
  - Potential misallocation of talent [Hsieh Hurst Jones Klenow 2019, Bell Chetty Jaravel Petkova Van Reenen 2019]

- Large cross-sectional literature on entrepreneurs, but relatively little known about star founders. [Evans Leighton 1989, Hamilton 2000, Robb 2002, Lazear 2005, Fairlie Robb 2009, Hurst Pugsley 2011, Guzman Stern 2015,2020, Levine Rubinstein 2017, Azoulay Jones Kim Miranda 2020]
  - Small samples, lack of longitudinal data

This Paper

- Use population tax records on firms linked to individuals, tracking all ents in U.S. 2000–19

- Study determinants of entrepreneurship in four steps:

  1. Descriptive analysis of the characteristics of founders
  2. Examine 3 key causal mechanisms determining entrepreneurship:
     - Labor market **experience**
     - Liquid **wealth** constraints
     - Childhood **exposure** to entrepreneurship
  3. Use this "lifecyle" approach to analyze both overall entrepreneurship & reasons for lower levels in under-represented groups
  4. Investigate GE effects using a model that extends Hsieh Hurst Jones Klenow (2019)

Main Findings

1. Large, persistent disparities in entrepreneurship rates by gender, parental income

Main Findings

1. Large, persistent disparities in entrepreneurship rates by gender, parental income

2. Use alternative **causal** designs suggesting all three drivers matter
    - **Experience:** Local shocks into (high entrepreneurship) industries when young individuals enter labor market $\rightarrow$ entrepreneurship $\uparrow$
    - **Liquidity**: Cash windfalls for early employees at IPO firms $\rightarrow$ entrepreneurship $\uparrow$
    - **Exposure:** Kids moving to areas with more entrepreneurs $\rightarrow$ entrepreneurship $\uparrow$

Main Findings

1. Large, persistent disparities in entrepreneurship rates by gender, parental income
2. Use alternative **causal** designs suggesting all three drivers matter
   - **Experience:** Local shocks into (high entrepreneurship) industries when young individuals enter labor market $\rightarrow$ entrepreneurship $\uparrow$
   - **Liquidity**: Cash windfalls for early employees at IPO firms $\rightarrow$ entrepreneurship $\uparrow$
   - **Exposure:** Kids moving to areas with more entrepreneurs $\rightarrow$ entrepreneurship $\uparrow$
3. Quantitatively, labor market experience matters the most, liquid wealth the least, exposure in the middle in explaining overall entrepreneurship & between group gaps

Main Findings

1. Large, persistent disparities in entrepreneurship rates by gender, parental income
2. Use alternative **causal** designs suggesting all three drivers matter
   - **Experience:** Local shocks into (high entrepreneurship) industries when young individuals enter labor market $\rightarrow$ entrepreneurship $\uparrow$
   - **Liquidity**: Cash windfalls for early employees at IPO firms $\rightarrow$ entrepreneurship $\uparrow$
   - **Exposure:** Kids moving to areas with more entrepreneurs $\rightarrow$ entrepreneurship $\uparrow$
3. Quantitatively, labor market experience matters the most, liquid wealth the least, exposure in the middle in explaining overall entrepreneurship & between group gaps
4. Entrepreneurs from under-represented groups have lower rates of return

Main Findings

1. Large, persistent disparities in entrepreneurship rates by gender, parental income
2. Use alternative **causal** designs suggesting all three drivers matter
   - **Experience:** Local shocks into (high entrepreneurship) industries when young individuals enter labor market $\rightarrow$ entrepreneurship $\uparrow$
   - **Liquidity**: Cash windfalls for early employees at IPO firms $\rightarrow$ entrepreneurship $\uparrow$
   - **Exposure:** Kids moving to areas with more entrepreneurs $\rightarrow$ entrepreneurship $\uparrow$
3. Quantitatively, labor market experience matters the most, liquid wealth the least, exposure in the middle in explaining overall entrepreneurship & between group gaps
4. Entrepreneurs from under-represented groups have lower rates of return
5. New GE model of entrepreneurship as a (dynamic) occupational choice
   - Could double number of ents by removing barriers against female entrepreneurs (6x more female, but 20% fewer male ents)

Main Findings

1. Large, persistent disparities in entrepreneurship rates by gender, parental income
2. Use alternative **causal** designs suggesting all three drivers matter
   - **Experience:** Local shocks into (high entrepreneurship) industries when young individuals enter labor market $\rightarrow$ entrepreneurship $\uparrow$
   - **Liquidity**: Cash windfalls for early employees at IPO firms $\rightarrow$ entrepreneurship $\uparrow$
   - **Exposure:** Kids moving to areas with more entrepreneurs $\rightarrow$ entrepreneurship $\uparrow$
3. Quantitatively, labor market experience matters the most, liquid wealth the least, exposure in the middle in explaining overall entrepreneurship & between group gaps
4. Entrepreneurs from under-represented groups have lower rates of return
5. New GE model of entrepreneurship as a (dynamic) occupational choice
   - Could double number of ents by removing barriers against female entrepreneurs (6x more female, but 20% fewer male ents)
6. **Policy implications:** Our results point toward "pipeline problem"
   - Early labor market experience can explain meaningful portion of entry gaps
   - Closing gaps requires policy to target experience (and exposure), not just liquidity

1. Data

Assembling the Data

Firms: New firms from C-corp, S-corp, partnership tax filings (1120, 1120-S, 1065) 1998-2019

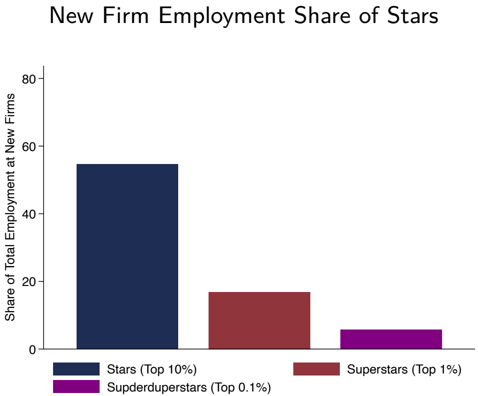
- Excludes unincorporated sole props; also exclude shells and spinoffs using W-2 data to isolate true new firms
- Founding date when first employ a non-owner employee (Census BDS definition)
- Validation: Number of new firms aligns with Census BDS ▸ Census Comparison

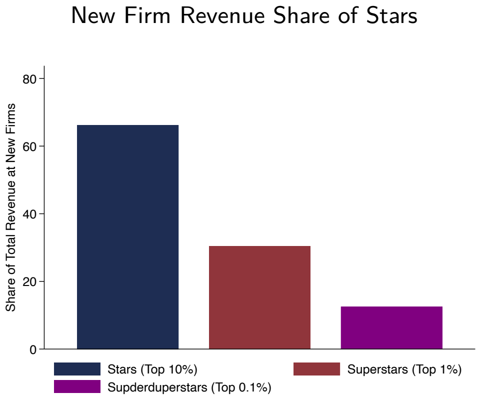
Founders: Owners in year firm started (K-1, 1125-E forms) [Smith Yagan Zidar Zwick 2019]

- For corps, exclude owners without W-2 income from the firm in first two years
- Validation: Use S-corps to test for Type I and II errors, outperforms W-2 method

Demographics: Parental background, childhood location, gender, race from individual level panel constructed in prior studies [Chetty Hendren Kline Saez 2014]

Importance of Star Entrepreneurs





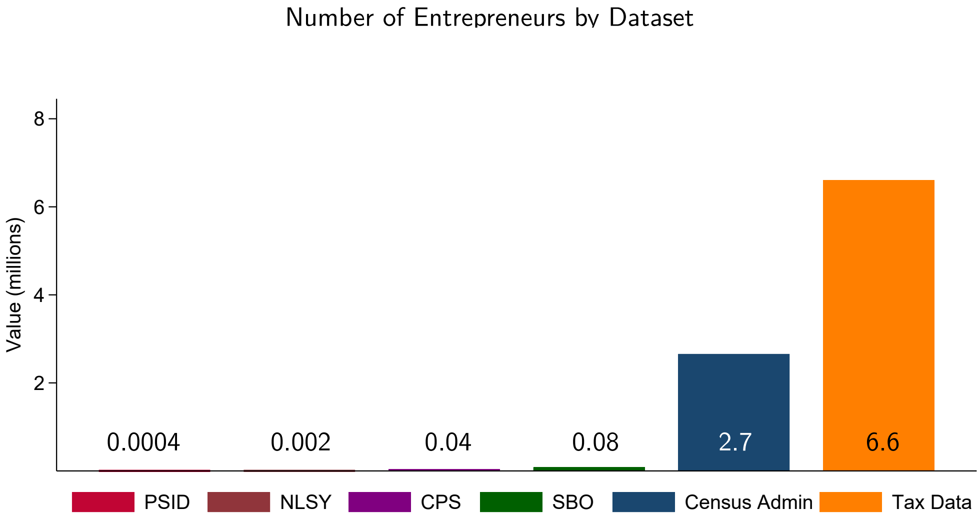
Type	Definition	Emp. Threshold (2015)	Rev. Threshold (2015)
Star	Top 10%	25	$1.7 million
Superstar	Top 1%	79	$8.7 million
Superduperstar	Top 0.1%	207	$35.1 million

Benchmarking

1. Founders are comparable to samples from other data on formal entrepreneurs
   - NLSY, CPS, PSID, SBO, Census Admin
   - Median age at founding: 40, Stars and superstars: 41–42
   - Female share: 29%, Stars and superstars: 20–25%
   - Median family income: $100K, Stars and superstars: $125K-$200K

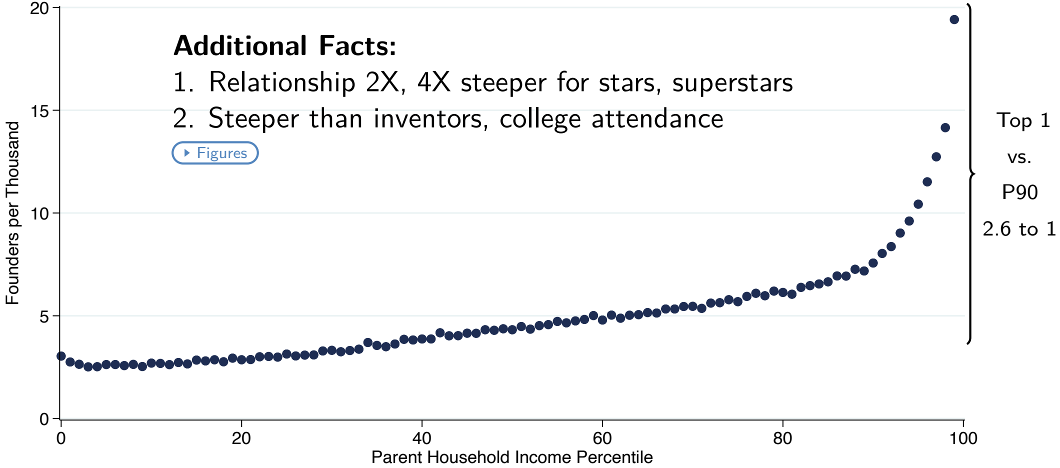
2. Improves sample size by multiple orders of magnitude

Benchmarking



2. Who are America’s Entrepreneurs?

Entrepreneurship Rates Lower if Born into a Low Income Family

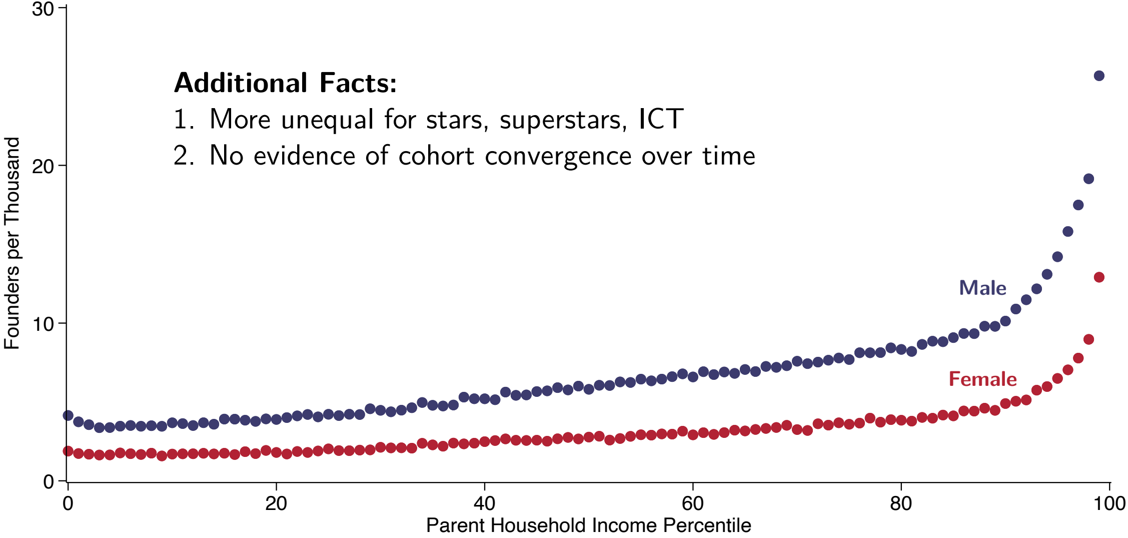


Additional Facts:

1. Relationship 2X, 4X steeper for stars, superstars
2. Steeper than inventors, college attendance

Figures

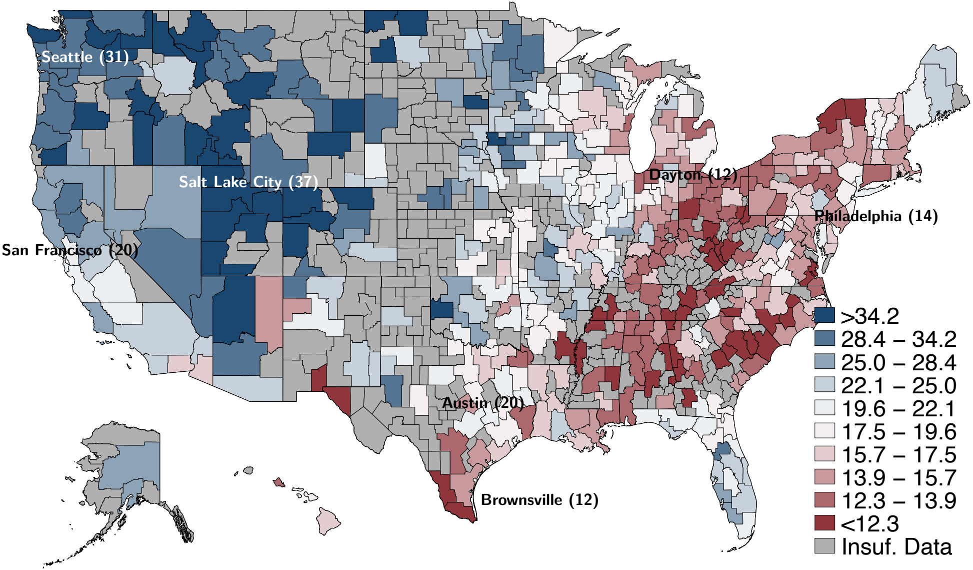
Entrepreneurship Rates by Family Income and Gender



Additional Facts:

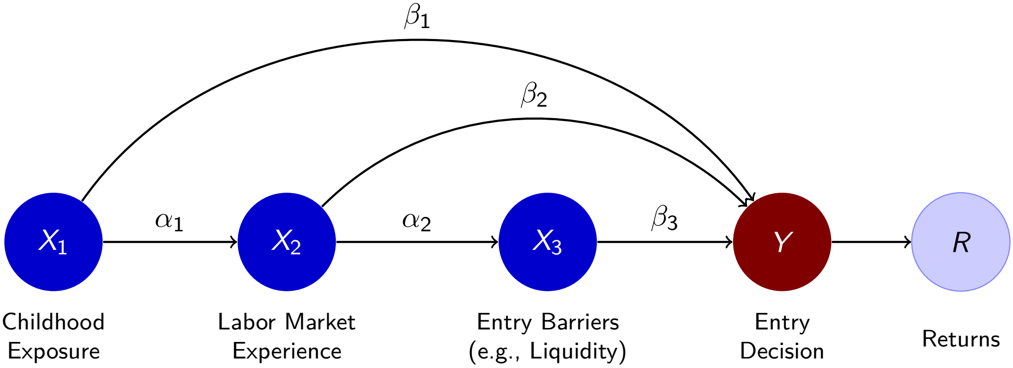
1. More unequal for stars, superstars, ICT
2. No evidence of cohort convergence over time

The Geographic Origins of Entrepreneurs



3. Determinants of Entrepreneurship

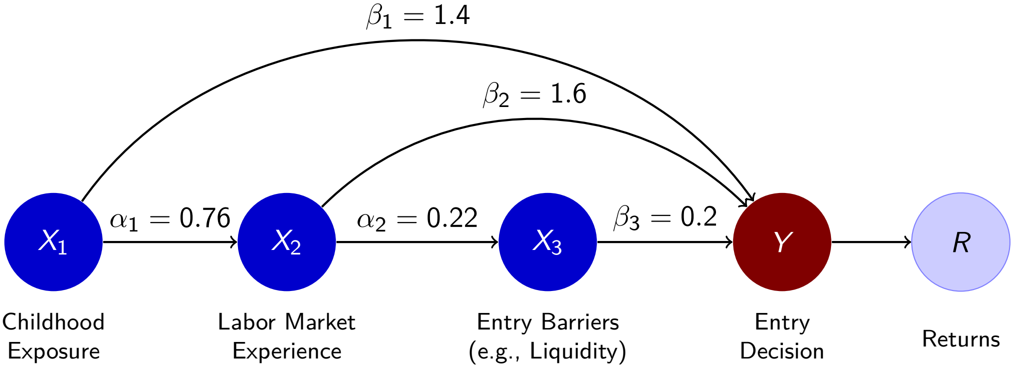
The Entrepreneurial Pipeline: Causal Graph



Note: βs give overall effect, αs are possible mechanisms.

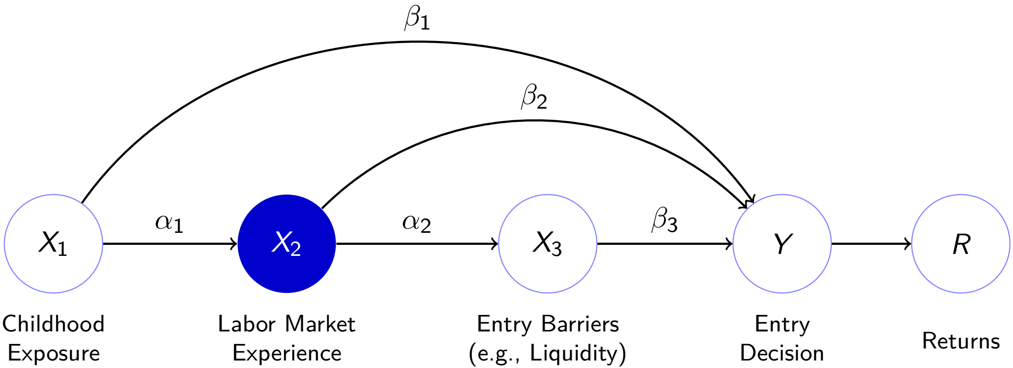
Question: How much does Y and ΔY (gaps) depend on X and ΔXs?

The Entrepreneurial Pipeline: Quantification



- Experience most important (wealth least); Exposure large, works via experience
- Experience accounts for material amount of URG gaps (at least a quarter)

The Entrepreneurial Pipeline I: Experience



Experience Effects from Prior Job

Goal: Estimate causal effect of industry experience on entry

- Labor market experience in particular sectors (e.g. software) when young generates entrepreneurial ideas, opportunity, networks
- Ideal experiment: randomly assign new workers to more entrepreneurial jobs

Experience Effects from Prior Job

Goal: Estimate causal effect of industry experience on entry

- Labor market experience in particular sectors (e.g. software) when young generates entrepreneurial ideas, opportunity, networks
- Ideal experiment: randomly assign new workers to more entrepreneurial jobs

Strategy: Compare workers with exposure to different industries

1. P(ent overall) = f(accumulated "entrepreneurial potential")
2. P(ent in industry n) = f(worker in industry n)
3. Use cohort design to isolate shocks to experience in an industry
4. Condition on factors driving demand for ents (narrow age range/current county)

Experience Effects from Prior Job

Goal: Estimate causal effect of industry experience on entry

- Labor market experience in particular sectors (e.g. software) when young generates entrepreneurial ideas, opportunity, networks
- Ideal experiment: randomly assign new workers to more entrepreneurial jobs

Strategy: Compare workers with exposure to different industries

1. P(ent overall) = f(accumulated "entrepreneurial potential")
2. P(ent in industry n) = f(worker in industry n)
3. Use cohort design to isolate shocks to experience in an industry
4. Condition on factors driving demand for ents (narrow age range/current county)

Mechanism: Test experience effects model [Hamilton 2000, Lazear 2005, Neal 1995]

Quantifying the Role of Labor Market Experience

Step 1: Instrument for experience in entrepreneurial jobs

1. Compute "entrepreneurship potential" E_n for each industry n
   - Share of workers in that industry that become entrepreneurs

Quantifying the Role of Labor Market Experience

Step 1: Instrument for experience in entrepreneurial jobs

1. Compute "entrepreneurship potential" E_n for each industry n
   - Share of workers in that industry that become entrepreneurs
2. Measure for each county-industry-year cell a worker share $\gamma_{c,n,y}$
   - Share of age 22 workers k in that county in industry n ($\frac{k_{c,n,y}}{k_{c,y}}$)

Quantifying the Role of Labor Market Experience

Step 1: Instrument for experience in entrepreneurial jobs

1. Compute "entrepreneurship potential" E_n for each industry n
   - Share of workers in that industry that become entrepreneurs
2. Measure for each county-industry-year cell a worker share $\gamma_{c,n,y}$
   - Share of age 22 workers k in that county in industry n ($\frac{k_{c,n,y}}{k_{c,y}}$)
3. Assign each person "entrepreneurship potential" from their age 22 county
   - Age 22 county-industry expected entrepreneurship rate $Z_{c(22),y} = \sum_n \gamma_{c(22),n,y} \cdot E_n$

Quantifying the Role of Labor Market Experience

Top Entrepreneurship Index Industries

	Industry (NAICS)	Ent Rate (%)
1	Residential building constr (2361)	4.02
2	Security contracts broker (5231)	3.50
3	Offices of real estate agents/brokers (5312)	3.12
4	Computer sys design/related svc (5415)	2.97
5	Personal care svc (8121)	2.86
6	Building equipment cntrctr (2382)	2.81
7	Building foundation/exterior cntrctr (2381)	2.76
8	Other information svc (5191)	2.66
9	Spectator sports (7112)	2.66
10	Nonresidential building constr (2362)	2.55
11	Sporting goods/musical instrument stores (4511)	2.34
12	Architectural/engineering svc (5413)	2.34
13	Cattle ranching/farming (1121)	2.32
14	Other specialty trade cntrctr (2389)	2.25
15	Accounting/bookkeeping svc (5412)	2.22

Quantifying the Role of Labor Market Experience

Bottom Entrepreneurship Index Industries

	Industry (NAICS)	Ent Rate (%)
1	Poultry/egg prodn (1123)	0.32
2	Animal slaughtering/processing (3116)	0.49
3	Child day care svc (6244)	0.58
4	Converted paper product mfg. (3222)	0.61
5	Investigation/security svc (5616)	0.66
6	Plastics product mfg. (3261)	0.66
7	Petroleum merch whlsl (4247)	0.69
8	Motor vehicle parts mfg. (3363)	0.72
9	Gambling industries (7132)	0.73
10	Home health care svc (6216)	0.74
11	Other food mfg. (3119)	0.77
12	Rubber product mfg. (3262)	0.77
13	Forging/stamping (3321)	0.77
14	Grocery stores (4451)	0.79
15	Employment svc (5613)	0.79

Quantifying the Role of Labor Market Experience

Research Design: Shocks to initial industry assignment (2 related approaches)

1. **Across-cohort**, within-county differences
   - e.g., 1979 versus 1980 cohort in Salt Lake City at age 22
   - Outcome is ever founding after 22
   - *Mechanism*: Shock in entry hiring in entrepreneurship-friendly industries in same county in adjacent years

2. **Specific labor market experience** $\rightarrow$ specific founding industry
   - e.g., 1979 versus 1980 cohort in Salt Lake City at 22 in NAICS 5415 (Computer Systems Design)
   - Outcome is founding after 22 in NAICS 5415
   - *Mechanism*: Early jobs enable particular founding trajectories

Quantifying the Role of Labor Market Experience

Step 2: Measuring experience in entrepreneurial jobs

1. Consider 1978-1982 birth cohorts.

Quantifying the Role of Labor Market Experience

Step 2: Measuring experience in entrepreneurial jobs

1. Consider 1978-1982 birth cohorts.

2. Use 1977 cohort to train a prediction model for entrepreneurship at age 35.
   - # years with wages $<$ \$5K, # years in each 4-digit NAICS for age 22-35
   - Partition # years variables into small vs. big firms, high vs. low relative wages
   - Occupation at age 35
   - $E_{i,1977} = f(\chi_{i,1}, ..., \chi_{i,M}) + \varepsilon_{i,1977}$

Quantifying the Role of Labor Market Experience

Step 2: Measuring experience in entrepreneurial jobs

1. Consider 1978-1982 birth cohorts.
2. Use 1977 cohort to train a prediction model for entrepreneurship at age 35.
    - # years with wages < $5K, # years in each 4-digit NAICS for age 22-35
    - Partition # years variables into small vs. big firms, high vs. low relative wages
    - Occupation at age 35
    - $E_{i,1977} = f(\chi_{i,1}, ..., \chi_{i,M}) + \varepsilon_{i,1977}$
3. Experience (X_2) $\equiv$ Predicted ent at age 35 for 1978-1982 cohorts.
    - $X_{2,i,c,y} = \hat{E}_{i,y}$ for $y \in \{1978 \ldots 1982\}$

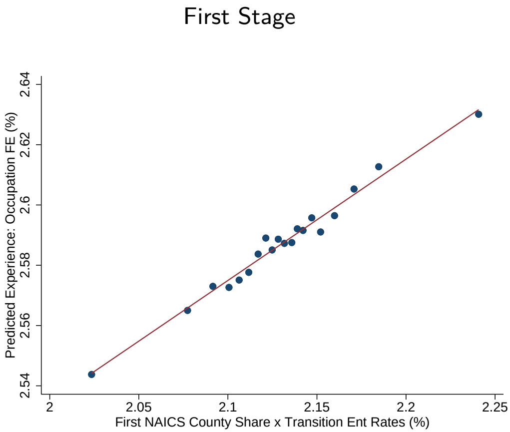
Quantifying the Role of Labor Market Experience

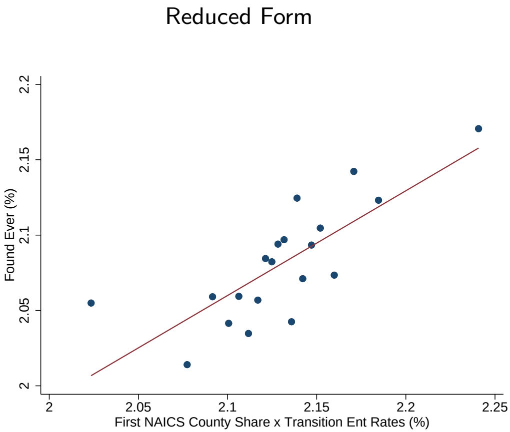
Step 2: Measuring experience in entrepreneurial jobs

1. Consider 1978-1982 birth cohorts.
2. Use 1977 cohort to train a prediction model for entrepreneurship at age 35.
   - # years with wages < $5K, # years in each 4-digit NAICS for age 22-35
   - Partition # years variables into small vs. big firms, high vs. low relative wages
   - Occupation at age 35
   - $E_{i,1977} = f(\chi_{i,1}, ..., \chi_{i,M}) + \varepsilon_{i,1977}$
3. Experience (X_2) $\equiv$ Predicted ent at age 35 for 1978-1982 cohorts.
   - $X_{2,i,c,y} = \hat{E}_{i,y}$ for $y \in \{1978 \ldots 1982\}$
4. Estimate causal effect of experience on level of entrepreneurship with 2SLS.
   - Use "entrepreneurship potential" as IV for experience
   - Use county (at 22) fixed effects & cohort fixed effects to isolate shocks (D)

2nd Stage: $Y_{i,c,y} = \beta_2 X_{2,i,c,y} + \alpha_{c(22)} + \alpha_y + \eta_{i,c,y}$ 1st Stage: $X_{2,i,c,y} = \delta Z_{c(22),y} + \alpha_{c(22)} + \alpha_y + \zeta_{i,c,y}$

The Causal Effect of Experience in Entrepreneurial Jobs





Result: Causal effect of industry experience on entrepreneurial entry

1. Design passes placebo tests with lagged and leading outcomes
2. Founding effect driven by entry well after the initial shock (post age 30)

The Effect of Early Labor Market Experience on Entrepreneurship

Dependent Variable	Experience	P(Ent)	P(Ent)
Age 22 County-Cohort Potential (IV)	0.3491	0.5683	
	(0.0131)	(0.0996)	
Experience			1.6278
			(0.2766)
Experience × Male			
Experience × Female			
Experience × ORG			
Experience × URG			
Experience × Par. Inc. Top 10			
Experience × Par. Inc. Bot 90			
Age 22 County + Cohort Fixed Effects	x	x	x
Age 22 County Clustered SE	x	x	x
Observations	10,074,353	10,074,353	10,074,353
F-statistic			715.04

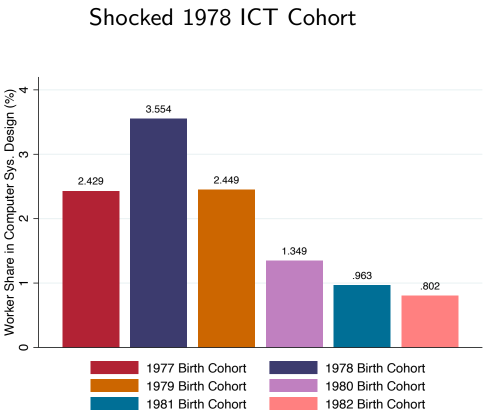
Notes: Mean entrepreneurship 2.6%, mean experience 2.1%. Implied elasticity = 2.0

