Competing for Inventors: Market Concentration and the Misallocation of Innovative Talent

Andrea Manera

October 5, 2021

Research Questions and Main Idea

I ask three main questions:

1. What are the boundaries of labor markets for *inventors?*
2. Does *competition for scarce inventors* between *different* product markets affect R&D allocation and productivity?
3. If so, which policies can restore efficiency?

Across product markets, more market power gives:

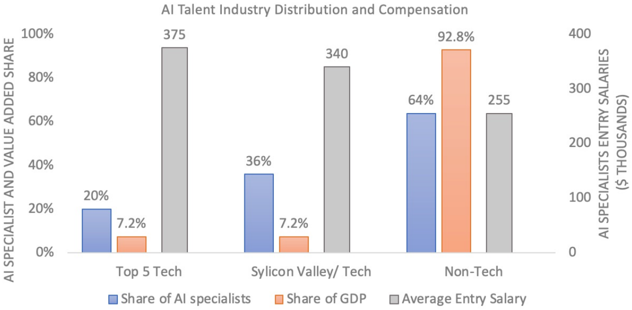
- Higher *private* returns to R&D, demand for inventors
- Lower *social* returns to R&D, less growth per inventor
- Uneven growth in concentration leads to *misallocation across sectors*

Complementary explanation for observed decline in R&D

Paper in a Slide

1. Build data on inventors' flows (USPTO) across product markets (NAICS 3-4 digits)
   - Competition for inventors extends beyond product markets
2. Long regressions (EC and regulations data, 1997-2012)
   - *Across product markets:* $\uparrow$ market concentration $\uparrow$ inventors
   - *Within product markets* that gained inventors*:* $\uparrow$ top 10% share, $\uparrow$ incremental/defensive innovation, $\downarrow$ growth/inventor
   - Misallocation explains up to .45pp lower annual growth
3. Schumpeterian model with defensive patenting:
   - Uneven markup increases lead to misallocation
   - Policy: entrant subsidies in *less competitive sectors*
   - Cost-neutral gives .155pp higher annual growth

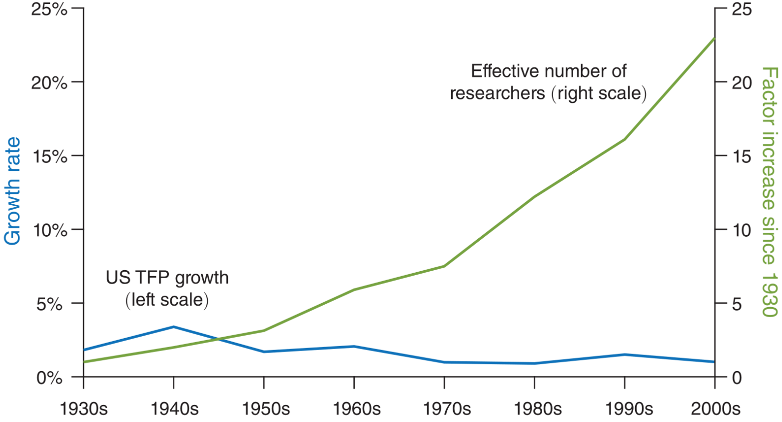
Motivating Example: The Allocation of AI Talent



Source: Global AI Talent Report (TalentSeer, 2020), BEA

- AI is a GPT, but top Tech attracts a disproportionate share of specialists (and offer higher wages)
    - Anecdotal widespread shortage in other sectors and smaller companies "outcompeted" by big tech ▸ Headlines

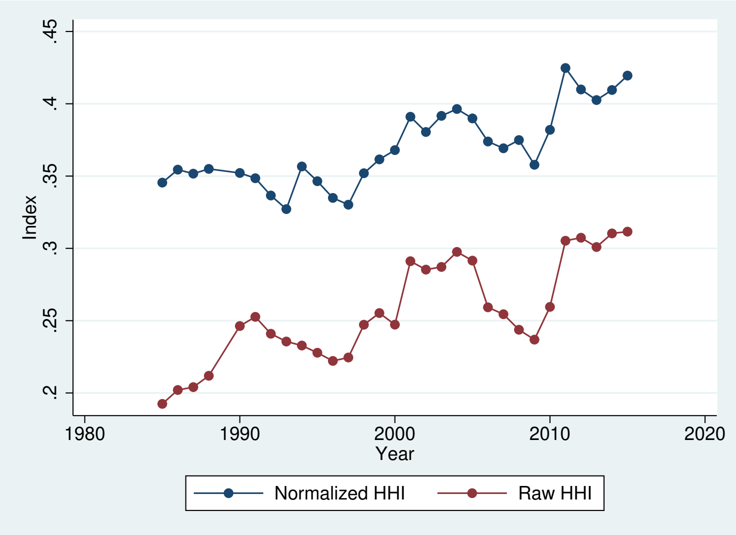
Are Ideas Harder to Find?



