Stanford Big-Data Initiative in International Macro-Finance

Text-Based Methods

Tarek Hassan, Associate Professor, Boston University
Department of Economics

Job Postings and Patents

Monday, August 31, 2020
Text-based Methods in Macro

Module 4:
Job postings and patents
Readings

Agenda

• Construct text-based measures of exposure to 20 different technologies at the firm, patent, and job-level, 2002-19.
• Use these novel data to study the spread of new technologies across firms, regions, occupations, and skill-levels.
Five Key Findings on New Technologies

1. Initial hiring is concentrated in high-skilled jobs and geographically.

2. Over time skill level in new jobs declines sharply: “skill broadening”.

3. Over time new jobs spread out across the US: “region broadening”.

4. Geographic hub retains an advantage, particularly for high skilled jobs.

5. Tech hubs are more likely around universities and educated areas.
Relates to a Large Prior Literature

- Clustering in entrepreneurial activity and innovation (Gompers et al. 2005, Glaeser et al., 2015, Moretti 2019);
- Diffusion of new technologies (Griliches 1959, Popp 2002, Acemoglu and Linn 2004, Hall 2006, Moscona 2019);
- Effects of research and entrepreneurial finance (Bronzini and Iachini 2014, Criscuolo et al. 2017, Howell 2017); and
- Earnings call and news text as data (Baker et al. 2016, Hassan et al. 2019).
Outline

1. Define Technologies
2. Exposure Measures
3. Key Employment Patterns
4. The Geographic Evolution of Nascent Technologies
5. Academia and Technology Hubs
Data Sources

1. Full text of USPTO patents (1976-2016)
   • Typically follow a research paper format – invention title, abstract, claim, description.

2. Transcripts of Earnings Conference Calls (2002-19)
   • Discussions of quarterly earnings by 12k publicly listed firms.
   • Typically contains management presentation followed by analyst Q & A.

   • Scraped from job forums (Eg: Glassdoor.com) and employer websites.
   • Geo-coded and assigned to SOC Codes

4. Full text from about 2000 US Newspapers from Newsbank
Step 1: Define Technology Keywords from Patents

Begin by identifying two-word combinations (bigrams) that are indicative of discussion of novel technologies.

1. Extract all (17 mil+) bigrams from most cited US patents for top 30 technology classes and years (1976-2016) -- 89,218 patents.
2. Keep only 15,716 ‘technical bigrams’ by removing common words and phrases.
   a. Remove stop words (eg: ‘of’)
   b. Remove common phrases using Corpus of Historical American English: a collection of text in common use (fiction, non-fiction books, etc.) from 1910-2010.
3. Cross-reference with full text of earnings conference calls and sort by frequency of occurrence in conference calls.
<table>
<thead>
<tr>
<th>Bigram</th>
<th># ECs</th>
<th>Bigram</th>
<th># ECs</th>
<th>Bigram</th>
<th># ECs</th>
</tr>
</thead>
<tbody>
<tr>
<td>floating rate</td>
<td>7,824</td>
<td>combined cycle</td>
<td>1,786</td>
<td>product support</td>
<td>1,229</td>
</tr>
<tr>
<td>renewable energy</td>
<td>6,267</td>
<td>wind energy</td>
<td>1,734</td>
<td>data warehouse</td>
<td>1,076</td>
</tr>
<tr>
<td>sheet items</td>
<td>5,045</td>
<td>data storage</td>
<td>1,700</td>
<td>printed circuit</td>
<td>1,018</td>
</tr>
<tr>
<td>pass code</td>
<td>4,831</td>
<td>mobile platform</td>
<td>1,636</td>
<td>disk drives</td>
<td>1,012</td>
</tr>
<tr>
<td>cell phone</td>
<td>4,320</td>
<td>drug delivery</td>
<td>1,619</td>
<td>mobile computing</td>
<td>965</td>
</tr>
<tr>
<td>burn rate</td>
<td>4,053</td>
<td>fiber optic</td>
<td>1,527</td>
<td>virtual reality</td>
<td>928</td>
</tr>
<tr>
<td>user interface</td>
<td>3,479</td>
<td>data package</td>
<td>1,442</td>
<td>wind turbine</td>
<td>912</td>
</tr>
<tr>
<td>growth mode</td>
<td>3,119</td>
<td>disk drive</td>
<td>1,421</td>
<td>building product</td>
<td>893</td>
</tr>
<tr>
<td>debit card</td>
<td>2,756</td>
<td>knowledge base</td>
<td>1,361</td>
<td>flow stream</td>
<td>871</td>
</tr>
<tr>
<td>solid organic</td>
<td>2,582</td>
<td>sheet structure</td>
<td>1,360</td>
<td>cable television</td>
<td>857</td>
</tr>
<tr>
<td>base stations</td>
<td>2,577</td>
<td>monoclonal antibody</td>
<td>1,296</td>
<td>processing system</td>
<td>816</td>
</tr>
<tr>
<td>individual segments</td>
<td>2,194</td>
<td>fiber network</td>
<td>1,280</td>
<td>computer system</td>
<td>803</td>
</tr>
<tr>
<td>energy management</td>
<td>1,936</td>
<td>video game</td>
<td>1,266</td>
<td>solar cell</td>
<td>786</td>
</tr>
<tr>
<td>flat panel</td>
<td>1,864</td>
<td>network performance</td>
<td>1,237</td>
<td>core component</td>
<td>777</td>
</tr>
</tbody>
</table>
Step 2: Define Technology Keywords

