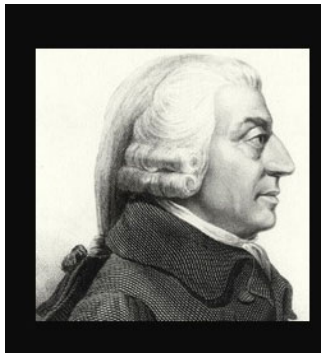
U.S. Entrepreneurship over the Long-run

Innessa Colaiacovo (Harvard Business School)
Dan Gross (Duke University & NBER)
Jorge Guzman (Columbia University)

NBER Innovation Information Initiative
Dec 2021

Entrepreneurship is assumed to be at the core of economic efficiency and growth

For economic efficiency
(local / main street entrepreneurship)

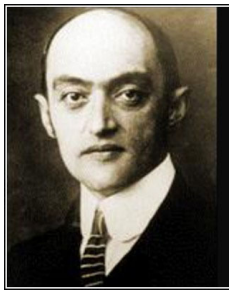


Adam Smith

The butcher, the baker, and the brewer…

Hurst and Pugsley (2011)

And translating innovations into economic growth
(high growth entrepreneurship)



Schumpeter

Growth through Heterogeneous Innovations

Ufuk Akcigit

University of Chicago, National Bureau of Economic Research, and Centre for Economic Policy Research

William R. Kerr

Harvard University and National Bureau of Economic Research

We build a tractable growth model in which multiproduct incumbents invest in internal innovations to improve their existing products, while new entrants and incumbents invest in external innovations to acquire new product lines. External and internal innovations generate heterogeneous innovation qualities, and firm size affects innovation incentives. We analyze how different types of innovation contribute to economic growth and the role of the firm size distribution. Our model aligns with many observed empirical regularities, and we quantify our framework with Census Bureau and patent data for US firms. Internal innovation scales moderately faster with firm size than external innovation.

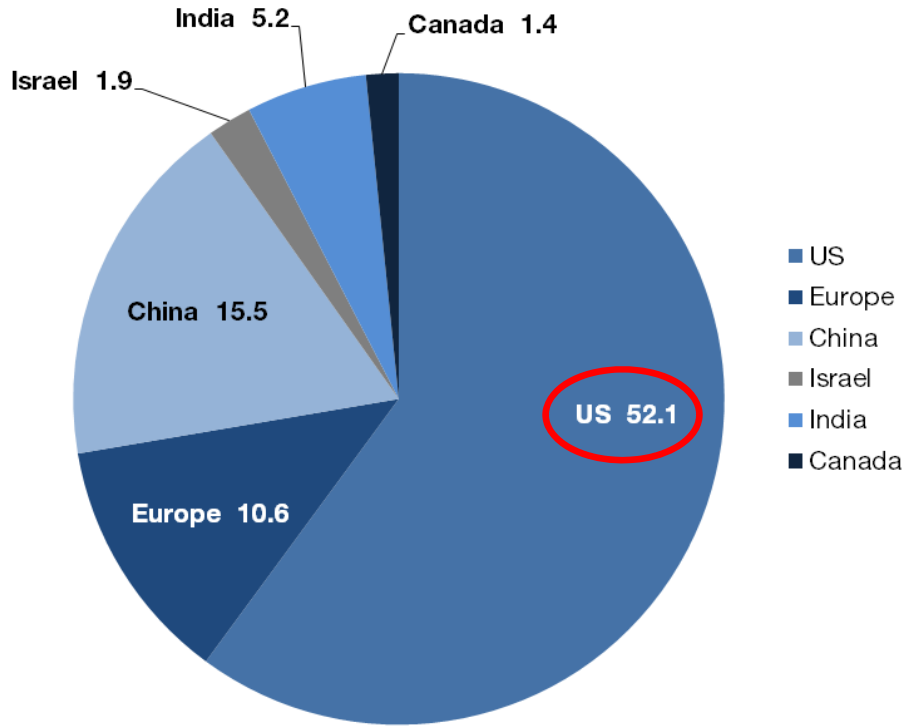
Nowhere do people think this link is truer than in the United States.

"As to whatever may depend on enterprise, we need not fear to be outdone by any people on earth. **It may almost be said that enterprise is our element.**"

Alexander Hamilton, 1795

Top countries for total venture capital invested

Share of total venture capital invested 2014, $ billions



Source: Ernst & Young

Yet, although economic growth is intrinsically a long term phenomenon…

We have almost no systematic evidence about entrepreneurship before ca. 1980

- VentureXpert (~1970), Crunchbase (2003), US Census LBD (1977), Startup Cartography Project (1988), NETS (1989)

U.S. Entrepreneurship in the Long-run

- How has U.S. entrepreneurship changed over the long run—both for Main Street and high growth startups?
- Has the rate of firm creation accelerated or slowed over long horizons?
- How does entrepreneurship evolve during major events in U.S. history?
  - WW2, manufacturing decline in US, etc.
- What is the role of innovation in propelling entrepreneurship?

What we're doing in this project

Data

- Use business registration records (Guzman & Stern 2020) to study the evolution of firm registrations per capita since 1900

- Use D&B data to train and validate a procedure for classifying these firms to economic sectors, based on their names
  - General: The required input for training is a list of firms with SICS/NAICS codes, and the required input for classification is a list of firms (that's all)
  - Flexible: Returns a classification with the probability that a firm belongs to each sector (or "sector share"—allows firms to span boundaries)
  - Meaningful: Procedure validates empirically + passes smell test

- Measure add'l economic phenomena via registrations

What we're doing in this project

Facts (a subset shown here)

- There has been a substantial growth in the aggregate rate of firm creation over the last century—both in absolute and per-capita
- World War II marked a structural break in U.S. entrepreneurship: startup growth is flat before, and (permanently) growing after
- Long-view insights into industry and regional dynamics
  - Decline of manufacturing in the Rust Belt—or return to trend?
- Dynamics of entrepreneurship around new technology
  - When important technologies enter patent record, we see firms with those terms after—some diffuse faster, others slower; some rise & fall, others perpetually grow
  - New views on Gort & Klepper (EJ 1982) or Suarez & Utterback (SMJ 1995) at scale

Data and Approach

The Startup Cartography Project

Startup Cartography Project

New USA business registrations since 1988. See methodology.