Source: Bloom et al. (2021)

Increased Concentration (3d NAICS)

TODO: use Census



Related Literature

- *Trends in innovation and R&D.*

  Akcigit and Kerr (2018), Akcigit and Ates (2020), **Bloom et al.** (2020, 2021), Goldschlag et al. (2016)

- *Increasing Concentration Facts and Measurement*

  Barkai (2020), De Loecker et al. (2020), **Gutiérrez and Philippon** (2017, 2018), Grullon et al. (2019), Keil (2017)

- *Competition and Innovation*

  Aghion et al. (2005, 2009, 2019), **Argente et al. (2020)**, Gutiérrez and Philippon (2017), Autor et al. (2021)

- *Models of innovation and growth*

  Aghion and Howitt (1992), Aghion et al. (2001), Acemoglu and Akcigit (2012), **Abrams et al. (2018)**, **Jo (2019)**

How is This Paper Different?

Empirical Literature:

- Focus on R&D *output* (patents and citations)
- Allocation of R&D expenditure *within* markets

This paper:

- Focus on R&D *input* (inventors)
- Allocation of *relevant* inventors *across* markets

Theory Literature:

- R&D activity is (usually) non-rival
- Competition and innovation *within* product markets

This paper:

- R&D is *rival* (scarce inventors and defensive innovation)
- Competition and innovation *across* different product markets

Plan of the Talk

1. Data construction
2. Regression analysis
3. Model
4. Calibration and Policy

Empirics Objectives

- Understand boundaries of markets for inventors
  - Identify "knowledge markets" as sets of product markets that hire the same type of inventors
  - Use patent data to build a network of flows of inventors across sectors
  - Identify connected sectors maximizing network's modularity
- Look *within knowledge markets* to see how product markets' share of inventors relate to concentration

Data Sources

- USPTO (patent-year) and Goldschlag et al. (2016):
    - patent citation and disambiguated inventor id's, 1975-present;
    - patent classification by NAICS of application (1978-2016)
- Economic Census and Keil (2017) (5-year-NAICS)
    - Concentration measure: HHI and HHI lower bound
    - Output per worker growth
- NBER-CES:
    - Constructed Lerner Index
- Mercatus RegData 4.0 (2021):
    - Sector-specific regulation counts
    - Extended using text similarity across NAICS for missing sectors

Dataset Structure

Patent ID	Inventor ID	Goldschlag et al. (2016) NAICS	Year
US00001	00001-1	1111	1980
US00001	00001-1	1112	1980
US00001	00001-2	1111	1980
US00001	00001-2	1112	1980
US00002	00001-1	3111	1981

Dataset Structure

Patent ID	Inventor ID	Goldschlag et al. (2016) NAICS	Year
US00001	00001-1	1111	1980
US00001	00001-1	1112	1980
US00001	00001-2	1111	1980
US00001	00001-2	1112	1980
US00002	00001-1	3111	1981

$\Downarrow \quad \Downarrow \quad \Downarrow$

Inventor ID	NAICS 1	NAICS 2	Year	Total Flow
00001-1	1111	1112	1980	2
00001-2	1111	1112	1980	2
00001-1	1112	3111	1981	1

"Knowledge Markets"

- *Knowledge Market:* set of NAICS (product markets) that employ the same type of inventors
  - To capture similar *required knowledge* to innovate
- From data, undirected network:
  - NAICS (4-digit) as nodes
  - *Minimal share of inventor flows* as edge weights, W

"Effective inventors"