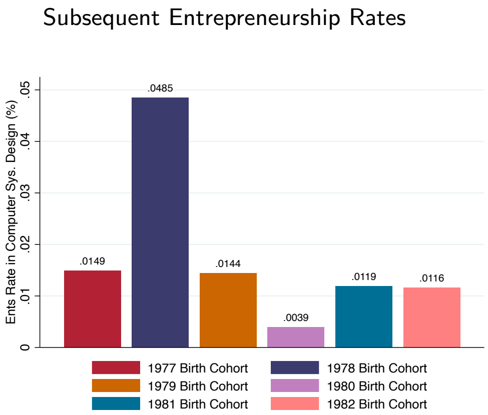
The Effect of Early Labor Market Experience on Entrepreneurship

Dependent Variable	Experience	P(Ent)	P(Ent)	P(Ent)	P(Ent)	P(Ent)
Age 22 County-Cohort Potential (IV)	0.3491	0.5683				
	(0.0131)	(0.0996)				
Experience			1.6278			
			(0.2766)			
Experience × Male				1.7740		
				(0.2672)		
Experience × Female				1.3311		
				(0.2923)		
Experience × ORG					2.2486	
					(0.4952)	
Experience × URG					1.9016	
					(0.5901)	
Experience × Par. Inc. Top 10						1.9885
						(0.3908)
Experience × Par. Inc. Bot 90						1.4224
						(0.4251)
Age 22 County + Cohort Fixed Effects	x	x	x	x	x	x
Age 22 County Clustered SE	x	x	x	x	x	x
Observations	10,074,353	10,074,353	10,074,353	10,074,353	4,763,500	5,662,260
F-statistic			715.04	345.77	165.04	165.60

Notes: Mean entrepreneurship 2.6%, mean experience 2.1%. Implied elasticity = 2.0

First Job Design: Illustrative Example from San Francisco CZ



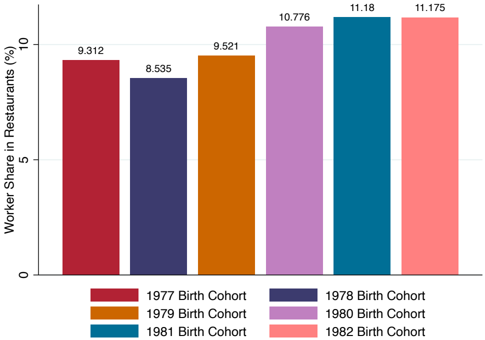


Challenge: Industry relation does not isolate supply effects of experience → Cohort design

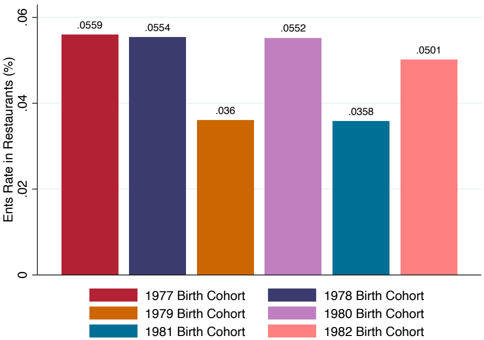
1. Compare 22-year-old workers in adjacent cohorts in a CZ
2. High employment shares in an industry as a proxy for **experience shocks**
3. Subsequent entry rate in same industry identifies **causal experience effects**

First Job Design: Illustrative Example from San Francisco CZ

Placebo 1978 Restaurant Cohort



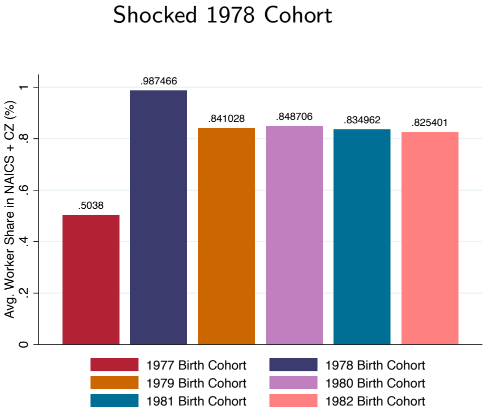
Subsequent Entrepreneurship Rates

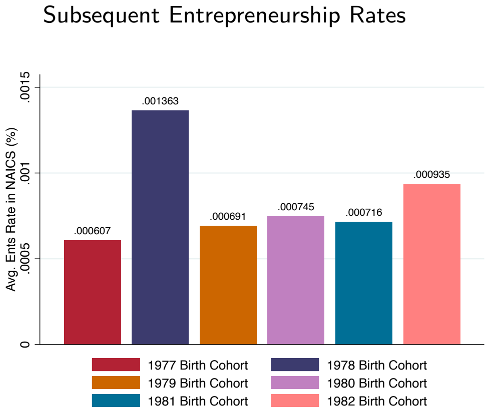


Challenge: Industry relation does not isolate supply effects of experience → Cohort design

1. Compare 22-year-old workers in adjacent cohorts in a CZ
2. High employment shares in an industry as a proxy for **experience shocks**
3. Subsequent entry rate in same industry identifies **causal experience effects**

Pooled First Job Shocks (Top Quartile)

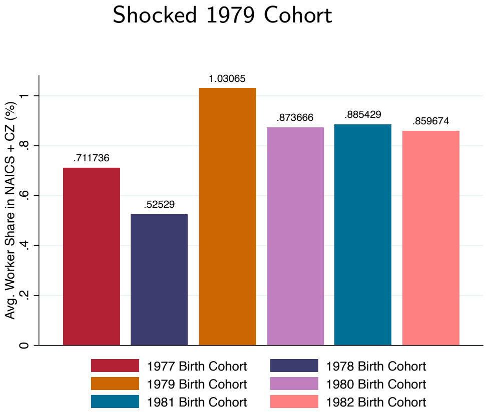


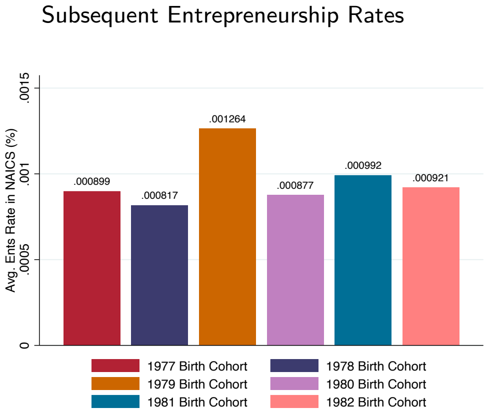


Challenge: Industry relation does not isolate supply effects of experience → Cohort design

1. Compare 22-year-old workers in adjacent cohorts in a CZ
2. High employment shares in an industry as a proxy for **experience shocks**
3. Subsequent entry rate in same industry identifies **causal experience effects**

Pooled First Job Shocks (Top Quartile)

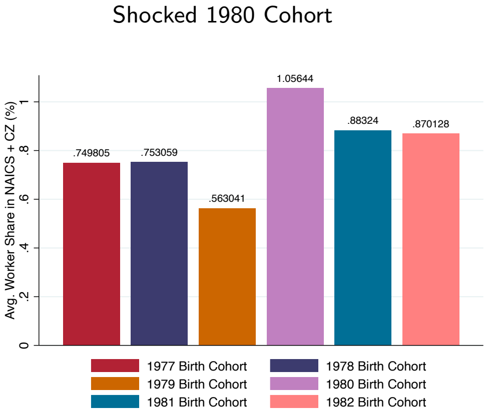


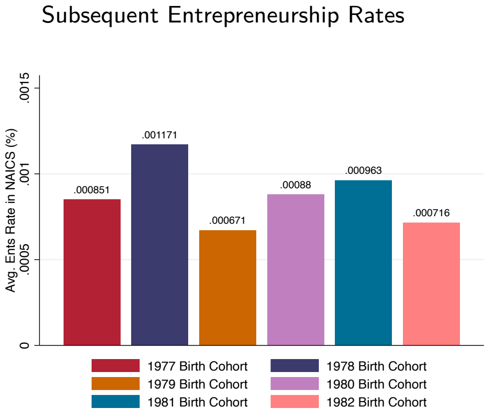


Challenge: Industry relation does not isolate supply effects of experience → Cohort design

1. Compare 22-year-old workers in adjacent cohorts in a CZ
2. High employment shares in an industry as a proxy for **experience shocks**
3. Subsequent entry rate in same industry identifies **causal experience effects**

Pooled First Job Shocks (Top Quartile)

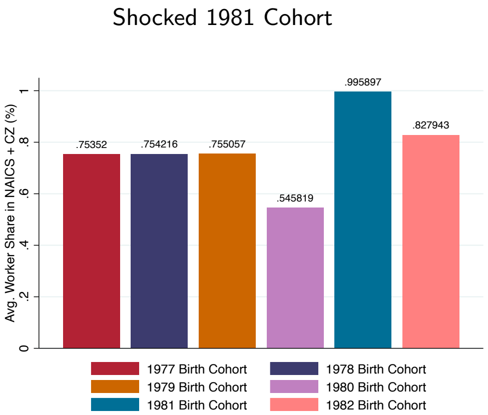


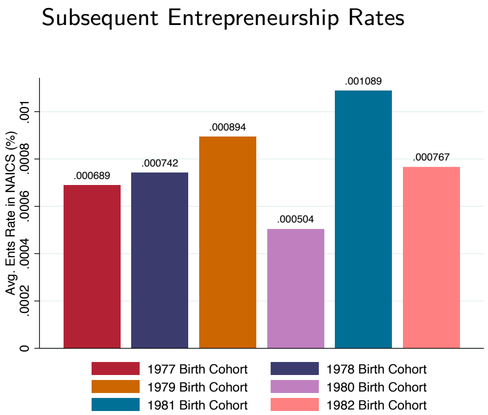


Challenge: Industry relation does not isolate supply effects of experience → Cohort design

1. Compare 22-year-old workers in adjacent cohorts in a CZ
2. High employment shares in an industry as a proxy for **experience shocks**
3. Subsequent entry rate in same industry identifies **causal experience effects**

Pooled First Job Shocks (Top Quartile)



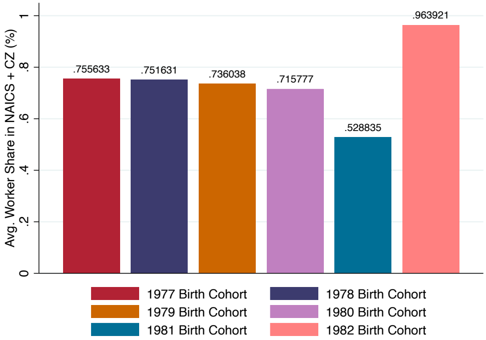


Challenge: Industry relation does not isolate supply effects of experience → Cohort design

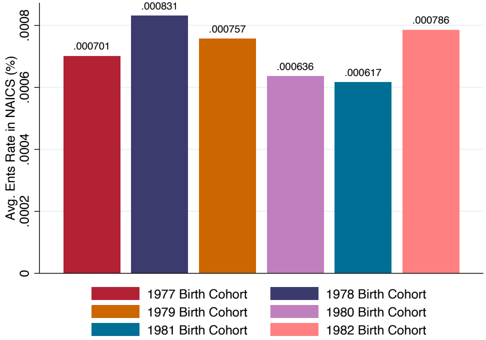
1. Compare 22-year-old workers in adjacent cohorts in a CZ
2. High employment shares in an industry as a proxy for **experience shocks**
3. Subsequent entry rate in same industry identifies **causal experience effects**

Pooled First Job Shocks (Top Quartile)

Shocked 1982 Cohort



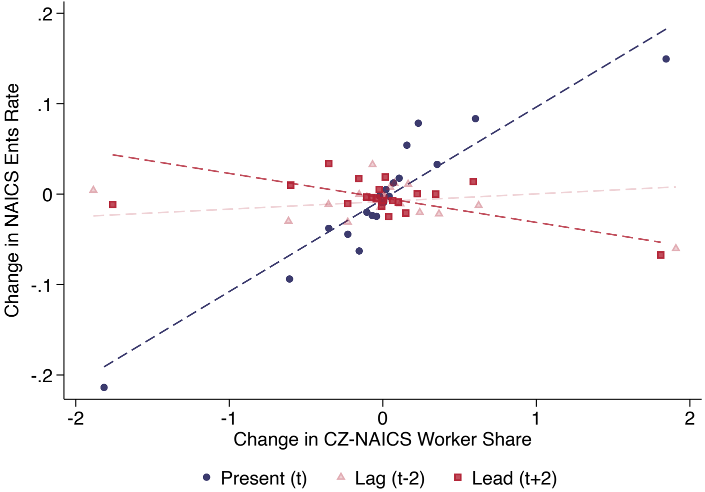
Subsequent Entrepreneurship Rates



Challenge: Industry relation does not isolate supply effects of experience → Cohort design

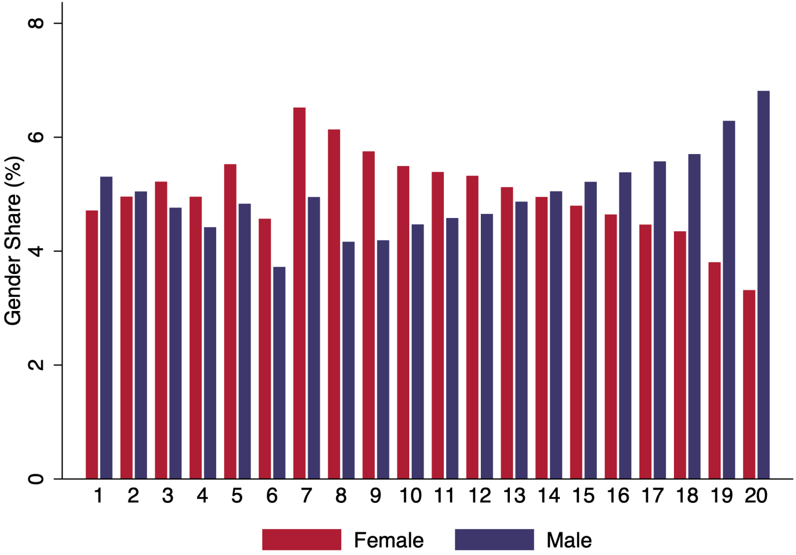
1. Compare 22-year-old workers in adjacent cohorts in a CZ
2. High employment shares in an industry as a proxy for **experience shocks**
3. Subsequent entry rate in same industry identifies **causal experience effects**

Pooled First Job Shocks



- Pooled diff-in-diff gives coefficient of 7.19×10^{-4} ($s.e. = 1.83 \times 10^{-4}$)
    - 1% increase in worker share in $k \rightarrow 0.72$ ents in k per 1,000 workers (60% of mean)

Gender Gaps in Labor Market Experience



Labor Market Experience Accounts for Large Share of Entrepreneurship Gap

Gap by sex:

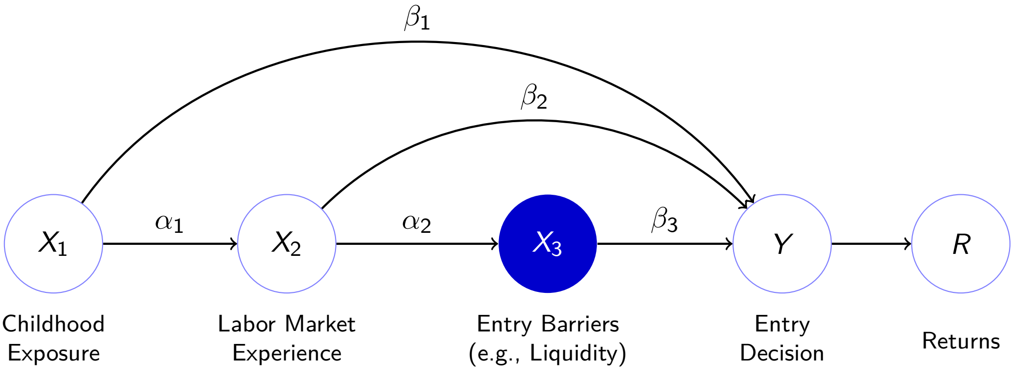
$$\begin{aligned}\beta_2 \times \Delta X_2/\Delta Y &= (1.33 \text{ or } 1.77) \times 0.31/1.73 \\ &= 24\% \text{ or } 32\%\end{aligned}$$