- Business registrations are public records recorded in the Secretary of State offices.

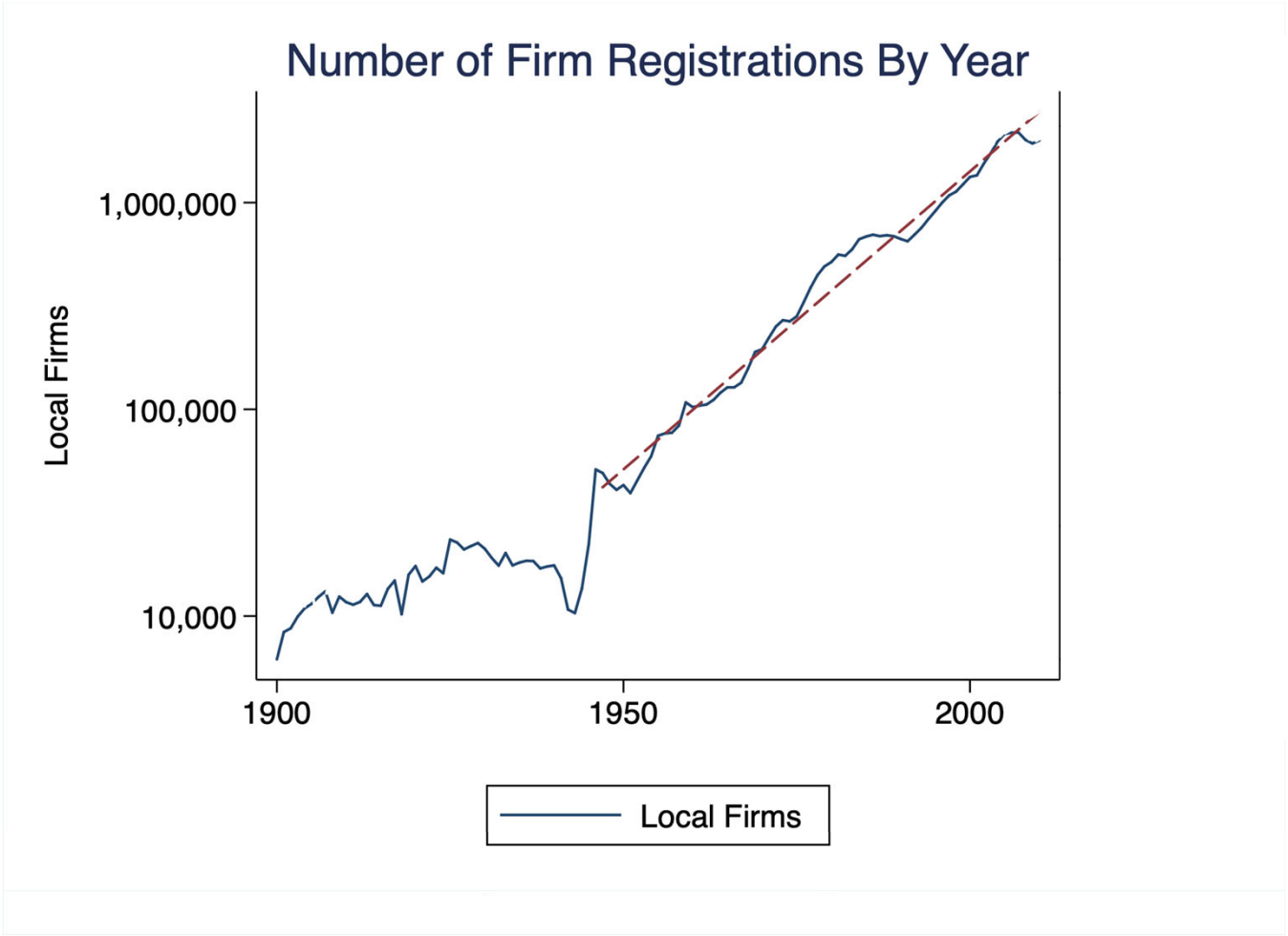
- The Startup Cartography Project (Andrews et al, 2021) is an effort to request and process business registration data from all 50 US states (and DC).