- *Effective inventors:*
  - "Productivity-adjusted" inventor. Fixed effect α_i in regression:

$$\#\text{Patents}_{cfit} = \alpha_i + \alpha_{cft} + \varepsilon_{cfit}$$

  - α_{cft}: CPC class 1-digit, c, by firm (assignee), f, by year, t
  - Raw number of inventors for robustness

Effective Inventor Flows

- Strength of connection between two sectors
- Build directed flows for each inventor i (avoid double counting):

$$\tilde{\text{flow}}_{1\to 2,i,t} \equiv \frac{\sum 1\{i \text{ moves } 1 \to 2 \text{ in } t\}}{\sum_{j,k} 1\{i \text{ moves } j \to k \text{ in } t\}} \times \alpha_i$$

- Compute total outflows and inflows for each NAICS 4-digit sector:

$$\tilde{\text{inflow}}_{\text{NAICS}} = \sum_n \sum_t \sum_i \tilde{\text{flow}}_{n\to \text{NAICS},i,t},$$

Network Weights

- Compute share of inflows and outflows, e.g.:

$$\text{share in}_{1\leftarrow 2} = \frac{\sum_t \sum_i \tilde{\text{flow}}_{\,2\rightarrow 1,i,t}}{\tilde{\text{inflow}}_1}$$

- Define weight:

$$W_{12} = W_{21} = \min\left\{\frac{\text{share in}_{1\leftarrow 2} + \text{share out}_{1\rightarrow 2}}{2}, \frac{\text{share in}_{2\leftarrow 1} + \text{share}}{2}\right.$$

- Can use average, but risk of overstating flows from small sectors to large

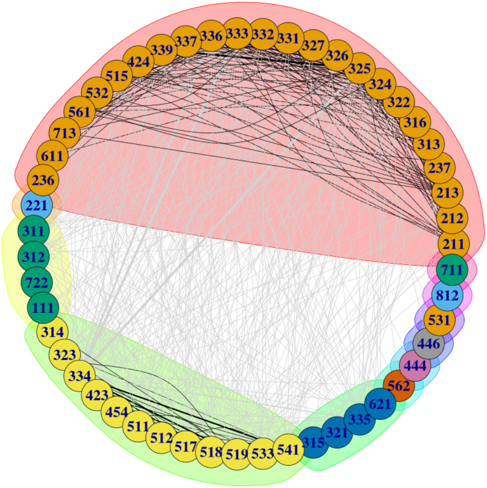
Detecting Knowledge Markets

- Run a weighted community detection algorithm:
  - Maximizes modularity of the network
  - Finds N *non-overlapping* communities to maximize:

$$Q = \sum_{c=1}^{N} \left[W_{cc} - \left(\sum_{j} W_{cj} \right)^2 \right],$$

  where W_{cj} is the weight edges that have one end in community c and the other in community j.
  - "How much more the community is connected internally than externally".
  - Result: 10 non-singleton sets of NAICS 4-digit that share inventors

Visualization at 3-digit NAICS



Features of Flows and Knowledge Markets

- Many connections across product markets even at 3 digits!
- Same inventors are employed by firms in *highly different* product markets
- Broad communities
- Reasonable?
  - Green Cluster collects "Food and Agriculture": Crop Production, Food Manufacturing and Services, Beverage and Tobacco;
  - Orange Cluster is mostly "Mining" and "Heavy Industry": e.g. Petroleum and Coal Products, Chemical, Machinery Manufacturing;
  - Yellow Cluster collects "Communications", "Electronics" and "Publishing":e.g. Computer and Electronic Products, Telecommunications, Data Processing;

Additional Data

- Sector's share of effective inventors in knowledge market, k, employed by sector p:

$$\text{Share}^k_{p,t} \equiv \frac{\sum_{p_i(t)=p} \alpha_i}{\sum_{k_i(t)=k} \alpha_i}.$$

- Baseline concentration measure is lower bound of HHI from Keil (2017):

$$\underline{\text{HHI}}_{p,t} = 4\left[\frac{\text{Top-4 Share}_{p,t}}{4}\right]^2 + 4\left[\frac{\text{Top-8 Share}_{p,t} - \text{Top-4 Share}_{p,t}}{4}\right]^2$$

  where top shares come from the Economic Census (corr. with actual HHI .93 when available)
- Regulation measure from Mercatus RegData 4.0: counts of regulation affecting NAICS 4d using text analysis
  - Extended to all sectors with HHI using cos-similarity Details