Gap by parental income:

$$\begin{aligned}\beta_2 \times \Delta X_2/\Delta Y &= (1.42 \text{ or } 1.99) \times 0.47/2.58 \\ &= 26\% \text{ or } 36\%\end{aligned}$$

Note: $X_2 \in \{$URG coefficient, ORG coefficient$\}$

The Entrepreneurial Pipeline II: Liquid Wealth



Liquidity from IPO Windfalls and Entrepreneurial Entry

Goal: Measure impact of large liquid wealth shocks on subsequent entry by group

- Outcomes include entry and proxies for returns conditional on entry
- Follow shock recipients several years after the shock
- Focus on population with relatively high baseline entry rates

Liquidity from IPO Windfalls and Entrepreneurial Entry

Goal: Measure impact of large liquid wealth shocks on subsequent entry by group

- Outcomes include entry and proxies for returns conditional on entry
- Follow shock recipients several years after the shock
- Focus on population with relatively high baseline entry rates

Strategy: Compare workers within IPO firms using pre-IPO wage rank

1. Wage rank $\approx$ Amount of stock held by non-founding, early workers
2. IPO $\rightarrow$ Illiquid stock becomes liquid, windfalls can be large
3. Condition on characteristics known to affect entry (earnings, age, geo)
4. Exclude top wage rank workers and any founding owners we can identify

Liquidity from IPO Windfalls and Entrepreneurial Entry

Goal: Measure impact of large liquid wealth shocks on subsequent entry by group

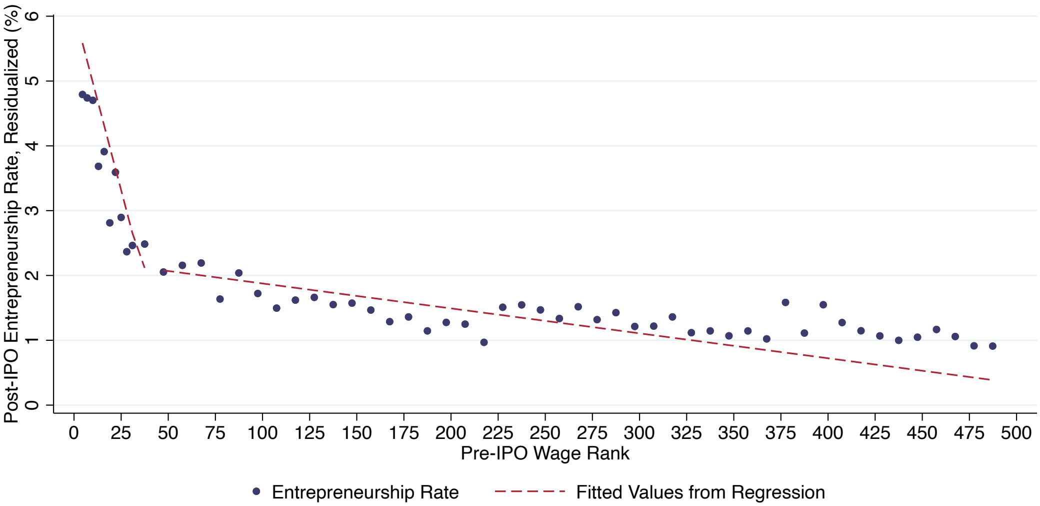
- Outcomes include entry and proxies for returns conditional on entry
- Follow shock recipients several years after the shock
- Focus on population with relatively high baseline entry rates

Strategy: Compare workers within IPO firms using pre-IPO wage rank

1. Wage rank $\approx$ Amount of stock held by non-founding, early workers
2. IPO $\rightarrow$ Illiquid stock becomes liquid, windfalls can be large
3. Condition on characteristics known to affect entry (earnings, age, geo)
4. Exclude top wage rank workers and any founding owners we can identify

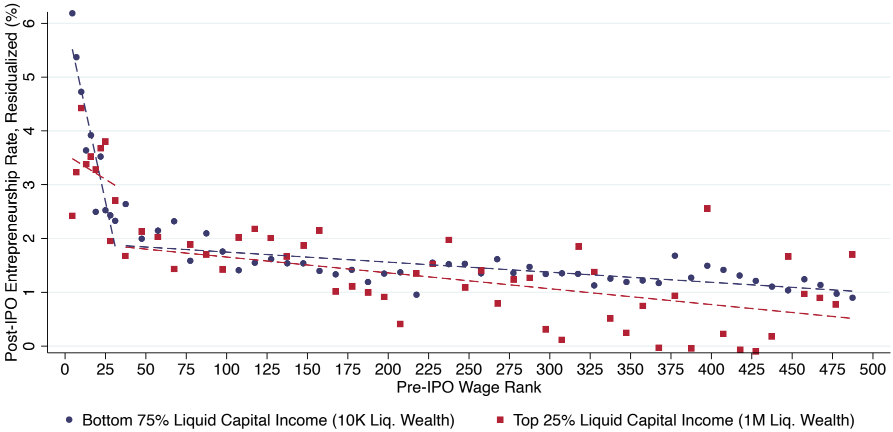
Mechanism: Test liquidity constraints entry model [Evans Jovanovic 1989]

IPO Cash Windfalls and Early Employee Entrepreneurship



Result: Larger windfall from IPO → Higher probability of becoming entrepreneur in the future

Effect Driven by Those with Less Pre-IPO Wealth

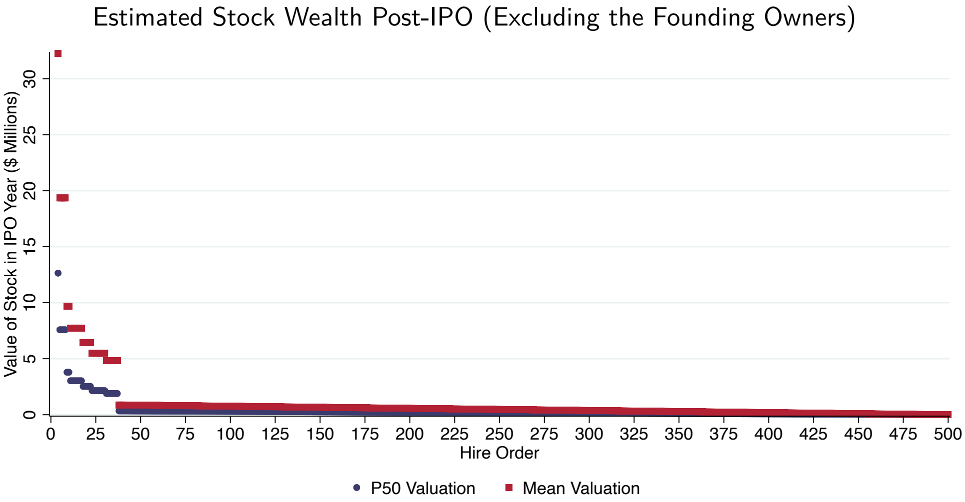


Result: Relationship goes to zero for workers with high liquid wealth prior to IPO

- Can reject equality of slopes among top workers with p-value $< .001$

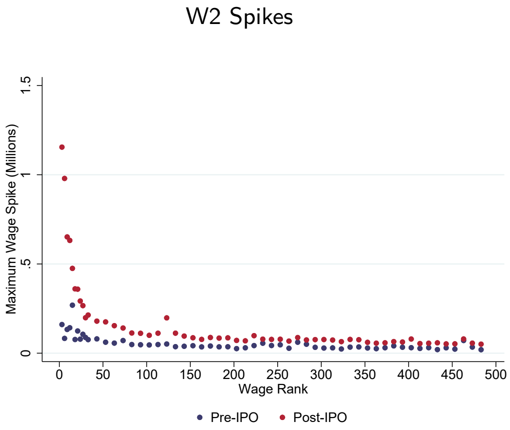
First Stage IPO Windfall

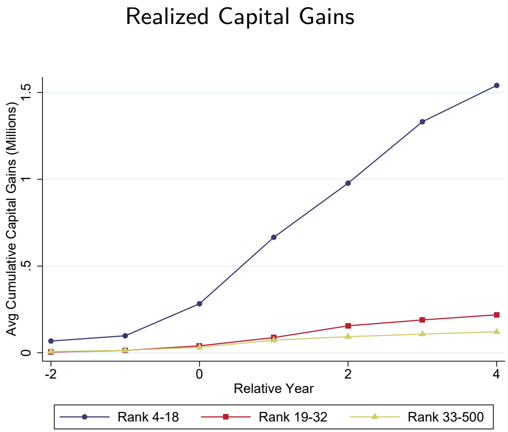
Data from Compustat and “The Holloway Guide to Equity Compensation”



Result: Nonlinear stock windfall pattern aligns with the reduced form ($\epsilon_{E,W} \approx 0.06$)

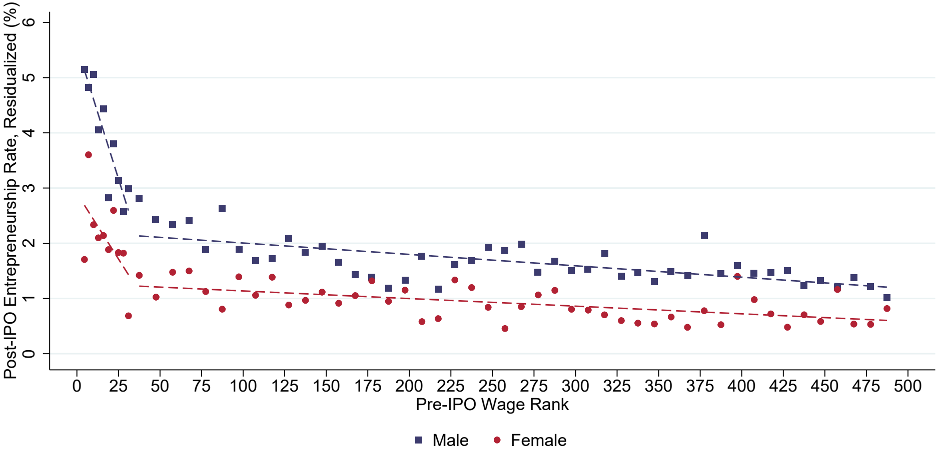
First Stage IPO Windfall: Alternative Approaches





Note: W2 spikes are deviations from mean W2 income → Proxy for stock option realizations

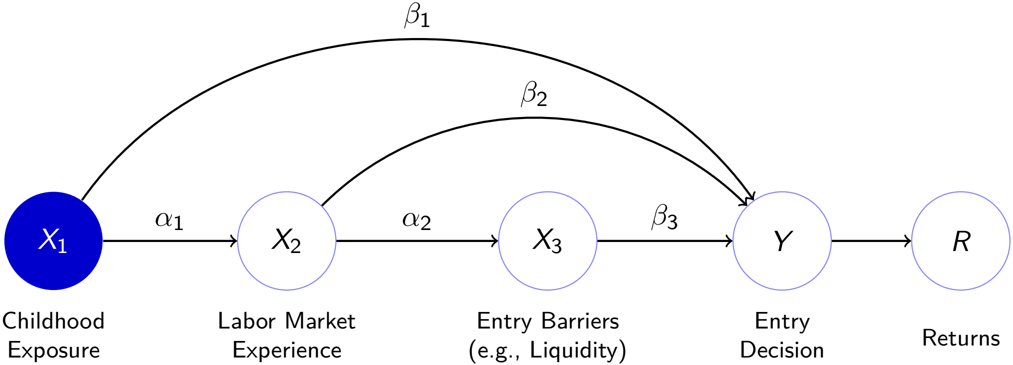
Heterogeneous Effects by Gender



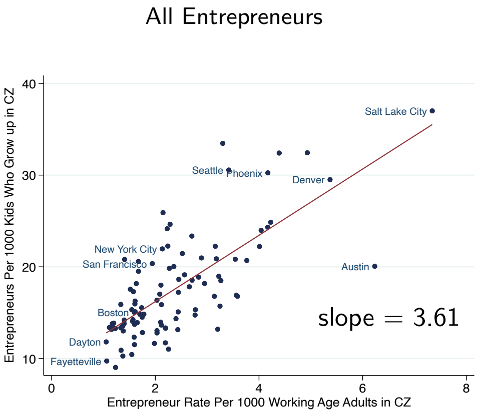
Result: Large gender gap remains at the top → Liquidity unlikely to be the key factor

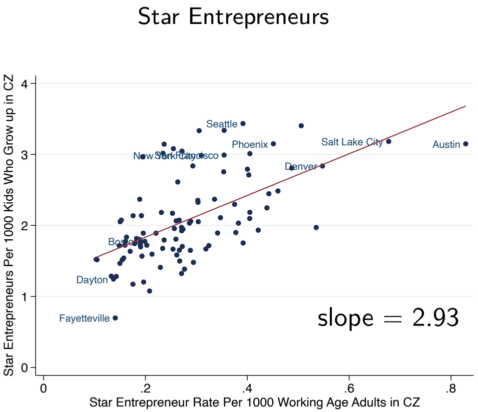
- Gender gap closes w/occupation FX, female share of top workers at firm → experience

The Entrepreneurial Pipeline III: Childhood exposure



Exposure Effects: Childhood Location and Future Entrepreneurship





Result: Childhood exposure to CZs with high entrepreneurship rates → greater future eship

- Causal effect using a movers research design
- Indirect effect through experience accounts for most of the exposure effect

Decomposition of Direct Effects

Panel A. Differences in Characteristics (1980 Cohort)				Panel B. Share of Outcome Difference Explained			
	Characteristics						
	Wealth ($K)	Experience (%)	Exposure (%)		Wealth	Experience	Exposure
Pooled Sample	59.2	2.62	0.095				
Men	61.2	2.86	0.095				
Women	57.5	2.55	0.094	*Men versus Women*			
Difference	3.7	0.31	0.001	P(ent)	0.0%	23.8%	0.1%
T10 Par. Inc.	211.7	3.22	0.129				
B90 Par. Inc.	40.3	2.75	0.122	*Top 10 versus Bottom 90 Parent Income*			
Difference	171.4	0.47	0.007	P(ent)	1.4%	25.8%	0.3%

Decomposition of Direct Effects

Panel A. Differences in Characteristics (1980 Cohort)				Panel B. Share of Outcome Difference Explained			
	Characteristics						
	Wealth ($K)	Experience (%)	Exposure (%)		Wealth	Experience	Exposure
Pooled Sample	59.2	2.62	0.095				
Men	61.2	2.86	0.095				
Women	57.5	2.55	0.094	*Men versus Women*			
Difference	3.7	0.31	0.001	P(ent)	0.0%	23.8%	0.1%
T10 Par. Inc.	211.7	3.22	0.129				
B90 Par. Inc.	40.3	2.75	0.122	*Top 10 versus Bottom 90 Parent Income*			
Difference	171.4	0.47	0.007	P(ent)	1.4%	25.8%	0.3%