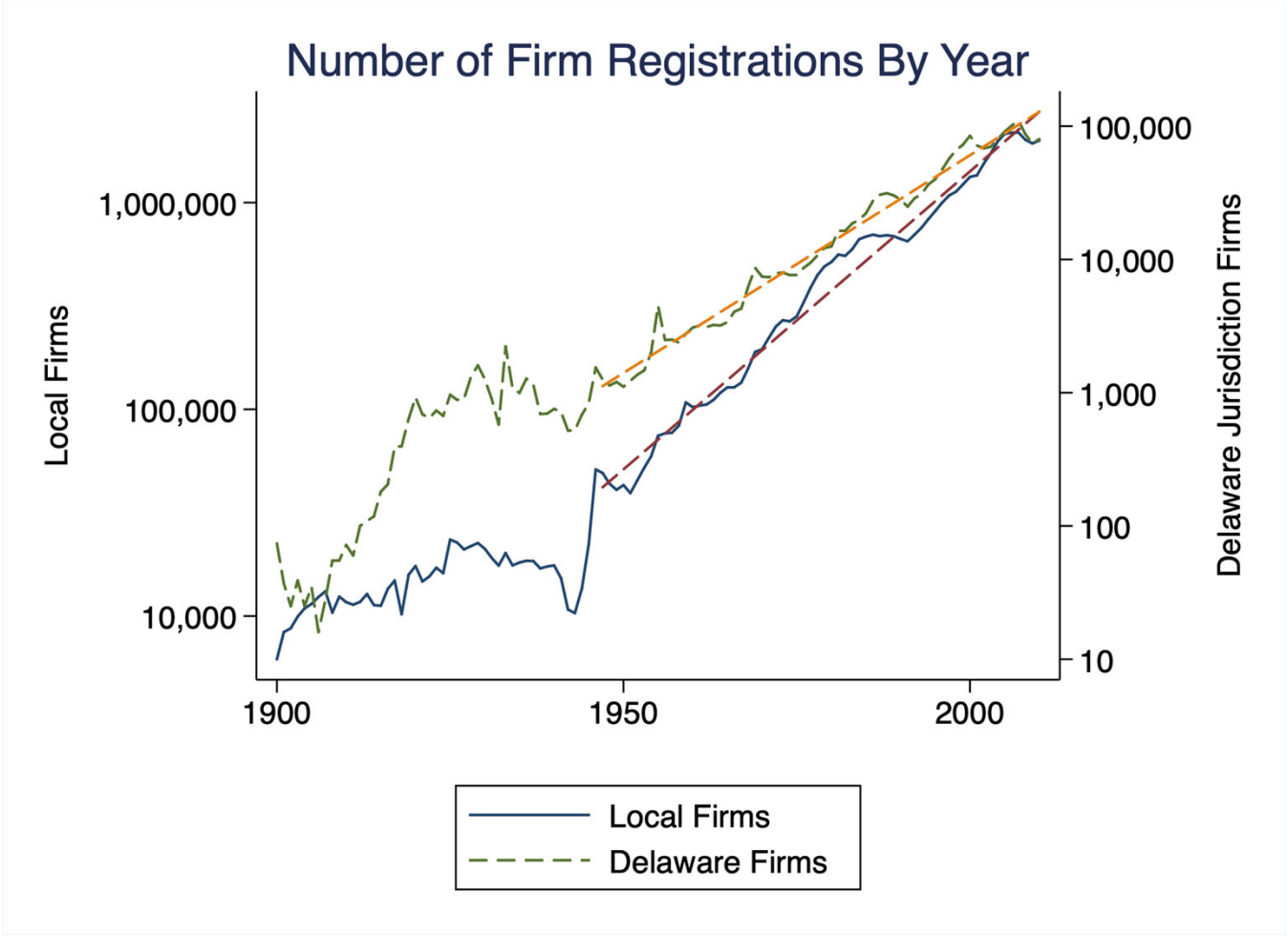
- We received data from 1900 (or state founding) for all but three states (Nebraska, North Carolina, Illinois)