Select bigrams that unambiguously reflect specific technological advances that changed the way businesses operate.

<table>
<thead>
<tr>
<th>bigram</th>
<th># ECs</th>
<th>bigram</th>
<th># ECs</th>
<th>bigram</th>
<th># ECs</th>
</tr>
</thead>
<tbody>
<tr>
<td>floating rate</td>
<td>7,824</td>
<td>combined cycle</td>
<td>1,786</td>
<td>product support</td>
<td>1,229</td>
</tr>
<tr>
<td>renewable energy</td>
<td>6,267</td>
<td>wind energy</td>
<td>1,734</td>
<td>data warehouse</td>
<td>1,076</td>
</tr>
<tr>
<td>sheet items</td>
<td>5,045</td>
<td>data storage</td>
<td>1,700</td>
<td>printed circuit</td>
<td>1,018</td>
</tr>
<tr>
<td>pass code</td>
<td>4,831</td>
<td>mobile platform</td>
<td>1,636</td>
<td>disk drives</td>
<td>1,012</td>
</tr>
<tr>
<td>cell phone</td>
<td>4,320</td>
<td>drug delivery</td>
<td>1,619</td>
<td>mobile computing</td>
<td>965</td>
</tr>
<tr>
<td>burn rate</td>
<td>4,053</td>
<td>fiber optic</td>
<td>1,527</td>
<td>virtual reality</td>
<td>928</td>
</tr>
<tr>
<td>user interface</td>
<td>3,479</td>
<td>data package</td>
<td>1,442</td>
<td>wind turbine</td>
<td>912</td>
</tr>
<tr>
<td>growth mode</td>
<td>3,119</td>
<td>disk drive</td>
<td>1,421</td>
<td>building product</td>
<td>893</td>
</tr>
<tr>
<td>debit card</td>
<td>2,756</td>
<td>knowledge base</td>
<td>1,361</td>
<td>flow stream</td>
<td>871</td>
</tr>
<tr>
<td>solid organic</td>
<td>2,582</td>
<td>sheet structure</td>
<td>1,360</td>
<td>cable television</td>
<td>857</td>
</tr>
<tr>
<td>base stations</td>
<td>2,577</td>
<td>monoclonal antibody</td>
<td>1,296</td>
<td>processing system</td>
<td>816</td>
</tr>
<tr>
<td>individual segments</td>
<td>2,194</td>
<td>fiber network</td>
<td>1,280</td>
<td>computer system</td>
<td>803</td>
</tr>
<tr>
<td>energy management</td>
<td>1,936</td>
<td>video game</td>
<td>1,266</td>
<td>solar cell</td>
<td>786</td>
</tr>
<tr>
<td>flat panel</td>
<td>1,864</td>
<td>network performance</td>
<td>1,237</td>
<td>core component</td>
<td>777</td>
</tr>
</tbody>
</table>
Step 3: Group Selected Keywords into 20 Technologies