Decomposition of Direct Effects

	Panel A. Differences in Characteristics (1980 Cohort)			Panel B. Share of Outcome Difference Explained			
	Characteristics						
	Wealth ($K)	Experience (%)	Exposure (%)		Wealth	Experience	Exposure
Pooled Sample	59.2	2.62	0.095				
Men	61.2	2.86	0.095				
Women	57.5	2.55	0.094	*Men versus Women*			
Difference	3.7	0.31	0.001	P(ent)	0.0%	23.8%	0.1%
T10 Par. Inc.	211.7	3.22	0.129				
B90 Par. Inc.	40.3	2.75	0.122	*Top 10 versus Bottom 90 Parent Income*			
Difference	171.4	0.47	0.007	P(ent)	1.4%	25.8%	0.3%

4. Estimating Individual Entrepreneurial Returns

Estimating the Person-Level "Returns" to Entrepreneurship

Goal: Measure average returns to entrepreneurship and differences across groups

- Restrict to first-time entrepreneurs and those with no prior business income
- Outcomes that may reflect firm exits
  - Today: total income, In progress: wealth, after-tax returns
- Follow entrepreneurs from $t = -5$ through $t = 8$

Estimating the Person-Level "Returns" to Entrepreneurship

Goal: Measure average returns to entrepreneurship and differences across groups

- Restrict to first-time entrepreneurs and those with no prior business income
- Outcomes that may reflect firm exits
  - Today: total income, In progress: wealth, after-tax returns
- Follow entrepreneurs from $t = -5$ through $t = 8$

Strategy: High dimensional 1-1 match between entrepreneurs and workers

1. Income history: AGI quantiles in $t \in \{-2, -3, -4\}$, with top 1% split into P99-99.9 and top 0.1%
2. NAICS 2-digit industry for employer in $t = -1$
3. Geo: Census region plus California
4. Age (3-year bucket), gender, single vs. joint filing status

Estimating the Person-Level "Returns" to Entrepreneurship

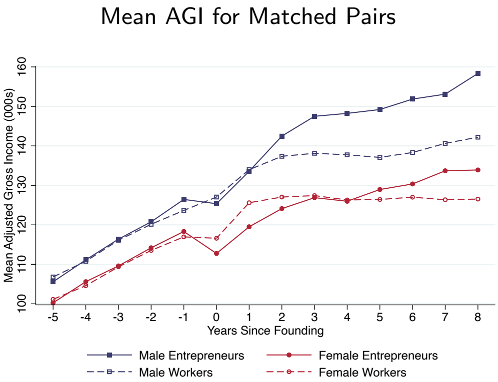
Goal: Measure average returns to entrepreneurship and differences across groups

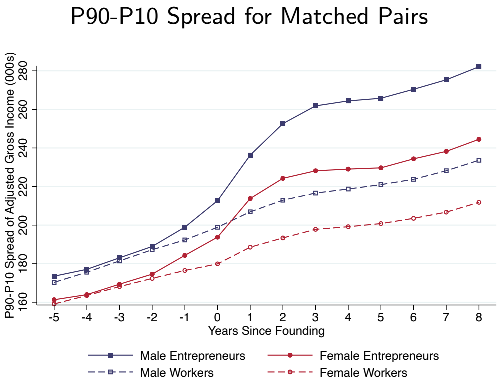
- Restrict to first-time entrepreneurs and those with no prior business income
- Outcomes that may reflect firm exits
  - Today: total income, In progress: wealth, after-tax returns
- Follow entrepreneurs from $t = -5$ through $t = 8$

Strategy: High dimensional 1-1 match between entrepreneurs and workers

1. Income history: AGI quantiles in $t \in \{-2, -3, -4\}$, with top 1% split into P99-99.9 and top 0.1%
2. NAICS 2-digit industry for employer in $t = -1$
3. Geo: Census region plus California
4. Age (3-year bucket), gender, single vs. joint filing status

The Distribution of Financial Returns

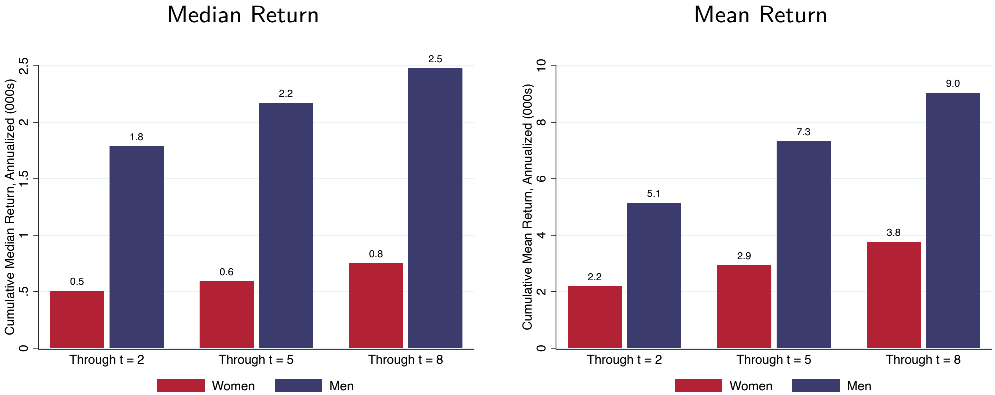




Result: Positive returns on average after 3-4 years

- Confirms findings on incorporated entrepreneurs [Levine Rubinstein 2017]
- Higher means come with higher variance → risk tolerance [Hall Woodward 2010]
- **Key:** Marginal returns in experience design indistinguishable from average returns

Return for Females is Half that for Males

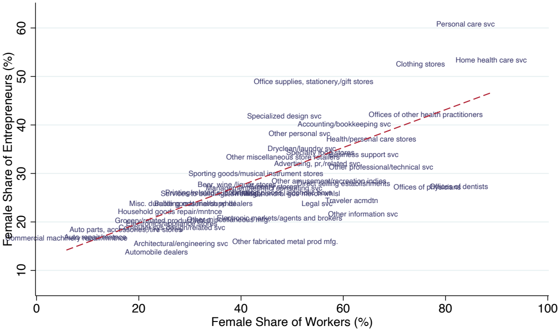


Result: Female entrepreneurs earn lower returns → Inconsistent with simple Roy model logic

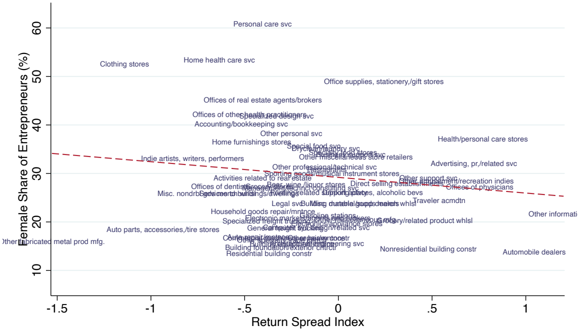
- Other factors/frictions cause women to earn less after entry
- Confirms SBO findings on incorporated+unincorporated ents (e.g., Robb 2002)
- Similar results for race URGs andlow parent income kids
- Conditioning on founding industry closes ≈1/3 of the gender gap

Experience Relation Appears More Important than Return Differences

Female Share of Workforce



Average Return Gap, Men vs. Women



Takeaways: Female share of workforce narrows the gender gap; returns to female ents does not

5. Modeling the Costs of Missing Entrepreneurs

GE Model of Entrepreneurship, Discrimination, and Talent Misallocation

Today: The impact on output and distribution from shrinking barriers to female ents

- Builds on Hsieh Hurst Jones Klenow (2019, HHJK) details
  - Roy model of sector choice based on heterogeneous talent (or preferences)
  - Agents make endogenous sector-specific human capital decisions (period 0 pre-labor market), then enter for three periods (young, middle-aged and old)

GE Model of Entrepreneurship, Discrimination, and Talent Misallocation

Today: The impact on output and distribution from shrinking barriers to female ents

- Builds on Hsieh Hurst Jones Klenow (2019, HHJK) details
  - Roy model of sector choice based on heterogeneous talent (or preferences)
  - Agents make endogenous sector-specific human capital decisions (period 0 pre-labor market), then enter for three periods (young, middle-aged and old)
- Two types of sector-specific frictions (τ) against URG (focus on women):
  - Human capital (τ^h) entry cost (e.g. lower exposure)
  - Labor market (τ^w) paid each period (e.g. income discrimination)

GE Model of Entrepreneurship, Discrimination, and Talent Misallocation

Today: The impact on output and distribution from shrinking barriers to female ents

- Builds on Hsieh Hurst Jones Klenow (2019, HHJK) details
  - Roy model of sector choice based on heterogeneous talent (or preferences)
  - Agents make endogenous sector-specific human capital decisions (period 0 pre-labor market), then enter for three periods (young, middle-aged and old)

- Two types of sector-specific frictions (τ) against URG (focus on women):
  - Human capital (τ^h) entry cost (e.g. lower exposure)
  - Labor market (τ^w) paid each period (e.g. income discrimination)

- $\tau \uparrow$ implies lower proportion of women in sector and larger gender income gap (although depends on selection)

GE Model of Entrepreneurship, Discrimination, and Talent Misallocation

Today: The impact on output and distribution from shrinking barriers to female ents

- Builds on Hsieh Hurst Jones Klenow (2019, HHJK) details
  - Roy model of sector choice based on heterogeneous talent (or preferences)
  - Agents make endogenous sector-specific human capital decisions (period 0 pre-labor market), then enter for three periods (young, middle-aged and old)
- Two types of sector-specific frictions (τ) against URG (focus on women):
  - Human capital (τ^h) entry cost (e.g. lower exposure)
  - Labor market (τ^w) paid each period (e.g. income discrimination)
- $\tau \uparrow$ implies lower proportion of women in sector and larger gender income gap (although depends on selection)
- Two extensions to allow for entrepreneurship:
  - **Extension I:** Entrepreneurship as another career with different τ's
  - **Extension II:** In addition, agents can start in one sector & switch mid-career to become entrepreneurs (at a sector specific transition cost)

Policy Counterfactual: Removing Barriers for Women

Panel A: % of entrepreneurs					
	Baseline	Remove τs, entrepreneurs		Remove τs, all sectors	
	(1)	(2)	(3)	(4)	(5)
		%	Δ p.p.	%	Δ p.p.
Men	1.6%	1.3%	-0.3%	1.5%	-0.1%
Women	0.5%	2.8%	2.3%	1.6%	1.1%
Total	**2.1%**	**4.2%**	**2.1%**	**3.1%**	**1.0%**

Panel B: GDP p.c.			
	Baseline	Remove τs, entrepreneurs	Remove τs, all sectors
	(1)	(2)	(3)
		% change	% change
Labor earnings per worker	$53,390	2.7%	30%

Notes. This table reports the results of a counterfactual estimation for the last period of the model where τ^h and τ^w are removed. Income moments come from the IRS. Employment shares come from the CPS. GDP p.c. is computed as total labor income divided by total number of workers (in 2017 USD).

Conclusions

Conclusions

We find:

1. Large, persistent disparities in entrepreneurship rates by gender and parental income
2. Early labor market experience matters more than liquidity and exposure in determining number of entrepreneurs and explaining gaps
3. Could be substantial increases in entrepreneurship (& output) from reducing frictions

Policy implications:

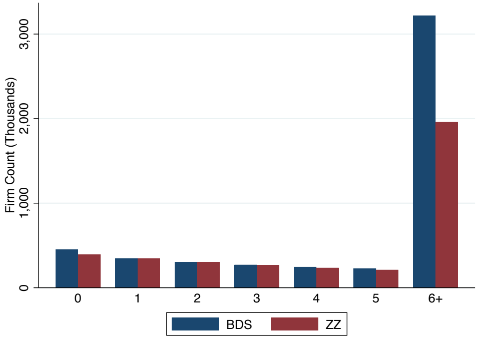
- Closing gaps requires policy to target experience and exposure, not just liquidity
- Earlier interventions focused on occupational choice/labor markets forces
- Finance targeted to URGs paired with mentoring/incubators

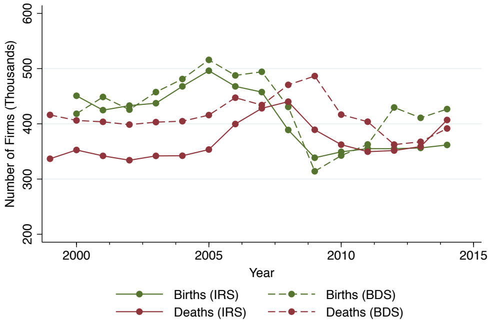
Thanks!

Census Comparison

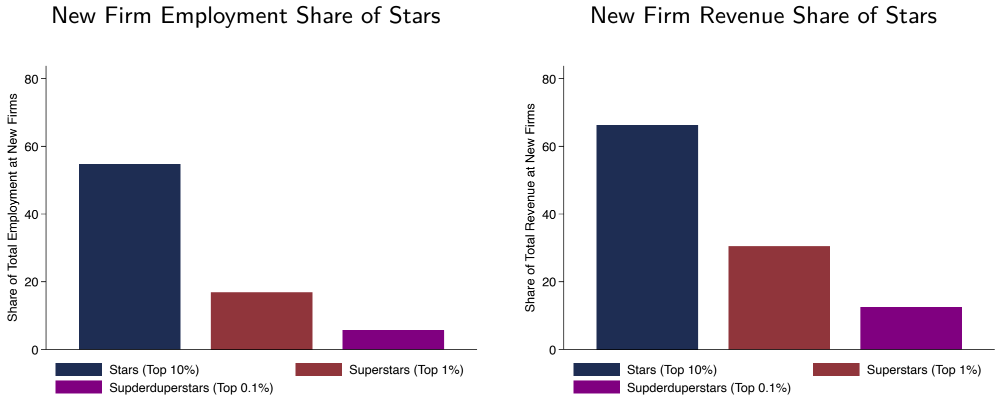
Back

Figure: Comparing IRS to Census Data





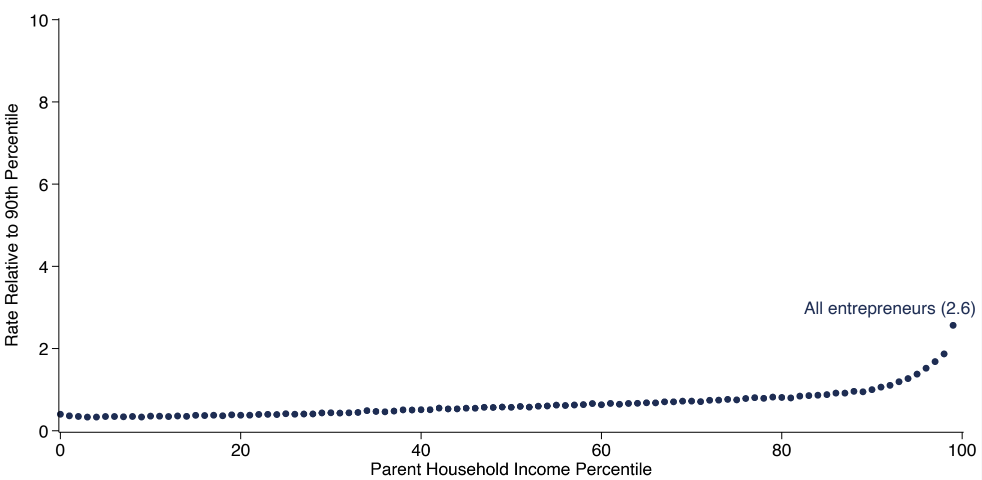
Importance of Star Entrepreneurs



Type	Definition	Emp. Threshold (2015)	Rev. Threshold (2015)
Star	Top 10%	25	\$1.7 million
Superstar	Top 1%	79	\$8.7 million
Superduperstar	Top 0.1%	207	\$35.1 million