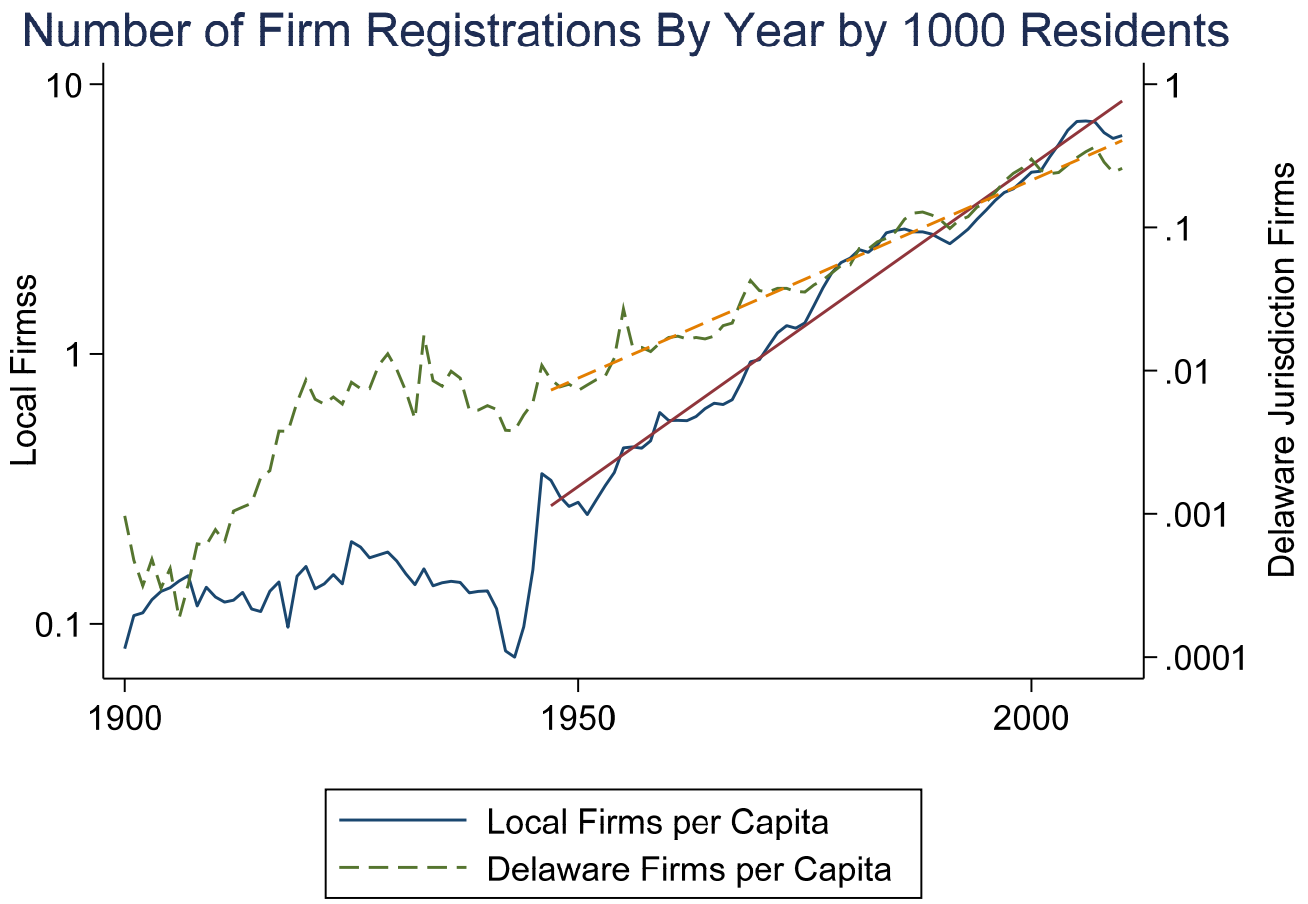
Available Data

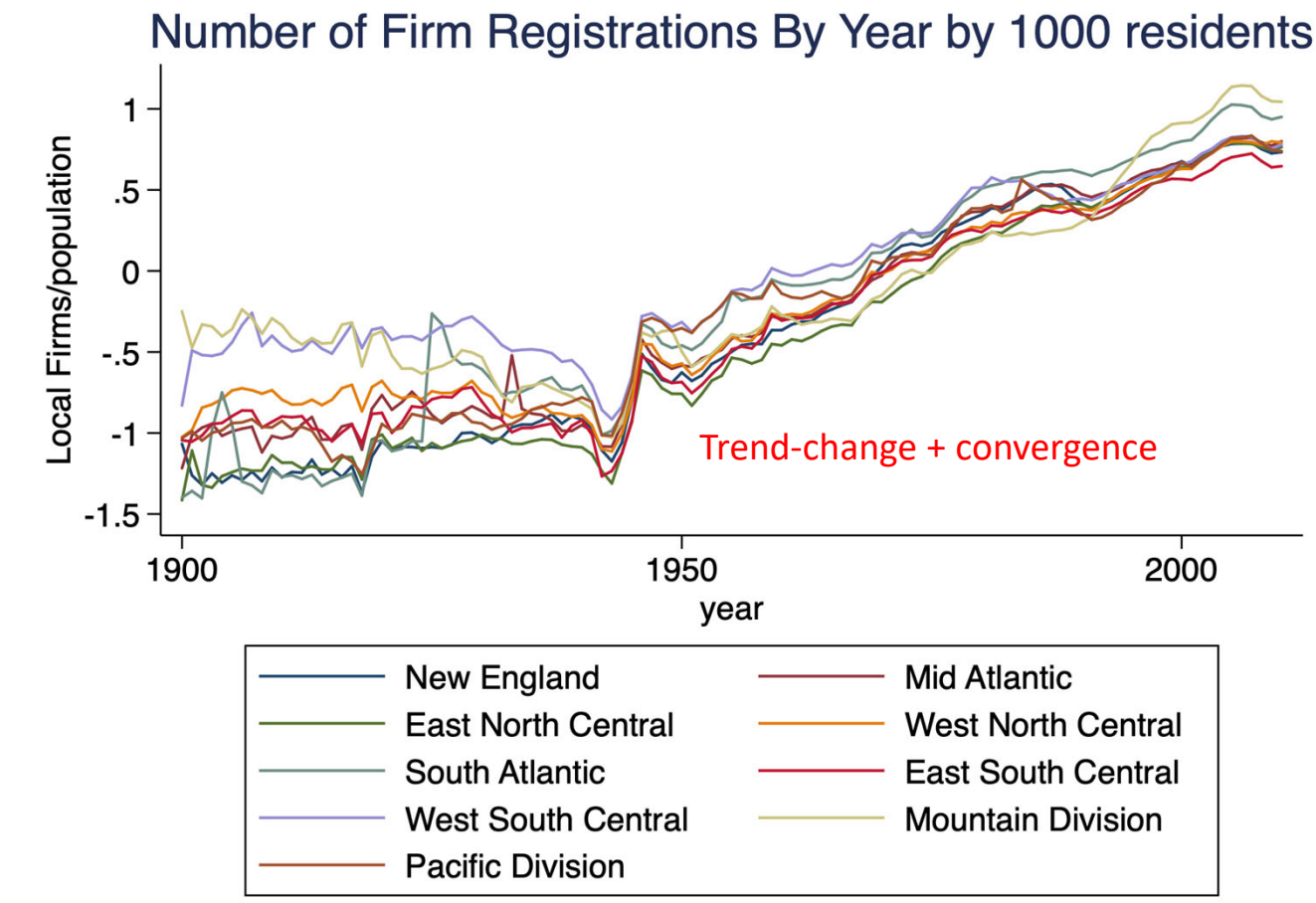
- Here we focus on four key variables:
  - Firm name, state, incorporation year, whether Delaware-registered

- How much can be learned from just these four things?


Number of Firm Registrations By Year
Local Firms
1,000,000
100,000
10,000
1900
1950
2000
Local Firms


Number of Firm Registrations By Year
Local Firms
1,000,000
100,000
10,000
1900
1950
2000
100,000
10,000
1,000
100
10
Delaware Jurisdiction Firms
Local Firms
Delaware Firms


Number of Firm Registrations By Year by 1000 Residents
Local Firmss
10
1
0.1
Delaware Jurisdiction Firms
1
.1
.01
.001
.0001
1900
1950
2000
Local Firms per Capita
Delaware Firms per Capita


Number of Firm Registrations By Year by 1000 residents
Local Firms/population
1
.5
0
-.5
-1
-1.5
1900
1950
2000
year
Trend-change + convergence
New England
East North Central
South Atlantic
West South Central
Pacific Division
Mid Atlantic
West North Central
East South Central
Mountain Division

But what *industries* are increasing and decreasing over time?