<table>
<thead>
<tr>
<th>bigram</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>flat panel</td>
<td>lcds</td>
</tr>
<tr>
<td>fiber optic</td>
<td>optical fiber</td>
</tr>
<tr>
<td>disk drive</td>
<td>disk drive</td>
</tr>
<tr>
<td>monoclonal antibody</td>
<td>monoclonal antibody</td>
</tr>
<tr>
<td>fiber network</td>
<td>optical fiber</td>
</tr>
<tr>
<td>disk drives</td>
<td>disk drive</td>
</tr>
<tr>
<td>mobile computing</td>
<td>cloud</td>
</tr>
<tr>
<td>virtual reality</td>
<td>virtual reality</td>
</tr>
<tr>
<td>solar cell</td>
<td>solar</td>
</tr>
<tr>
<td>panel display</td>
<td>lcds</td>
</tr>
<tr>
<td>monoclonal antibodies</td>
<td>monoclonal antibody</td>
</tr>
<tr>
<td>fiber optics</td>
<td>optical fiber</td>
</tr>
<tr>
<td>touch screen</td>
<td>touch screen</td>
</tr>
<tr>
<td>coalbed methane</td>
<td>fracking</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>bigram</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>smart card</td>
<td>smart card</td>
</tr>
<tr>
<td>solar panel</td>
<td>solar</td>
</tr>
<tr>
<td>hydraulic fracturing</td>
<td>fracking</td>
</tr>
<tr>
<td>hybrid electric</td>
<td>electric cars</td>
</tr>
<tr>
<td>optical fiber</td>
<td>optical fiber</td>
</tr>
<tr>
<td>additive manufacturing</td>
<td>3dprinting</td>
</tr>
<tr>
<td>fiber networks</td>
<td>optical fiber</td>
</tr>
<tr>
<td>machine vision</td>
<td>machine vision</td>
</tr>
<tr>
<td>cellular phone</td>
<td>smartphone</td>
</tr>
<tr>
<td>unmanned aerial</td>
<td>drone</td>
</tr>
<tr>
<td>solar modules</td>
<td>solar</td>
</tr>
<tr>
<td>global positioning</td>
<td>gps</td>
</tr>
<tr>
<td>solar module</td>
<td>solar</td>
</tr>
<tr>
<td>autonomous vehicle</td>
<td>driverless</td>
</tr>
</tbody>
</table>
Step 4: Iterate to Reduce Measurement Error

1. False Negative: Use embedding vectors trained on earnings calls to obtain additional keywords (synonyms) for each technology.
   a. Embedding vectors uses neighboring bigrams (context) to obtain similarity between two bigrams. (Eg: ‘synonyms’ to artificial intelligence – ‘machine learning’, ‘deep learning’)

2. False Positive: Perform human audit of Earnings snippets to rule out false positives, and include additional keywords.

<table>
<thead>
<tr>
<th>Firm Name</th>
<th>EC date</th>
<th>Excerpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy Conversion Devices</td>
<td>Sep-2002</td>
<td>that nancy bacon ecd senior vice president michael because <strong>united solar system</strong> is consolidated into ecd those are the sales that</td>
</tr>
<tr>
<td>Ferro Corporation</td>
<td>May-2007</td>
<td>inventory this decline was offset by the growing demand from <strong>solar cell</strong> manufacturers and demand for our surface finishing materials as</td>
</tr>
</tbody>
</table>
## Final List of Technology Keywords

