NBER WORKING PAPER SERIES

THE DIFFUSION OF DISRUPTIVE TECHNOLOGIES

Nicholas Bloom
Tarek Alexander Hassan
Aakash Kalyani
Josh Lerner
Ahmed Tahoun

Working Paper 28999
http://www.nber.org/papers/w28999

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2021

We thank audiences at the Applied Machine Learning, Economics, and Data Science webinar, Bocconi University, Columbia University, Dartmouth University, the London School of Economics, Nova Business School, the Toulouse Network on Information Technology, the University of Maryland, and Yeshiva University for helpful comments, Lisa Kahn for data sharing, and William Hartog and Jared Simpson for excellent research assistance. We thank Scarlett Chen, Nick Short, Corinne Stephenson, and Michael Webb for assistance in conceptualizing and researching early versions of this project. Funding for this research was provided by the Institute for New Economic Thinking, the Ewing Marion Kauffman Foundation, Harvard Business School's Division of Research and Doctoral Programs, the Alfred P. Sloan Foundation, and the Toulouse Network on Information Technology. Lerner has received compensation from advising institutional investors in venture capital funds, venture capital groups, and governments designing policies relevant to venture capital. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Diffusion of Disruptive Technologies
Nicholas Bloom, Tarek Alexander Hassan, Aakash Kalyani, Josh Lerner, and Ahmed Tahoun
NBER Working Paper No. 28999
July 2021
JEL No. O31,O32

ABSTRACT

We identify novel technologies using textual analysis of patents, job postings, and earnings calls. Our approach enables us to identify and document the diffusion of 29 disruptive technologies across firms and labor markets in the U.S. Five stylized facts emerge from our data. First, the locations where technologies are developed that later disrupt businesses are geographically highly concentrated, even more so than overall patenting. Second, as the technologies mature and the number of new jobs related to them grows, they gradually spread across space. While initial hiring is concentrated in high-skilled jobs, over time the mean skill level in new positions associated with the technologies declines, broadening the types of jobs that adopt a given technology. At the same time, the geographic diffusion of low-skilled positions is significantly faster than higher-skilled ones, so that the locations where initial discoveries were made retain their leading positions among high-paying positions for decades. Finally, these technology hubs are more likely to arise in areas with universities and high skilled labor pools.

Nicholas Bloom
Stanford University
Department of Economics
579 Serra Mall
Stanford, CA 94305-6072
and NBER
nbloom@stanford.edu

Tarek Alexander Hassan
Department of Economics
Boston University
270 Bay State Road
Boston, MA 02215
and NBER
thassan@bu.edu

Aakash Kalyani
Department of Economics
Boston University
270 Bay State Road
Boston, MA 02215
aakashk@bu.edu

Josh Lerner
Harvard Business School
Rock Center 214
Soldiers Field
Boston, MA 02163
and NBER
jlerner@hbs.edu

Ahmed Tahoun
London Business School
26 Sussex plc, Regent's Park
London NW1 4SA
United Kingdom
atahoun@london.edu

The dataset constructed as part of this paper, as well as relevant code is available at
www.techdiffusion.net

1. Introduction

The development of novel technologies, the degree to which they affect jobs, and the speed with which they spread across regions, firms, and industries are key elements in the study of economic growth, economic inequality, entrepreneurship, and firm dynamics. Many authors have sought to understand whether the benefits from the adoption of new technologies accrue primarily to inventors, early investors, highly skilled users, or to society more widely through, for instance, employment growth.[2] Other studies, as discussed below, have explored the geography of the development and diffusion of new technologies.

One key obstacle to resolving these questions is that it has proven difficult to measure the development and spread of multiple technological advances in a single framework, and to separate those innovations that affect jobs and businesses from those that do not.

In this paper, we make use of the full text of millions of patents and job postings and hundreds of thousands of earnings conference calls over the past two decades to make progress on this challenge. In particular, we develop a flexible methodology that allows us to determine which innovations or sets of innovations ("technologies") affect businesses, trace these back to the locations and firms where they emerged, and track their diffusion through regions, occupations, and industries over time. We then use our newly created data to establish five novel stylized facts about the development and diffusion of disruptive technologies across space, skill levels, and other dimensions.

The first step of our analysis is to develop a methodology for systematically identifying two-word phrases ("technical bigrams") associated with rapidly diffusing, or disruptive, technologies through a series of systematic rules, whose robustness we verify through various diagnostic tests. To this end, we intersect information from two large corpora of text. First, we use the full text of U.S. patents awarded between 1976 and 2016 to isolate two-word combinations that appear in influential patents but were not commonly used elsewhere. That is, we isolate language specific to

[2] See, for example, Katz and Murphy (1992), Goldin and Katz (2008), Autor, Katz, and Kearney (2008), Piketty and Saez (2013), and Song *et al*. (2019).

recent influential innovations. Second, we search for these bigrams in the full text of earnings conference calls held by more than 8,000 listed firms between 2002 and 2020, to identify those technical bigrams that feature prominently in discussions between firm executives and investors during our sample period. This procedure highlights a small set of 305 technical bigrams that describe recent technological advances that have disrupted a large number of businesses in the last two decades. The top three of these are "mobile devices," "machine learning," and "cloud computing."