Inferring Industrial Sectors from Firm Names

- There is no systematic measure of industry in firm registrations
- But: might there be information in firm names?
- Our conceptual approach:
  1. Begin with a dataset that measures firms and their sectors
  2. Use the distribution of firm names across industries to quantify the association of specific words with each sector
  3. Use word industry scores to infer any given firm's sector

Most common tokens in 1940s (D&B 1990)

Tokens in firm names			
1.	SERVICE	11.	CLUB
2.	CONSTRUCTION	12.	FURNITURE
3.	SHOP	13.	AGENCY
4.	SUPPLY	14.	OIL
5.	ELECTRIC	15.	FARM
6.	CHURCH	16.	LUMBER
7.	AUTO	17.	PLUMBING
8.	MOTOR	18.	PRODUCT
9.	HOME	19.	INSURANCE
10.	STORE	20.	TRUCKING

Most common tokens in 1940s (D&B 1990)

Tokens in firm names			
1.	SERVICE	11.	CLUB
2.	CONSTRUCTION	12.	FURNITURE
3.	SHOP	13.	AGENCY
4.	SUPPLY	14.	OIL
5.	ELECTRIC	15.	FARM
6.	CHURCH	16.	LUMBER
7.	AUTO	17.	PLUMBING
8.	MOTOR	18.	PRODUCT
9.	HOME	19.	INSURANCE
10.	STORE	20.	TRUCKING

Most common tokens in 1980s (D&B 1990)

Tokens in firm names			
1.	SERVICE	11.	PRODUCT
2.	CONSTRUCTION	12.	STORE
3.	SYSTEM	13.	FOOD
4.	AUTO	14.	VIDEO
5.	SHOP	15.	REPAIR
6.	CENTER	16.	HOME
7.	SUPPLY	17.	MARKET
8.	ELECTRIC	18.	PLUMBING
9.	DESIGN	19.	MANAGEMENT
10.	RESTAURANT	20.	INDUSTRY

Most common tokens in 1980s (D&B 1990)

Tokens in firm names			
1.	SERVICE	11.	PRODUCT
2.	CONSTRUCTION	12.	STORE
3.	SYSTEM	13.	FOOD
4.	AUTO	14.	VIDEO
5.	SHOP	15.	REPAIR
6.	CENTER	16.	HOME
7.	SUPPLY	17.	MARKET
8.	ELECTRIC	18.	PLUMBING
9.	DESIGN	19.	MANAGEMENT
10.	RESTAURANT	20.	INDUSTRY

Our specific approach

dun & bradstreet

- Using firm samples from Dun & Bradstreet for 1970, 1980, 1990, we:
  - Retrieve firm names, SIC codes, states, and reported founding years
  - Crosswalk SICs to 10 high-level economic sectors
  - Clean firm names of person names, place names, stop words, etc.
  - Calculate remaining tokens' sector distribution ("sector scores")
  - Aggregate tokens in firm names and average sector scores

- This returns a probability that each firm is in each sector
  - Alternative interpretation: the firm's association with each sector—a flexible classification that permits sectoral boundary spanning

Example: “ANDERSON HOME APPLIANCES”

Word	Agriculture	Construction	Fin./Ins./RE	Retail trade	Services
Anderson					
Home					
Appliances					
Average					

Example: “ANDERSON HOME APPLIANCES”

Word	Agriculture	Construction	Fin./Ins./RE	Retail trade	Services
Anderson	8%	17%	10%	14%	34%
Home					
Appliances					
Average					

Example: “ANDERSON HOME APPLIANCES”

Word	Agriculture	Construction	Fin./Ins./RE	Retail trade	Services
Anderson	8%	17%	10%	14%	34%
Home	0%	20%	10%	17%	48%
Appliances					
Average					

Example: “ANDERSON HOME APPLIANCES”

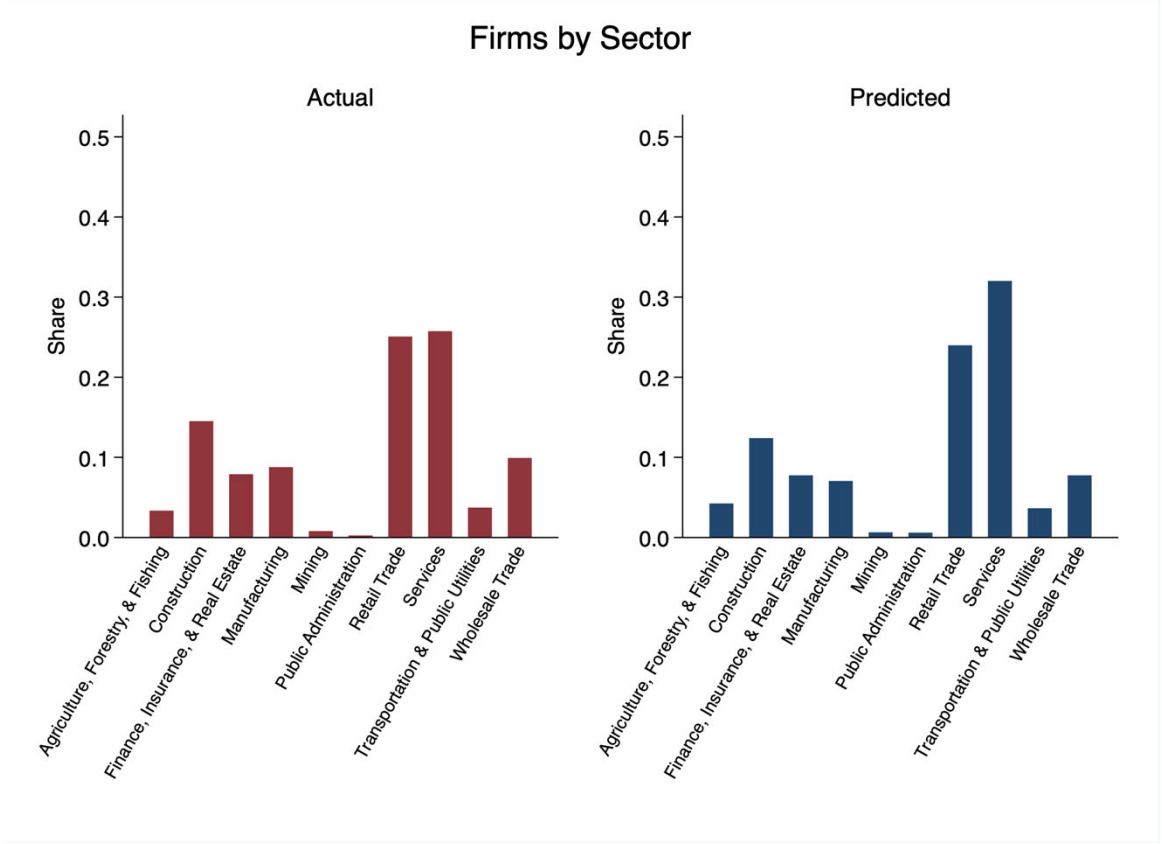
Word	Agriculture	Construction	Fin./Ins./RE	Retail trade	Services
Anderson	8%	17%	10%	14%	34%
Home	0%	20%	10%	17%	48%
Appliances	0%	2%	0%	71%	21%
Average					