<table>
<thead>
<tr>
<th>technology</th>
<th>keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>3dprinting</td>
<td>3d printing, 3d printed, 3d printer, 3d printers, 3d print, additive manufacturing, inkjet bioprinting, binder jetting, rapid prototyping</td>
</tr>
<tr>
<td>ai</td>
<td>artificial intelligence, machine learning, neural networks, deep learning, predictive analytics, language processing</td>
</tr>
<tr>
<td>fracking</td>
<td>fracking*, horizontal drilling, coalbed methane, unconventional gas, hydraulic fracturing, tight sandstones, shale rock, shale gas, shale oil, unconventional shale, oil shale, gas sands, shale plays, gas shale, hydraulic fracturing, fracturing activity</td>
</tr>
<tr>
<td>driverless</td>
<td>autonomous car, autonomous cars, selfdriving car, selfdriving cars, selfdriving truck, selfdriving trucks, autonomous trucks, autonomous vehicles, selfdriving vehicles, autonomous driving, autonomous vehicle, automated driving, driverless car, driverless cars, driverless vehicle, driverless vehicles</td>
</tr>
<tr>
<td>virtual reality</td>
<td>virtual reality, augmented reality, mixed reality</td>
</tr>
</tbody>
</table>

* all bigrams containing these unigrams are included.
Outline

1. Define Technologies
2. Exposure Measures
3. Key Employment Patterns
4. The Geographic Evolution of Nascent Technologies
5. Academia and Technology Hubs
Technology Exposure

1. Measure job and firm-level technology exposure as

   \[ \text{exposure}_{i,\zeta,t} = 1\{b_{\zeta} \in D_{i,t}\}, \]

where \( D_{i,t} \) is the set of bigrams contained in a job posting/earnings call posted at time \( t \) and \( b_{\zeta} \) is a bigram associated with technology \( \zeta \).
Example Job Exposed to Virtual Reality

Job Sample (Virtual Reality)

Sample Job Number: 38452265183
JobTitle
3D Artist
JobDomain
losangeles.craigslist.org

Job/Text
OUR COMPANY is a leading, digital creative agency. We are best known for transforming complex subject matter into compelling, interactive stories for our Fortune 100 and 500 clients. Our diverse team of artists, producers, programmers, and thinkers continuously develop pioneering, industry leading and award-winning content. In 2015 alone, the team brought home three Telly Awards Explore the World of Audi, Boeing Advanced F15 and ExxonMobil LNG Value Chain animation. Our interactives are stunning and utilize the latest in digital technology. We love what we do and we love collaborating across disciplines with some of the best of best. Every day brings new, exciting work: from immersive environments to virtual and

—AUGMENTED REALITY— to interactive murals and apps. We revel in pushing creative limits and always strive to deliver above and beyond.

POSITION
The 3D Generalist with minimum 3 years of experience in Video Game Cinematics, TV or Film. Must be proficient in 3DS Max, VRay or Lightwave or Maya. Job requires the ability to create highly photorealistic animation of characters, aircraft, vehicles and environments that match our highest standards.

QUALIFICATIONS
Minimum 3 years of experience as an animator using 3D Studio Max, Lightwave or Maya. Must be able to work on all aspects of a shot: modeling, texturing, lighting, cinematography, animation, rendering. Knowledge of character animation, rigging. Knowledge of Vue a plus. Knowledge of compositing, and color correct shots using AfterEffects or Fusion. Experience with 3rd party rendering tools, VRAY, Final Render. Must be able to manage render times while maintaining a high level of quality. Must be customerfocused, detail-oriented, and work well in a team environment. Must be resourceful with excellent followthrough skills. Must have strong multitasking ability. Working knowledge of interactive media e.g. display technologies, video, 2D/3D animation and interactive software. Experience working collaboratively with crossdisciplinary, art, programming, production and technology teams. Experience working on short timeline, fastpaced projects. Must be a strong leader and teammate. Must be punctual, reliable, highly flexible, and excellent at balancing multiple detail oriented projects. Must have excellent presentation and communication skills written, oral, internal, and clientfacing.
Technology Exposure

1. Measure job and firm-level technology exposure as

\[ \text{exposure}_{i, \zeta, t} = 1 \{ b_\zeta \in D_{i, t} \}, \]

where \( D_{i, t} \) is the set of bigrams contained in a job posting/earnings call posted at time \( t \) and \( b_\zeta \) is a bigram associated with technology \( \zeta \).