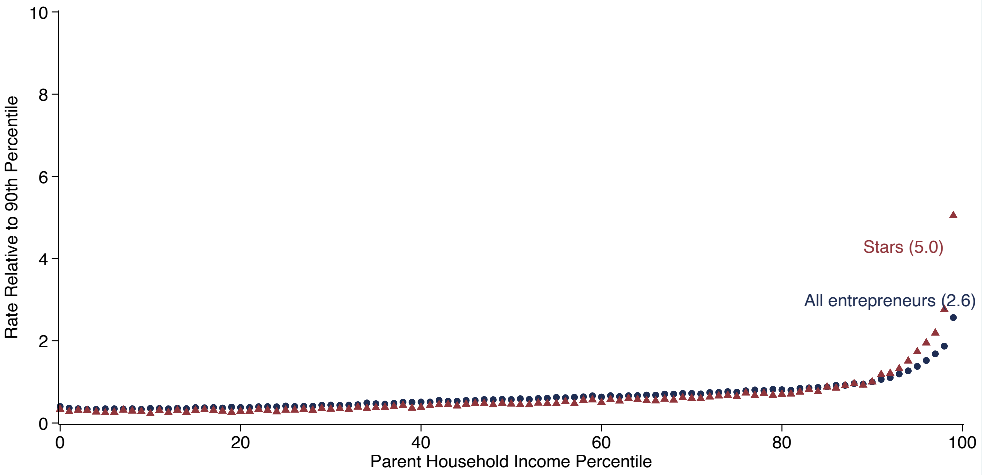
Entrepreneurship Rates Lower if Born into a Low Income Family

Back



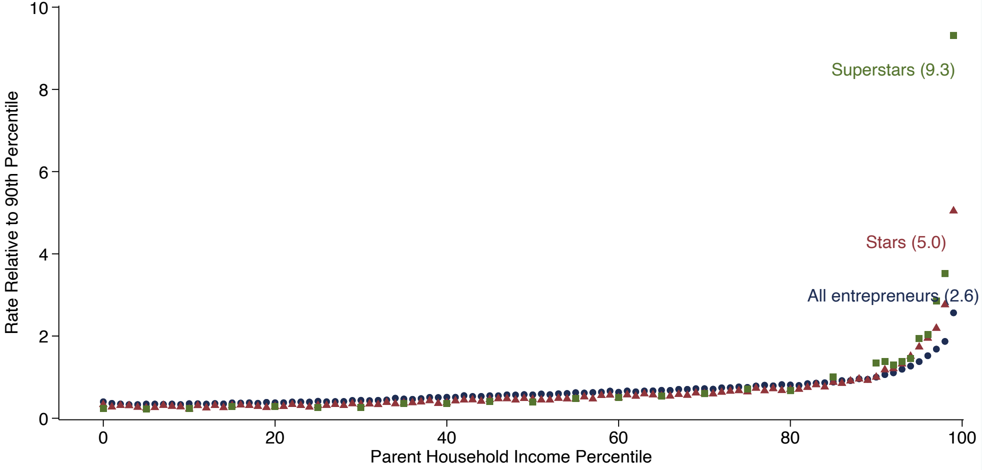
Entrepreneurship Rates Lower if Born into a Low Income Family

Back



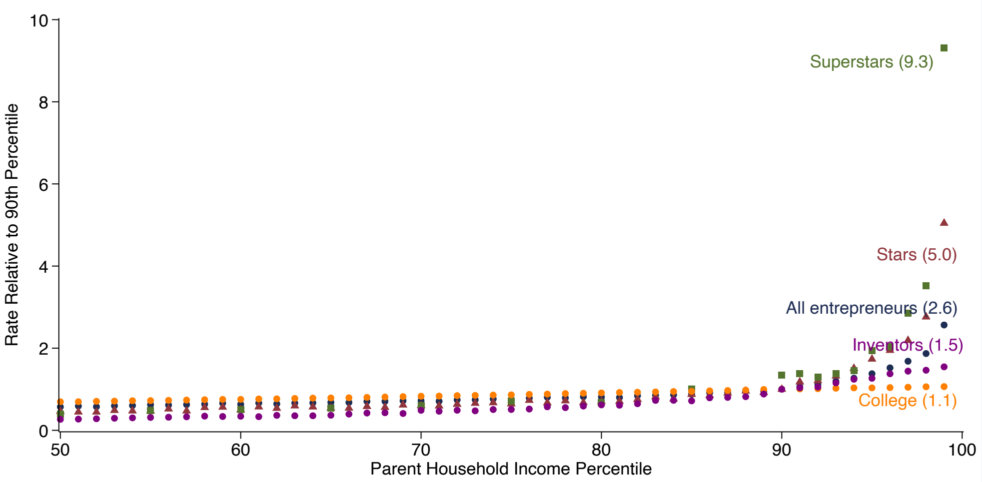
Entrepreneurship Rates Lower if Born into a Low Income Family

Back



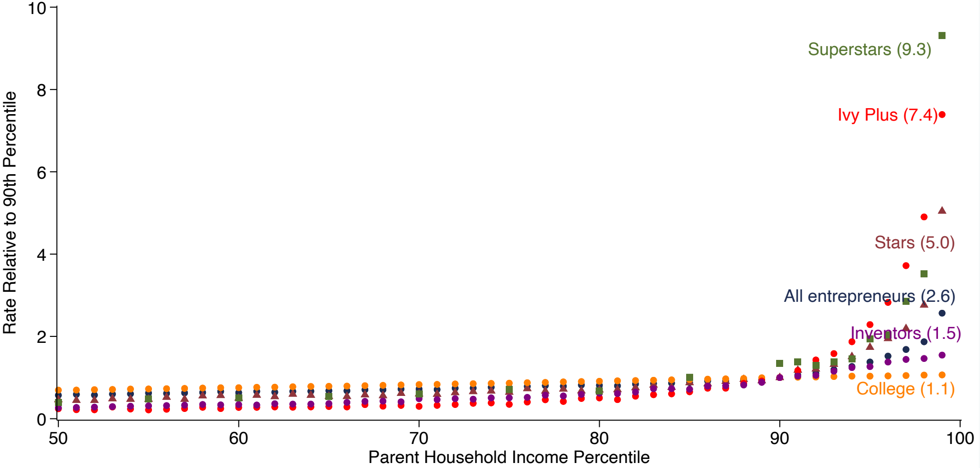
Entrepreneurship Rates Lower if Born into a Low Income Family

Back



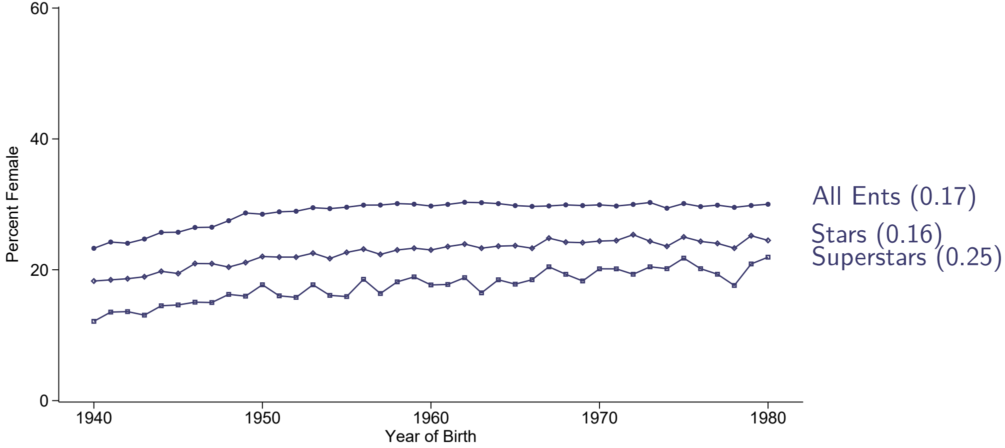
Entrepreneurship Rates Lower if Born into a Low Income Family

Back



Female Entrepreneurship Rates are Stagnant

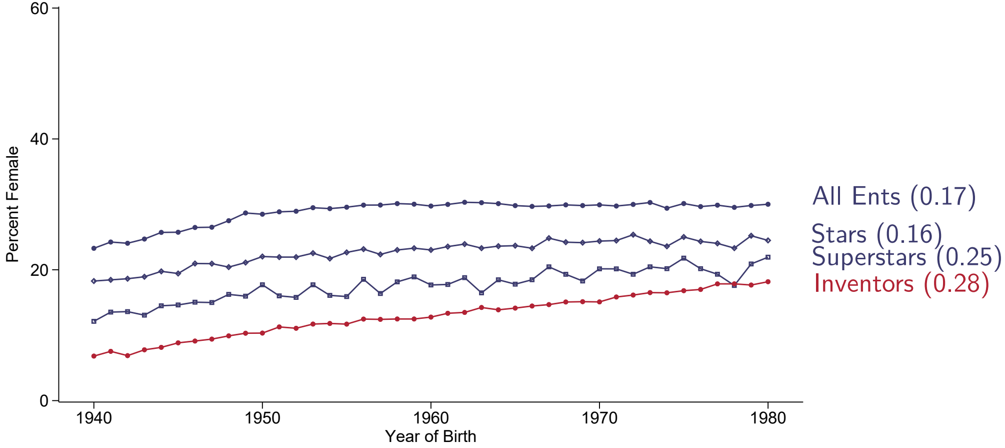
Back



Note: Growth rate in percentage points per year in parentheses.

Female Entrepreneurship Rates are Stagnant

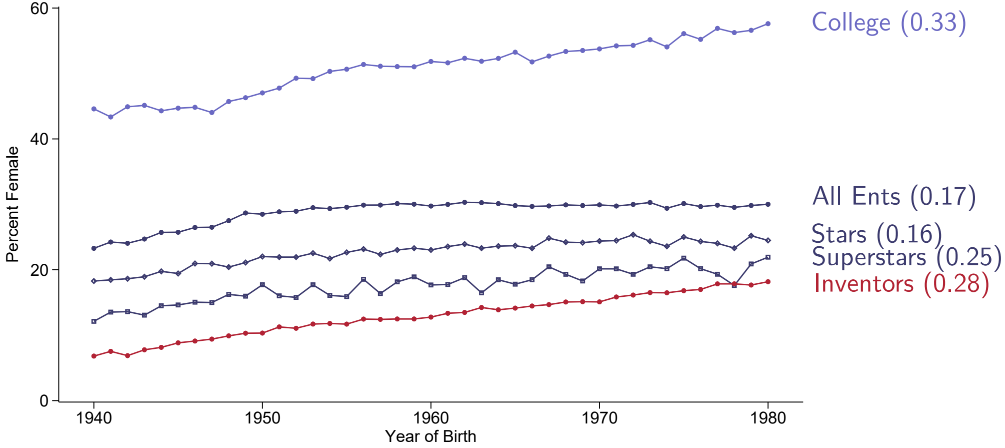
◂ Back



Note: Growth rate in percentage points per year in parentheses.

Female Entrepreneurship Rates are Stagnant

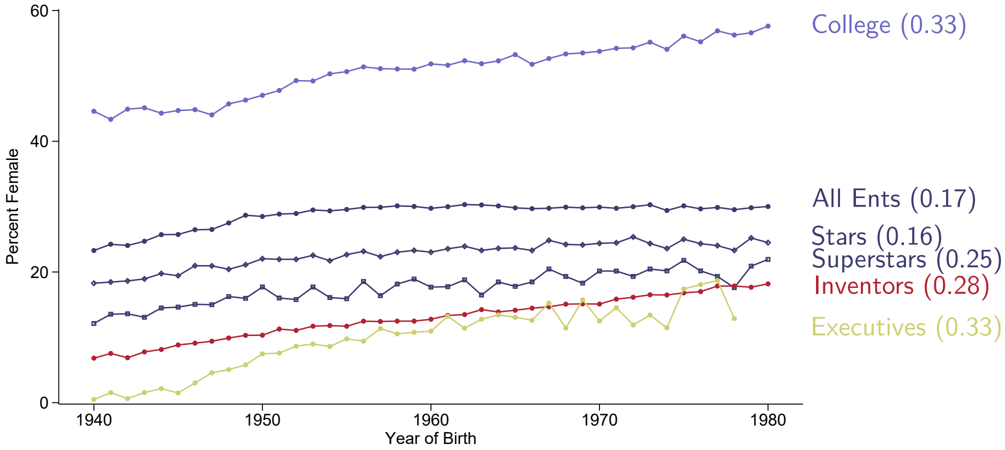
◂ Back



Note: Growth rate in percentage points per year in parentheses.

Female Entrepreneurship Rates are Stagnant

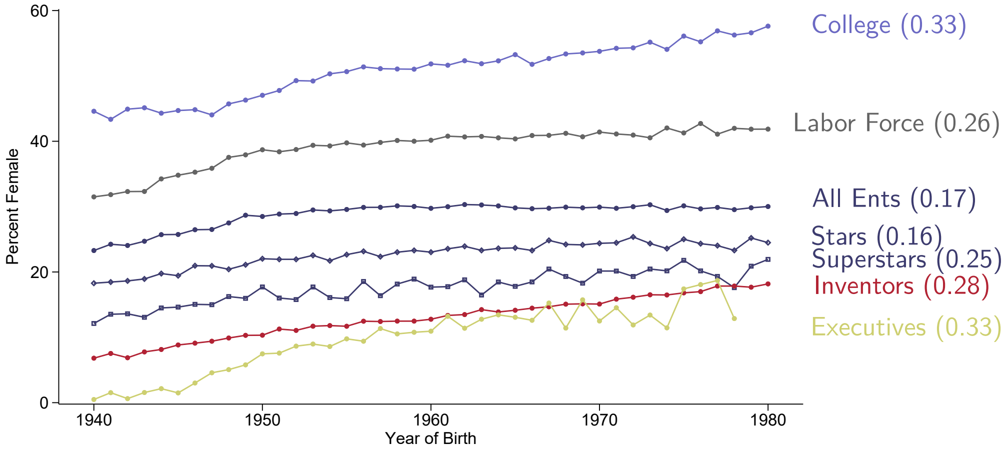
Back



Note: Growth rate in percentage points per year in parentheses.

Female Entrepreneurship Rates are Stagnant

Back



Note: Growth rate in percentage points per year in parentheses.