Example: “ANDERSON HOME APPLIANCES”

Word	Agriculture	Construction	Fin./Ins./RE	Retail trade	Services
Anderson	8%	17%	10%	14%	34%
Home	0%	20%	10%	17%	48%
Appliances	0%	2%	0%	71%	21%
Average	**3%**	**13%**	**7%**	**34%**	**34%**

- Actual top 1 & 2 sectors in D&B: Services + Retail trade
  - But we might think it’s a little bit construction, too

Validation: Predicted vs. actual distribution



Validation: When top-predicted sector is (row), what is probability that D&B sector is (column)?

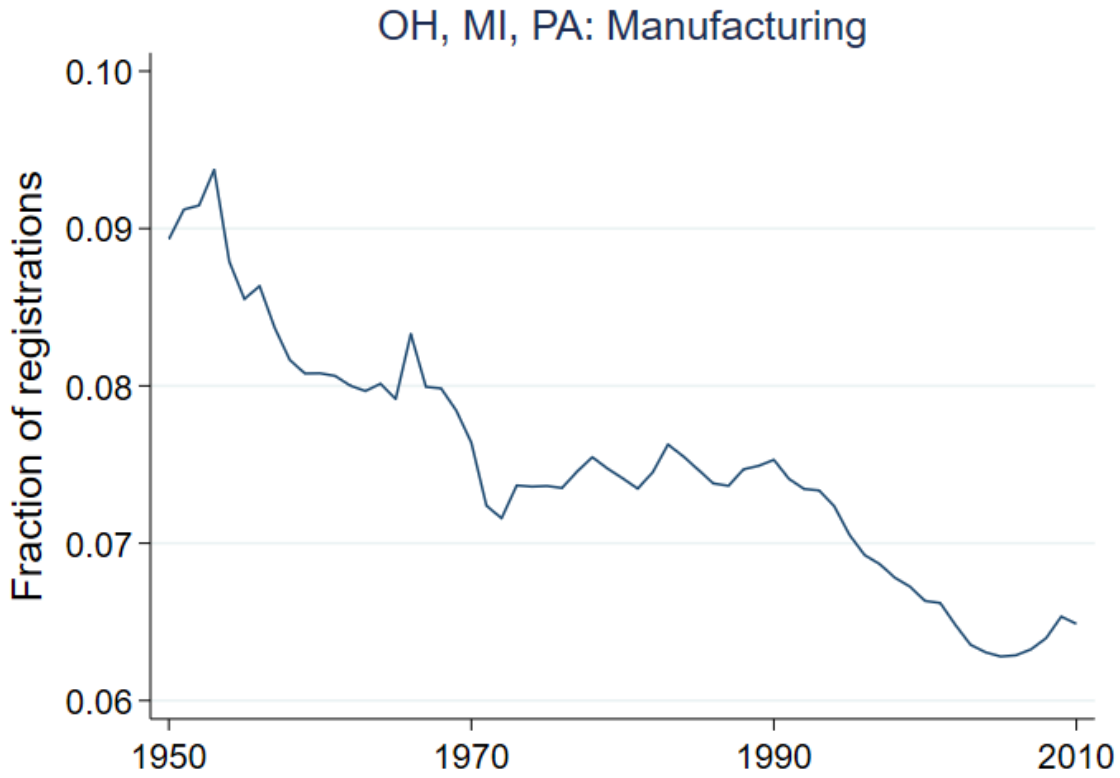
	Sector									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1: Agriculture, Forestry, & Fishing	78.1	1.7	1.5	1.2	0.1	0.1	7.4	4.8	0.7	4.4
2: Construction	0.6	83.5	1.2	2.5	0.8	0.0	3.0	4.4	0.8	3.0
3: Finance, Insurance, & Real Estate	0.7	3.5	82.2	0.9	0.3	0.2	3.1	6.7	0.8	1.6
4: Manufacturing	0.6	3.7	1.0	66.9	0.4	0.0	6.7	8.8	0.7	11.1
5: Mining	0.7	3.8	3.7	3.6	57.5	0.1	9.0	6.4	3.4	11.7
6: Public Administration	0.6	1.0	4.2	0.7	0.0	81.2	2.2	5.7	2.7	1.5
7: Retail Trade	1.5	1.8	1.8	4.8	0.4	0.2	72.7	8.7	1.2	6.9
8: Services	2.0	3.8	3.9	4.1	0.4	1.0	11.2	66.1	2.1	5.3
9: Transportation & Public Utilities	0.6	2.2	0.8	1.3	0.2	0.2	2.7	5.2	84.4	2.3
10: Wholesale Trade	2.2	2.5	1.4	8.1	2.1	0.1	18.0	6.5	2.2	57.0

Validation: When top-predicted sector is (row), what is probability that D&B sector is (column)?