2. Aggregate over various documents

\[ \text{share exposed}_{a, \zeta, t} = \frac{\sum_{i \in a, t} 1 \{ b_\zeta \in D_{i, t} \}}{\sum_{i \in a, t} 1 \{ D_{i, t} \}} \]

where \( a \) may be a firm, sector, urban region or occupation, and \( t \) is time.
## Example: Occupations Exposed to Virtual Reality

### Top Exposed Occupations to ‘Virtual Reality’

<table>
<thead>
<tr>
<th>SOCName</th>
<th>Exposed Jobs</th>
<th>Share Exposed (Pct)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Hardware Engineers</td>
<td>1000</td>
<td>1</td>
</tr>
<tr>
<td>Fine Artists, Including Painters, Sculptors, and Illustrators</td>
<td>658</td>
<td>1</td>
</tr>
<tr>
<td>Multimedia Artists and Animators</td>
<td>607</td>
<td>0.8</td>
</tr>
<tr>
<td>Computer and Information Research Scientists</td>
<td>1630</td>
<td>0.7</td>
</tr>
<tr>
<td>Art Directors</td>
<td>422</td>
<td>0.5</td>
</tr>
<tr>
<td>Sound Engineering Technicians</td>
<td>140</td>
<td>0.5</td>
</tr>
<tr>
<td>Interior Designers</td>
<td>382</td>
<td>0.4</td>
</tr>
<tr>
<td>Producers and Directors</td>
<td>576</td>
<td>0.4</td>
</tr>
<tr>
<td>Astronomers</td>
<td>45</td>
<td>0.4</td>
</tr>
<tr>
<td>Computer Science Teachers, Postsecondary</td>
<td>134</td>
<td>0.4</td>
</tr>
</tbody>
</table>
Share Firms Exposed & Technology Start Date

1. Measure the share of earnings calls mentioning a technology
Share Firms Exposed & Technology Start Date

1. Measure the share of earnings calls mentioning a technology
2. Define a “technology start date” as the time when business start taking note of the technology as measured in earnings conference calls; the year in which the time series first attains at least 10% of its maximum.

![Graph showing virtual reality over time](image-url)
Share Exposed Firms and Jobs – Corr. 68%
Outline

1. Define Technologies
2. Exposure Measures
3. **Key Employment Patterns**
4. The Geographic Evolution of Nascent Technologies
5. Academia and Technology Hubs
Skill Broadening of Jobs

Over the life cycle of a technology measure the skill component using SOC codes as

\[ Skill_t^\varsigma = \frac{\sum_o N_{o;\tau} x_{o;2015}}{\sum_o N_{o;\tau}} \]

where \( o \) is a Census SOC code, \( N_{o;\tau} \) is the number of postings exposed to technology \( \tau \) and SOC code \( o \) at time \( t \), and \( x_{o;2015} \) is the average skill level for SOC \( o \), as measured by the 2015 ACS sample. Plot this skill component for each technology over time.
Skill Broadening – Share college educated
Skill Broadening -- Regressions

\[ \text{Skill}^\varsigma_t = \alpha_0 + \text{year since start}^\varsigma_t + \delta_\varsigma + \delta_t + \epsilon_{\varsigma,t} \]

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>share college educated (pct.)</td>
<td>share phd (pct.)</td>
<td>share post graduate (pct.)</td>
<td>avg. wage</td>
<td>avg. schooling</td>
</tr>
<tr>
<td>years since start</td>
<td>-1.243*** (0.245)</td>
<td>-0.0698*** (0.0167)</td>
<td>-0.493*** (0.0963)</td>
<td>-1,248*** (248.7)</td>
<td>-0.0654*** (0.0130)</td>
</tr>
<tr>
<td>Observations</td>
<td>203</td>
<td>203</td>
<td>203</td>
<td>203</td>
<td>203</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.833</td>
<td>0.971</td>
<td>0.892</td>
<td>0.835</td>
<td>0.854</td>
</tr>
<tr>
<td>Tech FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Year FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