The Effect of Early Labor Market Experience on Entrepreneurship

Dependent Variable	Experience	P(Ent)	P(Ent)
Age 22 County-Cohort Potential (IV)	0.3491	0.5683	
	(0.0131)	(0.0996)	
Experience			1.6278
			(0.2766)
Experience × Male			
Experience × Female			
Experience × ORG			
Experience × URG			
Experience × Par. Inc. Top 10			
Experience × Par. Inc. Bot 90			
Age 22 County + Cohort Fixed Effects	x	x	x
Age 22 County Clustered SE	x	x	x
Observations	10,074,353	10,074,353	10,074,353
F-statistic			715.04

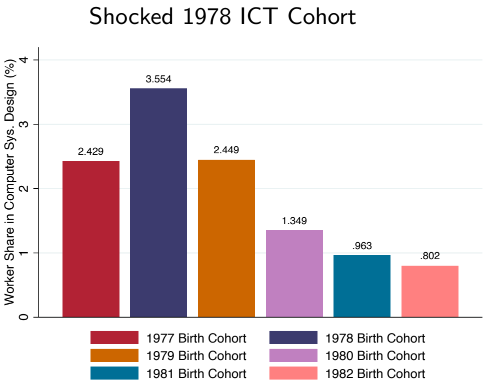
Notes: Mean entrepreneurship 2.6%, mean experience 2.1%. Implied elasticity = 2.0

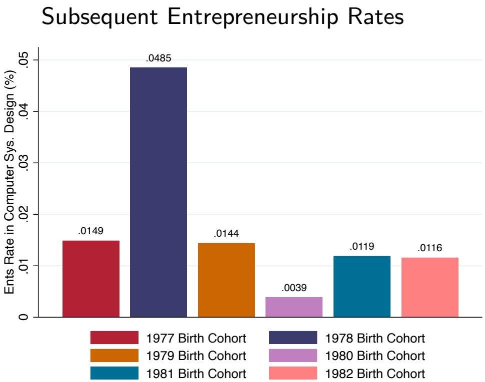
The Effect of Early Labor Market Experience on Entrepreneurship

Dependent Variable	Experience	P(Ent)	P(Ent)	P(Ent)	P(Ent)	P(Ent)
Age 22 County-Cohort Potential (IV)	0.3491	0.5683				
	(0.0131)	(0.0996)				
Experience			1.6278			
			(0.2766)			
Experience × Male				1.7740		
				(0.2672)		
Experience × Female				1.3311		
				(0.2923)		
Experience × ORG					2.2486	
					(0.4952)	
Experience × URG					1.9016	
					(0.5901)	
Experience × Par. Inc. Top 10						1.9885
						(0.3908)
Experience × Par. Inc. Bot 90						1.4224
						(0.4251)
Age 22 County + Cohort Fixed Effects	x	x	x	x	x	x
Age 22 County Clustered SE	x	x	x	x	x	x
Observations	10,074,353	10,074,353	10,074,353	10,074,353	4,763,500	5,662,260
F-statistic			715.04	345.77	165.04	165.60

Notes: Mean entrepreneurship 2.6%, mean experience 2.1%. Implied elasticity = 2.0

First Job Design: Illustrative Example from San Francisco CZ

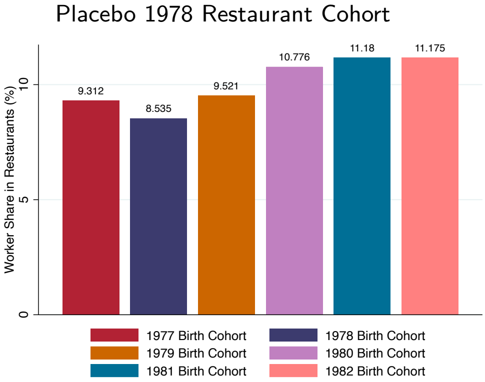


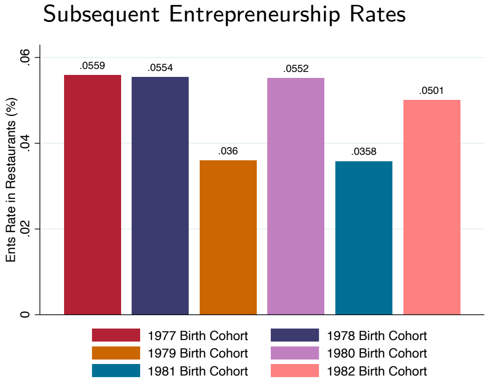


Challenge: Industry relation does not isolate supply effects of experience → Cohort design

1. Compare 22-year-old workers in adjacent cohorts in a CZ (e.g. 1978 cohort age 22 in 2000 internet boom, 1979 cohort 22 in 2001 bust)
2. High employment shares in an industry as a proxy for **experience shocks**
3. Subsequent ent. entry rate in same industry identifies **causal experience effects**

First Job Design: Illustrative Example from San Francisco CZ

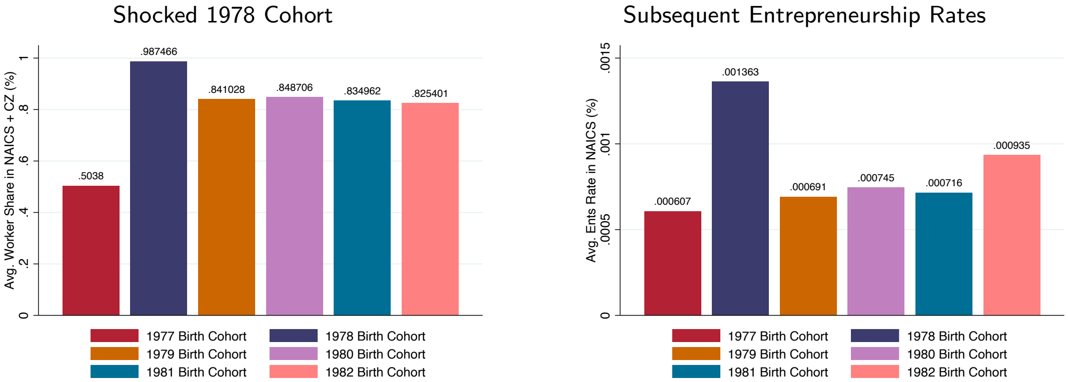




Challenge: Industry relation does not isolate supply effects of experience → Cohort design

1. Compare 22-year-old workers in adjacent cohorts in a CZ (e.g. 1978 cohort age 22 in 2000 internet boom, 1979 cohort 22 in 2001 bust)
2. High employment shares in an industry as a proxy for **experience shocks**
3. Subsequent ent. entry rate in same industry identifies **causal experience effects**

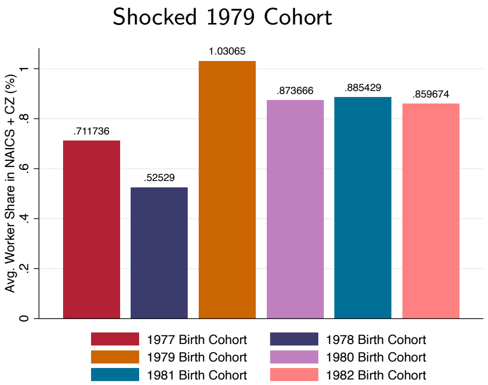
Pooled First Job Shocks (Top Quartile)

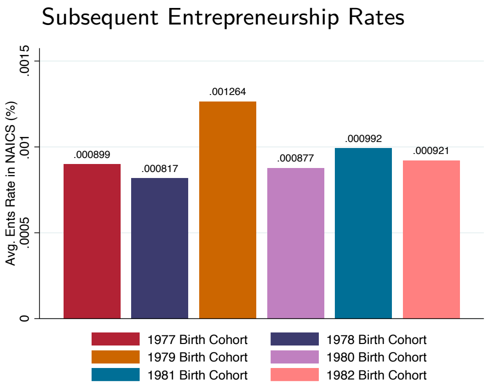


Challenge: Industry relation does not isolate supply effects of experience → Cohort design

1. Compare 22-year-old workers in adjacent cohorts in a CZ
2. High employment shares in an industry as a proxy for **experience shocks**
3. Subsequent entry rate in same industry identifies **causal experience effects**

Pooled First Job Shocks (Top Quartile)

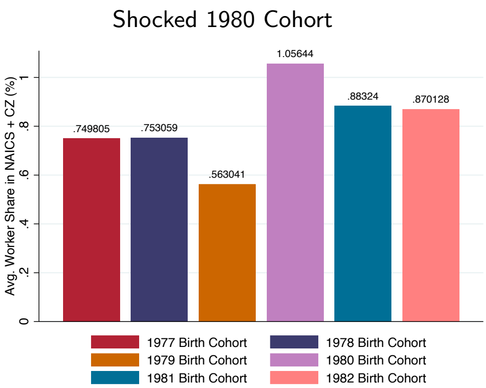


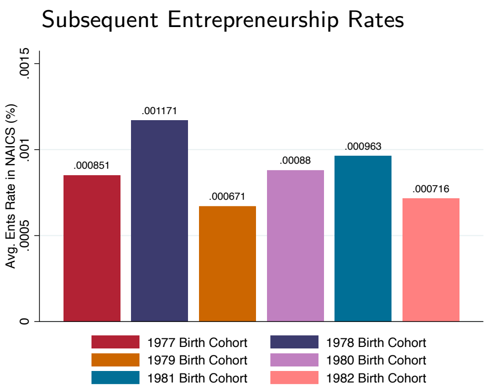


Challenge: Industry relation does not isolate supply effects of experience → Cohort design

1. Compare 22-year-old workers in adjacent cohorts in a CZ
2. High employment shares in an industry as a proxy for **experience shocks**
3. Subsequent entry rate in same industry identifies **causal experience effects**

Pooled First Job Shocks (Top Quartile)

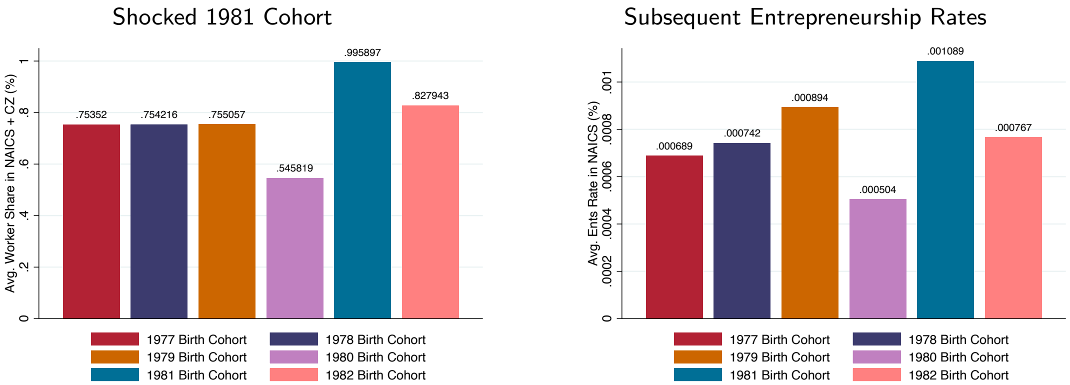




Challenge: Industry relation does not isolate supply effects of experience → Cohort design

1. Compare 22-year-old workers in adjacent cohorts in a CZ
2. High employment shares in an industry as a proxy for **experience shocks**
3. Subsequent entry rate in same industry identifies **causal experience effects**

Pooled First Job Shocks (Top Quartile)

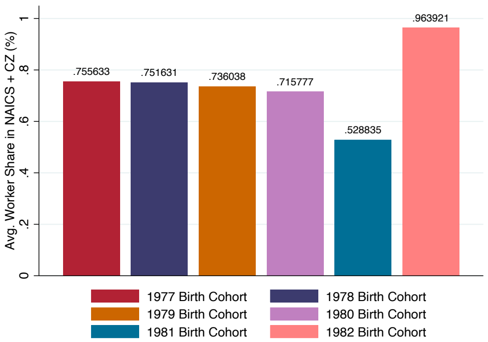


Challenge: Industry relation does not isolate supply effects of experience → Cohort design

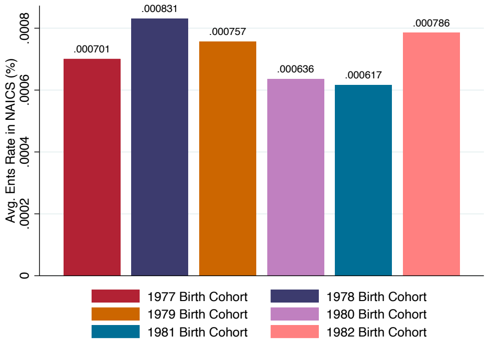
1. Compare 22-year-old workers in adjacent cohorts in a CZ
2. High employment shares in an industry as a proxy for **experience shocks**
3. Subsequent entry rate in same industry identifies **causal experience effects**

Pooled First Job Shocks (Top Quartile)

Shocked 1982 Cohort



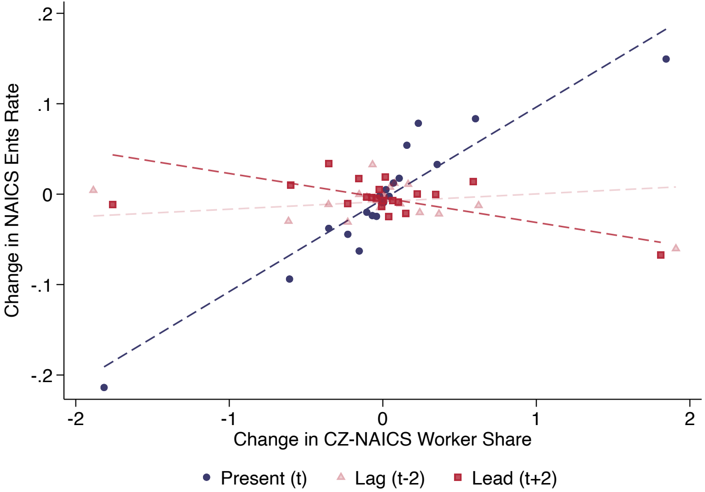
Subsequent Entrepreneurship Rates



Challenge: Industry relation does not isolate supply effects of experience → Cohort design

1. Compare 22-year-old workers in adjacent cohorts in a CZ
2. High employment shares in an industry as a proxy for **experience shocks**
3. Subsequent entry rate in same industry identifies **causal experience effects**

Pooled First Job Shocks: Placebos



Estimating the Person-Level "Returns" to Entrepreneurship

Goal: Measure average returns to entrepreneurship and differences across groups

- Restrict to first-time entrepreneurs and those with no prior business income
- Outcomes that may reflect firm exits
  - Today: total income, In progress: wealth, after-tax returns
- Follow entrepreneurs from $t = -5$ through $t = 8$

Estimating the Person-Level "Returns" to Entrepreneurship

Goal: Measure average returns to entrepreneurship and differences across groups

- Restrict to first-time entrepreneurs and those with no prior business income
- Outcomes that may reflect firm exits
  - Today: total income, In progress: wealth, after-tax returns
- Follow entrepreneurs from $t = -5$ through $t = 8$

Strategy: High dimensional 1-1 match between entrepreneurs and workers

1. Income history: AGI quantiles in $t \in \{-2, -3, -4\}$, with top 1% split into P99-99.9 and top 0.1%
2. NAICS 2-digit industry for employer in $t = -1$
3. Geo: Census region plus California
4. Age (3-year bucket), gender, single vs. joint filing status

Estimating the Person-Level "Returns" to Entrepreneurship

Goal: Measure average returns to entrepreneurship and differences across groups

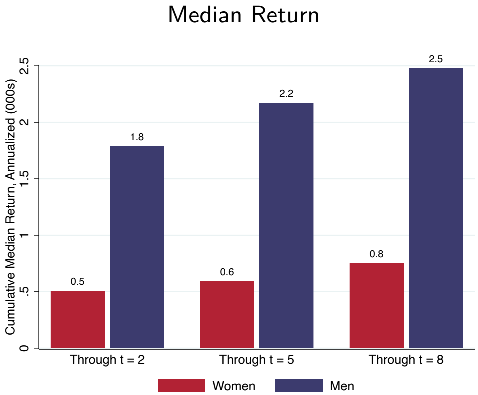
- Restrict to first-time entrepreneurs and those with no prior business income
- Outcomes that may reflect firm exits
  - Today: total income, In progress: wealth, after-tax returns
- Follow entrepreneurs from $t = -5$ through $t = 8$

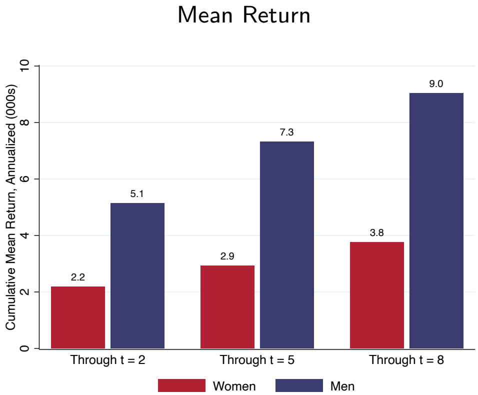
Strategy: High dimensional 1-1 match between entrepreneurs and workers