Specification

- At the NAICS 4-digit sector, p:

$$\Delta \text{Outcome}_p = f_k 1\{p \in k\} + \beta \Delta \text{Indep.Var.}_p + \gamma' \Delta \text{Controls}_p + \varepsilon_p,$$

- Δ denotes the long-difference operator:

$$\Delta \text{Outcome}_p = \text{Outcome}_{p,2012} - \text{Outcome}_{p,1997}$$

- $f_k 1\{p \in k\}$, indicator that sector p belongs to knowledge market k

Main Specification:

- $\Delta \text{Outcome}_p$: ΔShare_p^k
- $\Delta \text{Indep.Var.}_p$: $\Delta \underline{\text{HHI}}_p, \Delta \text{HHI}_p$
- $\Delta \text{Controls}_p$: Change in log-real sales; controls for sector size

$\beta > 0$: Sectors where concentration increased attracted more

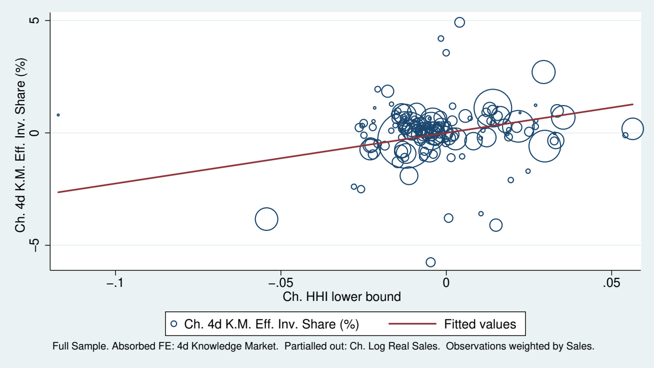
Main Specification Results

Robustness to Outliers | Raw Inventors

	Ch. 4d K.M. Eff. Inv. Share (%)	
	(1)	(2)
Ch. HHI lower bound	26.093*	22.509*
	(10.696)	(10.848)
Ch. Log Real Sales	0.914**	0.548*
	(0.278)	(0.243)
4D Knowledge Market FE		✓
Sample	Full Sample	Full Sample
Weight	Sales	Sales
Observations	157	153

0.1,* $p < 0.05$,** $p < .01$,*** $p < .001$

Graphically



Robustness to Individual Firm Size

▸ Robustness to Outliers

	Ch. 4d K.M. Eff. Inv. Share (%)	
	(1)	(2)
Ch. HHI lower bound	35.230**	20.783+
	(12.759)	(10.615)
Ch. Log Real Sales per company	0.175	-0.040
	(0.382)	(0.253)
4D Knowledge Market FE		✓
Sample	Full Sample	Full Sample
Weight	Sales	Sales
Observations	81	79

$0.1,^{*}\ p < 0.05,^{**}\ p < .01,^{***}\ p < .001$

IV Regression: Reduced Form and First Stage

	Ch. 4d K.M. Eff. Inv. Share (%)	Ch. HH
	(1)	
Ch. Log Restricitions (NAICS 4d)	0.478*	
	(0.220)	
Ch. Log Real Sales	0.539+	
	(0.274)	
4D Knowledge Market FE	✓	
Sample	Full Sample	Fu
Weight	Sales	
Observations	153	

0.1,* $p < 0.05$,** $p < .01$,*** $p < .001$

IV Regression: 2SLS Results

	Ch. 4d K.M. Eff. Inv. Share (%)	
	(1)	(2)
Ch. HHI lower bound	30.560+	30.096+
	(15.904)	(15.819)
Ch. Log Real Sales	0.544*	0.525*
	(0.244)	(0.247)
4D Knowledge Market FE	✓	✓
Sample	Full Sample	Mahalanobis 5%
Weight	Sales	Sales
Observations	157	150
First-Stage F	4.587229	4.753009
Anderson-Rubin p-value	.0281448	.0321185

0.1,* $p < 0.05$,** $p < .01$,*** $p < .001$

What Happens *within Knowledge Markets?*

An increase in inventor shares:

- Significantly increases
  - Top 10% firms' inventor shares (link to Table), Top 10%/Bottom 50% ratio
  - Self-citations (link to Table)
- Significantly decreases
  - Inventors' productivity
  - Patents' forward citations (link to Table)

Fall in Inventors' Productivity

	Ch. Avg. Output/Worker Growth/Inventor (%)			
	(1)	(2)	(3)	
Ch. 4d K.M. Eff. Inv. Share (%)	-0.007**	-0.005*	-0.007**	-
	(0.002)	(0.002)	(0.002)	(
Ch. Log Real Sales		-0.051*		-
		(0.021)		(
4D Knowledge Market FE	✓	✓	✓	
Sample	Full Sample	Full Sample	Mahalanobis 5%	Maha
Weight	Sales	Sales	Sales	
Observations	101	101	96	

$+\ p < 0.1,^{*}\ p < 0.05,^{**}\ p < .01,^{***}\ p < .001$

Sizable Growth Loss from Misallocation

Back-of-the-envelope:

Model Objectives

- Explain intuition on decrease in competition driving lower growth through misallocation
- Build a model that generates a positive relation between concentration and inventor demand
  - Schumpeterian model
    - Entrants give creative-destruction growth
    - Incumbents can engage in defensive innovation
    - Two sectors, one knowledge market
- Calibration matching R&D statistics to evaluate policy:
  - Optimal to subsidize entrants in concentrated sectors
  - Cost-neutral policy gives up to .155pp higher annual growth

Environment

Production and Competition

Incumbents' Values

Entrants

Growth

Comparative Statics: Markup Increase

Two-Sectors, Inventor Market Equilibrium

Model Properties

Calibration

Policy Table

Discussion

What Next?

- Include human capital and specific inventor types
- Empirically:
    - Look at new inventors in each year as a function of concentration
- Quantitative exploration in more sophisticated model

Sector-Specific Parameter Values

Back

Search Bloomberg Sign In

Deals

Big Tech Swallows Most of the Hot AI Startups

An acquisition spree by Apple, Amazon, Facebook, Google and Microsoft eliminated potential rivals and concentrated brain power in this critical field.

The New York Times https://www.nytimes.com/2017/10/22/technology/artificial-intelligence-experts-salaries.html

Tech Giants Are Paying Huge Salaries for Scarce A.I. Talent

Sector-Specific Parameter Values

Back

- For all pairs NAICS 4-d sectors:
  - Build cosine similarity between descriptions
- For each NAICS 4-d without missing data:
  - Rank 5 most similar sectors with regulation data
  - Attribute regulations of top 5 most similar sectors, weighted by cos.similarity
  - If highest cos-similarity is smaller than .2, use only most similar sector.

Main Specification: Robustness to Outliers

Back

	Ch. 4d K.M. Eff. Inv. Share (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
Ch. HHI lower bound	26.093*	22.509*	25.904*	22.716*	26.111*	22.554*
	(10.696)	(10.848)	(11.124)	(10.948)	(10.725)	(11.019)
Ch. Log Real Sales	0.914**	0.548*	0.881**	0.539*	0.918**	0.562*
	(0.278)	(0.243)	(0.275)	(0.242)	(0.283)	(0.261)
4D Knowledge Market FE		✓		✓		✓
Sample	Full Sample	Full Sample	Trim Outliers	Trim Outliers	Mahalanobis 5%	Mahalanobis 5%
Weight	Sales	Sales	Sales	Sales	Sales	Sales
Observations	157	153	155	152	150	139

Robustness to Individual Firm Size

Back

	Ch. 4d K.M. Eff. Inv. Share (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
Ch. HHI lower bound	35.230**	20.783+	35.230**	20.783+	35.154**	22.854*
	(12.759)	(10.615)	(12.759)	(10.615)	(12.647)	(11.197)
Ch. Log Real Sales per company	0.175	-0.040	0.175	-0.040	0.300	-0.055
	(0.382)	(0.253)	(0.382)	(0.253)	(0.460)	(0.346)
4D Knowledge Market FE		✓		✓		✓
Sample	Full Sample	Full Sample	Trim Outliers	Trim Outliers	Mahalanobis 5%	Mahalanobis 5%
Weight	Sales	Sales	Sales	Sales	Sales	Sales
Observations	81	79	81	79	75	67