Standard Errors are clustered by technology.
Outline

1. Define Technologies
2. Exposure Measures
3. Key Employment Patterns
4. The Geographic Evolution of Nascent Technologies
5. Academia and Technology Hubs
Region Broadening

For each region, define

\[
Normalized \ share_{cbsa,\tau,t} = \frac{share\ jobs\ expoed_{cbsa,\tau,t}}{share\ jobs\ expoed_{\tau,t}}
\]

Calculate the coefficient of variation in this normalized share to show how technology jobs spread over time.
Regional Broadening over Time
## Regional Broadening – Regressions

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N share top</td>
<td>SD</td>
<td>CV</td>
</tr>
<tr>
<td>years since</td>
<td>-0.150***</td>
<td>-0.0245***</td>
<td>-0.0625**</td>
</tr>
<tr>
<td>start</td>
<td>(0.0466)</td>
<td>(0.00503)</td>
<td>(0.0219)</td>
</tr>
<tr>
<td>Observations</td>
<td>203</td>
<td>203</td>
<td>203</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.835</td>
<td>0.878</td>
<td>0.929</td>
</tr>
<tr>
<td>Tech FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Year FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

Standard errors are clustered by technology.
Regional Broadening Concentrated in Low-Skill Jobs
### Regional Broadening Concentrated in Low-Skill Jobs (2)

<table>
<thead>
<tr>
<th>Panel A: Low Skill</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normalized share top 1%</td>
<td>SD</td>
<td>CV</td>
</tr>
<tr>
<td>years since start</td>
<td>-0.520***</td>
<td>-0.0727***</td>
<td>-0.106***</td>
</tr>
<tr>
<td></td>
<td>(0.0732)</td>
<td>(0.00701)</td>
<td>(0.0264)</td>
</tr>
<tr>
<td>Observations</td>
<td>202</td>
<td>202</td>
<td>200</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.832</td>
<td>0.831</td>
<td>0.891</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: High Skill</th>
<th>(3)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>years since start</td>
<td>-0.107</td>
<td>-0.0192*</td>
<td>-0.0532*</td>
</tr>
<tr>
<td></td>
<td>(0.0777)</td>
<td>(0.0107)</td>
<td>(0.0274)</td>
</tr>
<tr>
<td>Observations</td>
<td>203</td>
<td>203</td>
<td>202</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.866</td>
<td>0.890</td>
<td>0.945</td>
</tr>
</tbody>
</table>

Standard Errors are robust.
Regional Broadening Concentrated in Low-Skill Jobs

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normalized</td>
<td>SD</td>
<td>CV</td>
</tr>
<tr>
<td>share top 1%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(year since t0) * (skill == low)</td>
<td>-0.388***</td>
<td>-0.0552***</td>
<td>-0.0527***</td>
</tr>
<tr>
<td>(year since t0)</td>
<td>-0.134*</td>
<td>-0.0214**</td>
<td>-0.0568***</td>
</tr>
<tr>
<td></td>
<td>(0.0677)</td>
<td>(0.00920)</td>
<td>(0.0134)</td>
</tr>
<tr>
<td></td>
<td>(0.0682)</td>
<td>(0.0107)</td>
<td>(0.0210)</td>
</tr>
<tr>
<td>Observations</td>
<td>405</td>
<td>405</td>
<td>402</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.756</td>
<td>0.786</td>
<td>0.879</td>
</tr>
<tr>
<td>Tech FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Skill FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Year FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

Standard Errors are robust.
## Technology Hubs

Define a technology hub as locations which account for 80% of technology job postings in the starting year.