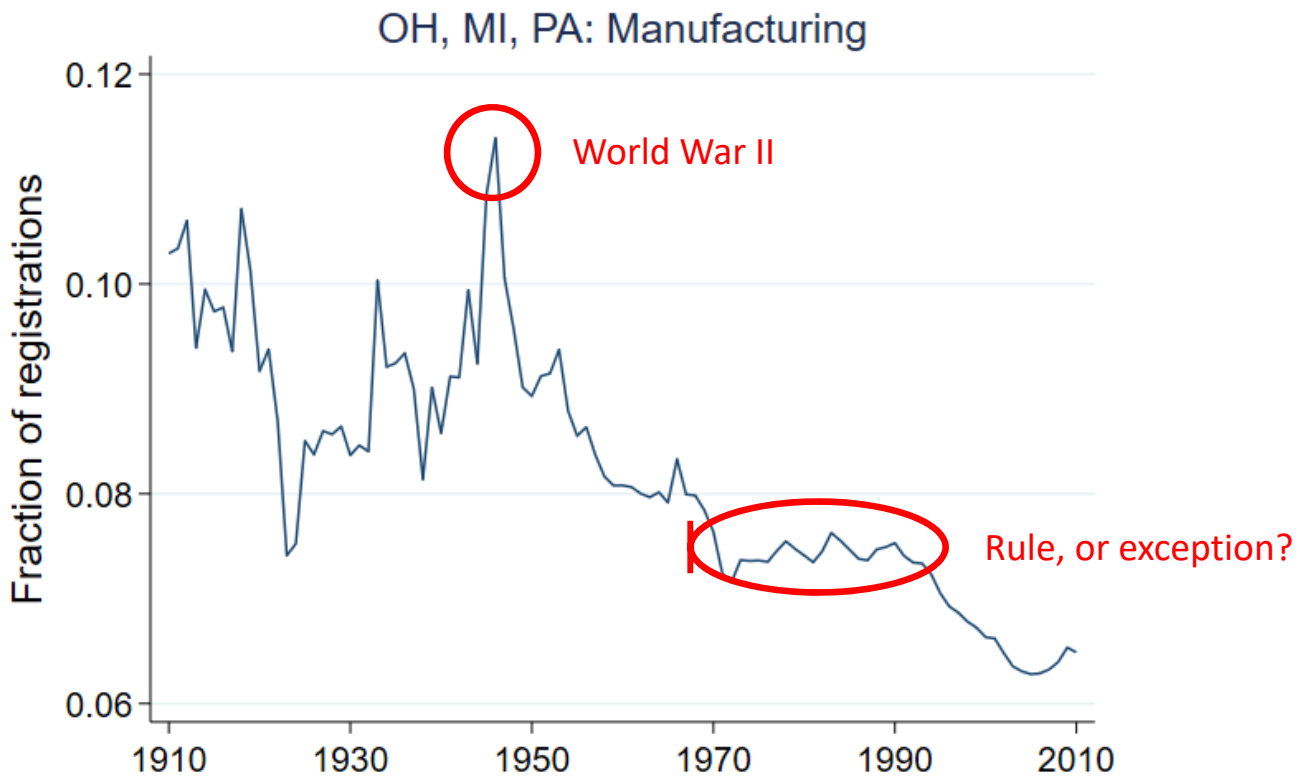
	Sector									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1: Agriculture, Forestry, & Fishing	78.1	1.7	1.5	1.2	0.1	0.1	7.4	4.8	0.7	4.4
2: Construction	0.6	83.5	1.2	2.5	0.8	0.0	3.0	4.4	0.8	3.0
3: Finance, Insurance, & Real Estate	0.7	3.5	82.2	0.9	0.3	0.2	3.1	6.7	0.8	1.6
4: Manufacturing	0.6	3.7	1.0	66.9	0.4	0.0	6.7	8.8	0.7	11.1
5: Mining	0.7	3.8	3.7	3.6	57.5	0.1	9.0	6.4	3.4	11.7
6: Public Administration	0.6	1.0	4.2	0.7	0.0	81.2	2.2	5.7	2.7	1.5
7: Retail Trade	1.5	1.8	1.8	4.8	0.4	0.2	72.7	8.7	1.2	6.9
8: Services	2.0	3.8	3.9	4.1	0.4	1.0	11.2	66.1	2.1	5.3
9: Transportation & Public Utilities	0.6	2.2	0.8	1.3	0.2	0.2	2.7	5.2	84.4	2.3
10: Wholesale Trade	2.2	2.5	1.4	8.1	2.1	0.1	18.0	6.5	2.2	57.0

First look: Sector composition of U.S. startups over 100 years, 1910-2010

Manufacturing in the Rust Belt (SCP)

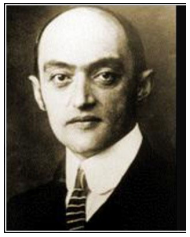


Manufacturing in the Rust Belt (SCP)



Connecting innovation and entrepreneurship

- Creative destruction: startups form around new innovations.