1. Income history: AGI quantiles in $t \in \{-2, -3, -4\}$, with top 1% split into P99-99.9 and top 0.1%
2. NAICS 2-digit industry for employer in $t = -1$
3. Geo: Census region plus California
4. Age (3-year bucket), gender, single vs. joint filing status

Mechanism: Test up-front barriers-to-entry model [Hsieh Hurst Jones Klenow 2019]

Return for Females is Half that for Males





Result: Female entrepreneurs earn lower returns → Inconsistent with Roy model logic

- Other factors/frictions cause women to earn less after entry
- Confirms SBO findings on incorporated+unincorporated ents (e.g., Robb 2002)
- Conditioning on founding industry closes ≈1/3 of the gender gap
- Similar results for low parent income kids

Additional Evidence on Mechanisms

1. **Experience effects**
   - Effects strongest for own-industry and stronger for technologically “close” industries
   - Experience effects possibly due to time, occupation in industry, networks
     - Female share of IPO workforce narrows the gender entry gap
     - Female IPO workers concentrate in less entrepreneurial occupations
2. Barriers to entry (incl. liquidity)
3. Exposure effects

Additional Evidence on Mechanisms

1. Experience effects

2. **Barriers to entry (incl. liquidity)**

   A. Returns evidence

   - Higher variance driven by high outside options, not large financial losses
   - Plausible levels/shape of risk aversion cannot generate entry patterns

   B. Liquidity evidence

   - Larger effects of wealth for men than women $\rightarrow$ Complementarity with experience?
   - Elasticities lower than relation between parental wealth and founding

3. Exposure effects

Additional Evidence on Mechanisms

1. Barriers to entry (incl. liquidity)
2. Experience effects
3. **Exposure effects**
    - Unlike inventors, less evidence of childhood dosage effects
    - Points toward mediation through local economy at working age
    - Unlikely to explain gender gap for workers at IPO firms

GE Model of Entrepreneurship, Discrimination and Talent Misallocation

- Two extensions to allow for entrepreneurship:
  - **Extension I:** Entrepreneurship as another career with different τ's
  - **Extension II:** In addition, agents can start in one sector & switch mid-career to become entrepreneurs (at a sector specific transition cost)

calibration

- Calibrate model with:
  - Moments from IRS (e.g. income gap between men and women from different sectors)
  - Moments from complementary datasets (e.g. ACS)
  - Existing literature (using HHJK as baseline)

- Consider different alternative policies to reducing frictions
- Focus today on removing frictions against
  - Female entrepreneurs
  - Women in all sectors (including entrepreneurs)

Policy Counterfactual: Removing Barriers for Women

Panel A: % of entrepreneurs			
	Baseline	Remove τs, entrepreneurs	
	(1)	(2)	(3)
		%	Δ p.p.
Men	1.6%	1.3%	-0.3%
Women	0.5%	2.8%	2.3%
Total	**2.1%**	**4.2%**	**2.1%**

Notes. This table reports the results of a counterfactual estimation for the last period of the model where τ^h and τ^w are removed. Income moments come from the IRS. Employment shares come from the CPS. GDP p.c. is computed as total labor income divided by total number of workers (in 2017 USD).

Policy Counterfactual: Removing Barriers for Women

- Removing barriers for female *entrepreneurs*:
  - $\uparrow$ GDP around 2.7%. (Higher if spillovers included)
  - % entrepreneurs doubles
  - % women entrepreneurs $\uparrow$ 6x, % male ents $\downarrow$ by $1/5$

Policy Counterfactual: Removing Barriers for Women

Panel A: % of entrepreneurs					
	Baseline	Remove τs, entrepreneurs		Remove τs, all sectors	
	(1)	(2)	(3)	(4)	(5)
		%	Δ p.p.	%	Δ p.p.
Men	1.6%	1.3%	-0.3%	1.5%	-0.1%
Women	0.5%	2.8%	2.3%	1.6%	1.1%
Total	**2.1%**	**4.2%**	**2.1%**	**3.1%**	**1.0%**

Notes. This table reports the results of a counterfactual estimation for the last period of the model where τ^h and τ^w are removed. Income moments come from the IRS. Employment shares come from the CPS. GDP p.c. is computed as total labor income divided by total number of workers (in 2017 USD).

Policy Counterfactual: Removing Barriers for Women in Basic Model

- Removing barriers for women in *all* sectors:
  - % entrepreneurs $\uparrow$ by 50% (less diversion)
  - % women ents triples, % male entrepreneurs essentially unchanged
  - $\uparrow$ GDP by 30%

Extension II: Labor Market Dynamics

- Novelty: When individuals enter labor market in period 1 in one sector, they can switch to entrepreneurs in period 2
- Timing
  - In t = 0: individuals choose a *path*: (sector i in t = 1 and sector i' in t = 2)
  - In t = 3: everyone stays in the same sector as t = 2
- Paths
  - Stayers: sector i in t = 1 $\Rightarrow$ sector i in t = 2
  - Switchers: sector i in t = 1 $\Rightarrow$ entrepreneurship in t = 2
- Switchers face a third friction: an entry barrier to entrepreneurship (τ^t)
  - τ^t depends on the sector at t = 1 (prior to founding a new firm)
  - τ^t computed from data: wage cut for women relative to men after founding a firm

Extension II: Labor Market Dynamics

- Novelty: When individuals enter labor market in period 1 in one sector, they can switch to entrepreneurs in period 2
- Timing
  - In t = 0: individuals choose a *path*: (sector i in t = 1 and sector i' in t = 2)
  - In t = 3: everyone stays in the same sector as t = 2
- Paths
  - Stayers: sector i in t = 1 $\Rightarrow$ sector i in t = 2
  - Switchers: sector i in t = 1 $\Rightarrow$ entrepreneurship in t = 2
- Switchers face a third friction: an entry barrier to entrepreneurship (τ^t)
  - τ^t depends on the sector at t = 1 (prior to founding a new firm)
  - τ^t computed from data: wage cut for women relative to men after founding a firm
- Preliminary Result: removing initial frictions to entrepreneurship friendly sectors highly effective (compared to removing pure ent τ or transition τ^t

Back

- (M+1) sectors: M market sectors + home sector
- Individuals live for three periods (young, middle age, old)

Workers

Back

- Preferences:

$$U = \left(c_{young} \cdot c_{middle} \cdot c_{old}\right)^{\beta} (1 - s)\, z\, \mu \tag{1}$$

- Human capital:

$$h = \bar{h} \gamma s^{\phi_i} e^{\eta} \tag{2}$$

- Consumption:

$$c = (1 - \tau^{w}) w \varepsilon h - (1 + \tau^{h}) e \tag{3}$$

- Talent: drawn from a multivariate Frechet ($\downarrow \theta = \uparrow$ talent dispersion)

$$F_g\left(\epsilon_1, \ldots, \epsilon_M\right) = \exp\left[-\sum_{i=1}^{M} \epsilon_i^{-\theta}\right] \tag{4}$$

Variable Description

◂ Back

U	Lifetime utility
c	Consumption
s	Schooling (normalized to 1, so (1-s) is leisure)
z	Group-specific utility from working in sector i
μ	Individual idiosyncratic utility from working in sector *i*
h	Human capital
$\bar{h}_{i,g}$	Differences in talent common to a group in a given sector i
γ	Return to experience
e	Education
ϕ	Return to time investment in human capital specific of sector *i*
η	Elasticity of human capital wrt to human capital expenditures.
w	Wage per efficiency unit
ε	Idiosyncratic talent
θ	Dispersion of talent across sectors
β	Trade-off btw consumption and time spent accumulating h

Solving the Model

Back

1. Workers' equilibrium:
   - Proposition 1: Sector choice P1
   - Proposition 2: Average quality of workers P2
   - Proposition 3: Average wages P3
   - Proposition 4: Relative propensities P4
   - Proposition 5: Relative LFP P5
2. Firm's equilibrium Firms
3. Workers + Firm Eq.

Estimation

- Setup
  - Focus on prime age workers (28 to 51) ▸ Cohort structure
  - 3 periods (2003, 2009, 2015). We estimate a counterfactual for last period (2015)
  - 24 market sectors (NAICS2 + entrepreneurship)
- Calibration
  - Income from IRS (2003-15)
  - Employment shares from CPS (1995-19) ▸ Data
  - Identification assumptions & parameter values from HHJK ▸ Parameters ▸ Assumptions
- Estimation results
  - Level of barriers (τs) faced by female ents is similar to lawyers and engineers
  - Removing τs for *women ents*: ↑ GDP by up to 10% & 2x number of ents
  - Removing τs for *all women*: ↑ GDP by up to 30% & 2x number of ents

Proposition 1: Sector Choice

Back

- The fraction of people in group g working in sector i equals:

$$p_{ig} = \frac{\tilde{w}_{ig}^{\theta}}{\sum_{s=1}^{M} \tilde{w}_{sg}^{\theta}} \tag{5}$$

- $\tilde{w}_{ig}$ (= return to working in a i for someone with average talent) is defined as:

$$\widetilde{w}_{ig} \equiv w_i s_i^{\phi_i} \left[1 - s_i\right]^{\frac{1-\eta}{3\beta}} \cdot \frac{\bar{h}_{ig} \tilde{z}_{ig}}{\tau_{ig}} \tag{6}$$

Proposition 2: Average Quality of Workers

Back

- The geometric average of worker quality in each sector is equal to:

$$\exp\left(\mathbb{E}\log\left[h_{igct}\,\epsilon_{igc}\right]\right) = \bar{\Gamma} s_{ic}^{\phi_{it}} \gamma(t-c) \left(\frac{\eta s_{ic}^{\phi_{ic}} \bar{\gamma} \bar{h}_{ig} w_{ic}\left[1-\tau_{igc}^{w}\right]}{1+\tau_{igc}^{h}}\right)^{\frac{\eta}{1-\eta}} \left(\frac{1}{p_{igc}}\right)^{\frac{1-\delta}{\theta(1-\eta)}} \tag{7}$$

Proposition 3: Average Wages

Back

- The geometric average of earnings in i by cohort c in period t of group g equals:

$$
\begin{aligned}
\overline{\text{wage}}_{igct} &\equiv \left(1-\tau^{w}_{igt}\right) w_{it} e^{\mathbb{E}\log[h_{igct}\epsilon_{ig}]} \\
&= \bar{\Gamma}\bar{\eta}\left[p^{\delta}_{igc} m_{gc}\right]^{\frac{1}{\sigma(1-\eta)}} \tilde{z}_{igc}^{-\frac{1}{1-\eta}} \left[1-s_{ic}\right]^{-\frac{1}{3\beta}} \times \frac{1-\tau^{w}_{igt}}{1-\tau^{w}_{igc}} \frac{w_{it}}{w_{ic}} \frac{\gamma(t-c)}{\bar{\gamma}} \frac{s_{ic}^{\phi_{it}}}{s_{ic}^{\phi_{ic}}}
\end{aligned}
\tag{8}
$$

Proposition 4: Relative Propensities

Back

- The fraction of group g employed in sector i relative to men equals:

$$\frac{p_{ig}}{p_{i,men}} = \left(\frac{\tau_{ig}}{\tau_{i,men}}\right)^{-\frac{\theta}{1-\delta}} \left(\frac{\bar{h}_{ig}}{\bar{h}_{i,men}}\right)^{\frac{\theta}{1-\delta}} \left(\frac{\overline{\mathrm{wage}}_{ig}}{\overline{\mathrm{wage}}_{i,men}}\right)^{-\frac{\theta(1-\eta)}{1-\delta}} \tag{9}$$

Proposition 5: Relative Labor Force Participation

Back

- The share of group g in the home sector relative to men (m) for equals:

$$\frac{1-\mathrm{LFP}_g}{1-\mathrm{LFP}_{men}} = \frac{m_{men}}{m_g} = \left(\frac{\overline{\mathrm{wage}}_{ig}}{\overline{\mathrm{wage}}_{i,men}}\right)^{-\theta(1-\eta)} \left(\frac{\tilde{z}_{ig}}{\tilde{z}_{i,men}}\right)^{-\theta} \left(\frac{p_{ig}}{p_{i,men}}\right)^{\delta} \ \forall \text{ market } i \tag{10}$$

where $\frac{m_{men}}{m_g} \equiv \frac{\sum_{i=1}^{M} \widetilde{w}_{i,\mathrm{men}}^{\theta}}{\sum_{i=1}^{M} \widetilde{w}_{ig}^{\theta}}$

Firms

◂ Back

- A representative firm produces final output Y from workers in M sectors:

$$Y = \left[\sum_{i=1}^{M} (A_i \cdot H_i)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \tag{11}$$

- H_i = total efficiency units of labor in sector i
- A_i = productivity of sector i (exogenously given)
- σ = elasticity of substitution across sectors in aggregate production.

Equilibrium

Back

- H_{it}^{demand} that satisfies the firm's profit maximization equals:

$$H_{it}^{demand} = \left(\frac{A_{it}^{\frac{\sigma-1}{\sigma}}}{w_{it}} \right) Y_t \tag{12}$$

- w_{it} clears the labor market in each sector so that $H_{it}^{supply} = H_{it}^{demand}$
- w_i is found numerically.

Comparison of IRS, CPS, and Census datasets

◂ Back

	IRS	CPS	Census
Time span (years)	2000 - 2015	1995 - 2019	1960 - 2010
Coverage	Universe of tax filers	Sample	Sample
Entrepreneurs	✓	✓	
Full income distribution	✓		
Industry	✓	✓	✓
Occupation		✓	✓
Home sector		✓	✓
Gender	✓	✓	✓
Race		✓	✓
Unemployed		✓	✓
Part-time workers		✓	✓

Notes. This table compares the information available in three distinct datasets: IRS, CPS, Census. The original model by HHJK is estimated using Census data.

Cohort Structure

Back

Year	Young (28 to 35)	Middle (36 to 43)	Old (44 to 51)
2003	3	4	5
2009	2	3	4
2015	1	2	3

Notes. This table shows the evolution of cohorts over time. For example, cohort 3 is young in 2003 (the first period of the model), middle-aged in 2009 (the second period of the model), and old in 2015 (the third (and last)period of the model). “Young" is defined as individuals aged 28 to 35 years old; “Middle” is defined as individuals aged 36 to 43; “Old” is defined as individuals aged 44 to 51.

Baseline Parameter Values

Back

Parameter	Value	Interpretation
α_0	0.5	Initial split between τ^h and τ^w
Lower constraint for τ^h	-0.8	
β	0.231	Consumption weight in utility
η	0.103	Elasticity of H wrt education spending
θ	2	Frechet shape
σ	3	Elasticity of substitution across sectors
δ	0	Fraction sorting on preferences

Notes. This table reports baseline parameter values (the same used by Hsieh Hurst Jones Klenow 2019).

Additional Estimation Assumptions

Back

- $\delta = 0$: sorting is entirely on talent (and *not* on preferences)
- $\frac{\bar{h}_{i,g}}{\bar{h}_{i,\mathrm{m}}} = 1$: we assume talent is distributed equally across sectors
- $\tau^h = 0$ and $\tau^w = 0$ for men in all sectors and all periods
- Home sector preference for all groups $= 1$
- The return to experience (γ) is the same for all sectors, groups, and cohorts.