<table>
<thead>
<tr>
<th>CBSA</th>
<th>State</th>
<th>Share (t0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>San Francisco-Oakland-Hayward</td>
<td>CA</td>
<td>13%</td>
</tr>
<tr>
<td>New York-Newark-Jersey City</td>
<td>NY-NJ-P A</td>
<td>11%</td>
</tr>
<tr>
<td>San Jose-Sunnyvale-Santa Clara</td>
<td>CA</td>
<td>9%</td>
</tr>
<tr>
<td>Seattle-Tacoma-Bellevue</td>
<td>WA</td>
<td>8%</td>
</tr>
<tr>
<td>Boston-Cambridge-Newton</td>
<td>MA-NH</td>
<td>6%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CBSA</th>
<th>State</th>
<th>Share (t0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Houston-The Woodlands-Sugar Land</td>
<td>TX</td>
<td>32%</td>
</tr>
<tr>
<td>Dallas-Fort Worth-Arlington</td>
<td>TX</td>
<td>19%</td>
</tr>
<tr>
<td>Denver-Aurora-Lakewood</td>
<td>CO</td>
<td>9%</td>
</tr>
<tr>
<td>Los Angeles-Long Beach-Anaheim</td>
<td>CA</td>
<td>3%</td>
</tr>
<tr>
<td>Washington-Arlington-Alexandria</td>
<td>DC-VA-MD-WV</td>
<td>2%</td>
</tr>
</tbody>
</table>

### Smartphone

<table>
<thead>
<tr>
<th>CBSA</th>
<th>State</th>
<th>Share (t0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>San Jose-Sunnyvale-Santa Clara</td>
<td>CA</td>
<td>31%</td>
</tr>
<tr>
<td>San Francisco-Oakland-Hayward</td>
<td>CA</td>
<td>8%</td>
</tr>
<tr>
<td>Los Angeles-Long Beach-Anaheim</td>
<td>CA</td>
<td>8%</td>
</tr>
<tr>
<td>Seattle-Tacoma-Bellevue</td>
<td>WA</td>
<td>6%</td>
</tr>
<tr>
<td>New York-Newark-Jersey City</td>
<td>NY-NJ-P A</td>
<td>5%</td>
</tr>
</tbody>
</table>

### Driverless Cars

<table>
<thead>
<tr>
<th>CBSA</th>
<th>State</th>
<th>Share (t0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detroit-Warren-Dearborn</td>
<td>MI</td>
<td>18%</td>
</tr>
<tr>
<td>San Jose-Sunnyvale-Santa Clara</td>
<td>CA</td>
<td>17%</td>
</tr>
<tr>
<td>Chicago-Naperville-Elgin</td>
<td>IL-IN-WI</td>
<td>13%</td>
</tr>
<tr>
<td>San Francisco-Oakland-Hayward</td>
<td>CA</td>
<td>6%</td>
</tr>
<tr>
<td>San Diego-Carlsbad</td>
<td>CA</td>
<td>4%</td>
</tr>
</tbody>
</table>
## Persistent Advantage of Technology Hubs

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Normalized share technology jobs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tech hub</td>
<td>0.565***</td>
<td>0.883***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0340)</td>
<td>(0.0614)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tech hub*(year since t0))</td>
<td>-0.0261***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00419)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ihs(IDW(tech hub))</td>
<td></td>
<td>0.216***</td>
<td>0.201***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0357)</td>
<td>(0.0452)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00115</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.00224)</td>
</tr>
<tr>
<td>Observations</td>
<td>140,537</td>
<td>140,537</td>
<td>134,092</td>
<td>134,092</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.171</td>
<td>0.173</td>
<td>0.112</td>
<td>0.112</td>
</tr>
<tr>
<td>Tech FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Year FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Year since FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>CBSA FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

Standard Errors are robust
### Advantage of Technology Hubs Particularly in High-Skill Jobs