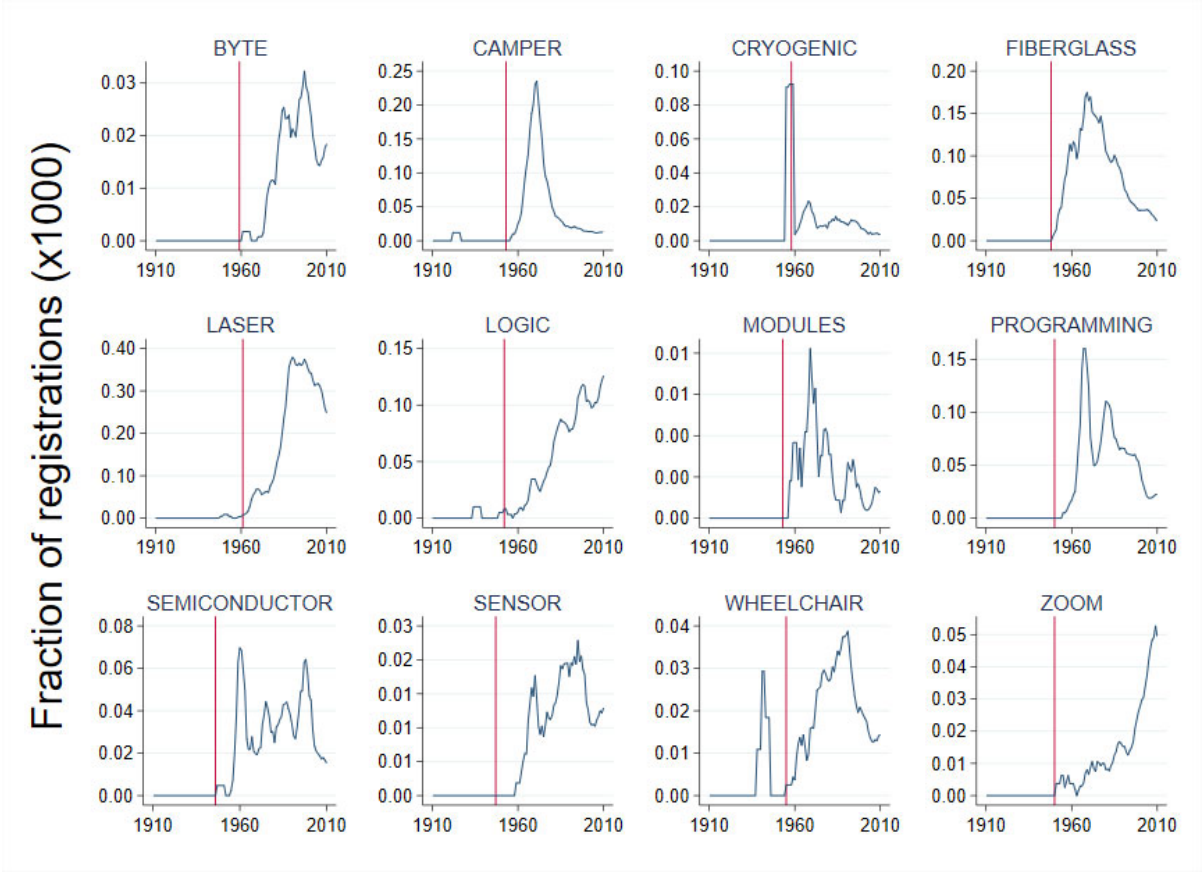


- How to link innovation to entrepreneurship?

  We take two approaches:

  1. Look for new (never-before used) words in firm names
  2. Look for words being introduced into the patent record, and trace their usage in firm names in the years before/after

Fraction of firms w/ new word from patent (SCP)



Notes:

- Sampled words are those introduced into patent record in the 1950s with >100 uses by 1980.
- Red line marks the year of the first patent filing with the given word.

New bigrams, by decade (D&B 1990)

1910s	…	1960s	1970s	1980s
PLUMBING_HEATING		HI_TECH	HOUR_PHOTO	BOXES_ETC
CREDIT_UNION		INFORMATION_SYSTEMS	CONVENIENCE_STORE	FAMILY_VIDEO
ELECTRIC_SERVICE		FITNESS_CENTER	SIR_SPEEDY	VIDEO_STORE
TOOL_DIE		CARPET_CARE	SPEEDY_PRINTING	VIDEO_RENTAL
WELL_DRILLING		CONCRETE_PUMPING	MINUTEMAN_PRESS	VIDEO_EXPRESS

Note: Removed “SONIC_DRIVE” from 1960s and “SUBWAY_SANDWICHES” from 1980s

Limitations:

- D&B 1990 sample approximates the universe of U.S. establishments in late 1980s/1990. Analysis to be re-run on SCP.
- Sample restricted to HQ and single-branch establishments; franchised companies will thus be overrepresented in the firm counts, or conversely, wholly-owned multi-establishment firms underrepresented relative to franchises.

Big-picture takeaways

- U.S. entrepreneurship since 1900: a mix of continuity and change
  - Conventional wisdom gets some crucial details wrong
  - The exact story also depends on the sample: DE firms vs. Main Street
- A large amount of information can be gleaned from firm names
  - Here we build on e.g. Belenzon-Chatterji-Daley 2017, 2019
- Our new data and approach provide a novel way to document the evolution of the U.S. entrepreneurial ecosystem
  - Data are still under development, but we're planning for an eventual release

Open questions requiring long-run data

- What sparked the post-war entrepreneurship take-off?
- Technological vs. technology-*enabled* entrepreneurship
- Industry dynamics and technology lifecycles
- Corporate tax rates and entrepreneurship over the long-run
- Persistence of regional entrepreneurship rates – is there a local culture of entrepreneurship?
- Many more…

Next steps

- (Assorted) additional investments to improve sector-tagging
- Extend D&B text-based analysis to SCP
- Extend to all patents 1836 to 2000ish
  - Other decision rules for which patent terms to focus on
    - Most heavily used or cited: current approach
    - Highest-value? (Kogan et al. 2017)
    - “Breakthrough” innovations? (Kelly et al. 2021)

Eventually: prepare data for public release!