<table>
<thead>
<tr>
<th></th>
<th>Panel A: Low Skill</th>
<th>Panel B: High Skill</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normalized Share</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tech jobs</td>
<td></td>
</tr>
<tr>
<td>tech hub</td>
<td>0.449***</td>
<td>0.660***</td>
</tr>
<tr>
<td></td>
<td>(0.0388)</td>
<td>(0.0438)</td>
</tr>
<tr>
<td>tech hub*(year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>since t0)</td>
<td>-0.0405***</td>
<td>-0.0151***</td>
</tr>
<tr>
<td></td>
<td>(0.00540)</td>
<td>(0.00451)</td>
</tr>
<tr>
<td>Observations</td>
<td>169,099</td>
<td>169,099</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.085</td>
<td>0.187</td>
</tr>
</tbody>
</table>

Standard Errors are robust
Outline

1. Define Technologies
2. Exposure Measures
3. Key Employment Patterns
4. The Geographic Evolution of Nascent Technologies
5. Academia and Technology Hubs
Where are Technology Hubs?
## Where are Technology Hubs? (2)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normalized Share Tech Jobs at t0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(1 + # research unis)</td>
<td>0.284***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0279)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(1+ R and D uni spend)</td>
<td></td>
<td>0.0314***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.00298)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share College Educated</td>
<td></td>
<td></td>
<td>2.852***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.260)</td>
<td></td>
</tr>
<tr>
<td>Share post graduate</td>
<td></td>
<td></td>
<td></td>
<td>6.496***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.649)</td>
</tr>
<tr>
<td>Observations</td>
<td>18,531</td>
<td>18,531</td>
<td>18,340</td>
<td>18,340</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.049</td>
<td>0.050</td>
<td>0.055</td>
<td>0.055</td>
</tr>
<tr>
<td>Tech FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

Standard Errors are clustered by CBSA.
Conclusion

Introduced flexible measures of technology exposure at the firm, job, occupation, and region-level. Find:

1. Initial hiring in new technologies is highly concentrated in high-skilled jobs.
2. Overtime the average skill level in new jobs declines “skill broadening”
3. New hiring also increases its geographic footprint “region broadening”
4. The initial tech hub retains an advantage, particularly for high skilled jobs.
5. Technology hubs most likely around universities and educated areas
## Top bigrams in most cited patents

<table>
<thead>
<tr>
<th>Bigram</th>
<th>patents</th>
<th>Bigram</th>
<th>patents</th>
<th>Bigram</th>
<th>patents</th>
</tr>
</thead>
<tbody>
<tr>
<td>electrically conductive</td>
<td>7548</td>
<td>pulse width</td>
<td>5415</td>
<td>conductive layer</td>
<td>4128</td>
</tr>
<tr>
<td>fluid communication</td>
<td>7415</td>
<td>conductive material</td>
<td>5027</td>
<td>silicon nitride</td>
<td>4092</td>
</tr>
<tr>
<td>user interface</td>
<td>7064</td>
<td>temperature sensor</td>
<td>5026</td>
<td>inlet port</td>
<td>4043</td>
</tr>
<tr>
<td>reaction zone</td>
<td>6951</td>
<td>dielectric layer</td>
<td>5016</td>
<td>feed stream</td>
<td>4025</td>
</tr>
<tr>
<td>vapor deposition</td>
<td>6344</td>
<td>image data</td>
<td>4971</td>
<td>output signals</td>
<td>3954</td>
</tr>
<tr>
<td>host computer</td>
<td>6014</td>
<td>matrix material</td>
<td>4587</td>
<td>subterranean formation</td>
<td>3922</td>
</tr>
<tr>
<td>integrated circuit</td>
<td>5999</td>
<td>fiber optic</td>
<td>4519</td>
<td>block copolymers</td>
<td>3834</td>
</tr>
<tr>
<td>substrate surface</td>
<td>5986</td>
<td>metal layer</td>
<td>4457</td>
<td>poly ethylene</td>
<td>3823</td>
</tr>
<tr>
<td>polymeric material</td>
<td>5772</td>
<td>drug delivery</td>
<td>4247</td>
<td>display system</td>
<td>3820</td>
</tr>
<tr>
<td>duty cycle</td>
<td>5643</td>
<td>adhesive layer</td>
<td>4200</td>
<td>specific binding</td>
<td>3797</td>
</tr>
</tbody>
</table>