To aid interpretation, we then group our technical bigrams into sets of technologies, recognizing the fact that, for example, "cloud computing" and "cloud services" refer to closely-related innovations. This approach partly relies on human judgment, aided by machine learning methodologies. Using this "supervised" process, we identify 29 disruptive technologies, which we use for the main analyses in the paper.[3] Taken together, 22% of all patents granted by the U.S. Patent and Trademark Office (USPTO) between 1992 and 2016 are involved with the development of at least one of our 29 technologies. In this sense, our disruptive technologies cover a significant part of recent innovative activity. While we make no claim of completeness, we argue that each of these 29 advances had significant implications for businesses and jobs in the United States in the past two decades.

After establishing our list of new technologies, we then identify patents, earnings calls, and job postings that mention these new technologies. We use patents to identify the locations where each of the technologies were developed and earnings calls to identify exposed firms and the year in which the technology started to feature prominently in the conversations between executives and investors (its commercial breakthrough). We then cross-reference our list of technical bigrams with the full text of online job postings to identify 13 million jobs advertised between 2007 and 2020 that use, produce, or develop our disruptive technologies. These granular data uniquely allow us to track the spread of disruptive technologies along a dimension of crucial importance to policymakers: employment. In particular, we examine the evolution of the number, location, and quality of job postings associated with these new technologies.

[3] In a second, "unsupervised" approach we use all technical bigrams "as is"—i.e., with no human processing. This alternative approach yields qualitatively identical results.

The key results of this analysis are as follows.

First, the locations where disruptive technologies are developed are geographically highly concentrated, both within and across technologies. Based on patenting activity ten years prior to each technology's commercial breakthrough, we show that the typical disruptive technology in our data emerged from only a handful of urban areas, which housed the majority of early patenting in the technology and the vast majority of its early employment up to the year in which it has its commercial breakthrough. We term these specific urban areas the technology's "pioneer locations."

Although 23 of the 50 U.S. states host at least one pioneer location according to this definition, their distribution across technologies is also remarkably skewed: a few super-clusters are the birthplace of a surprising number of disruptive technologies in our data. Collectively, locations in California alone host a remarkable 40.2% of our technology-pioneer location pairs. Another super-cluster along the Northeast Corridor from Washington to Boston accounts for additional 21.2%. More broadly, we find that the geographic distribution of patenting related to our 29 disruptive technologies is even more skewed than that of patenting in general.

Second, despite this highly skewed initial distribution, as technologies mature and the number of new jobs related to them grows, they gradually spread geographically. For example, the coefficient of variation across the 917 core-based statistical areas (CBSAs) in the United States drops by 24% in the first decade after a technology emerges. We see this pattern of "region broadening" in virtually every technology that we examine.

Third, while initial hiring is concentrated on high-skilled jobs, over time, the mean required skill levels of the jobs associated with the technologies declines, broadening the types of jobs that adopt a given technology. For example, the average earnings associated with job postings in a given new technology drop by about 15% within the first decade, falling from $70,468 per year to $60,608 per year on average, a drop of about one thousand dollars per year (all figures in 2015 dollars).

This pattern of an increasing share of low-skilled jobs that begin to use and adopt a given technology holds within most (though not all) of our disruptive technologies.[4]

Fourth, region and skill broadening interact: Low-skill jobs associated with a given technology spread out across space significantly faster than high-skill jobs. Our estimates suggest low-skilled jobs that use or produce new technologies are almost fully dispersed geographically within 20 years. For example, as technologies like the smart phone, cloud computing, and electric cars mature, the lower-skilled jobs associated with these – salespeople, technicians, repair specialists, etc. – spread across the United States at a fast clip.

Fifth, despite region and skill broadening, disruptive technologies appear to yield long-lasting benefits for their pioneer locations, particularly when it comes to high-skill employment. Our estimates imply it takes almost 40 years for high-skilled job postings to fully disperse from their original pioneer locations. Perhaps not surprisingly, these pioneer locations tend to be located around universities and areas with more educated populations. Thus, regions with strong local education, research institutions, and universities appear to benefit from successful disruptive innovation for substantial periods of time.

While the focus of our analysis is on documenting the major stylized facts about the spread of technologies, the granularity of our data also allows us to study the employment dynamics associated with disruptive technologies for individual locations and firms. As an example of such a more micro-focused analysis, we document a case study of the geographic footprints of two large Detroit-based car manufacturers, and how they evolve after the emergence of technologies relating to self-driving cars. In this instance, we show that both large incumbents shifted significant numbers of job postings relating to self-driving cars towards the technology's pioneer locations (particularly Silicon Valley) and away from their traditional hub in Detroit. We speculate that this kind of "re-homing" of established firms may form part of the reason for the long-lasting hiring advantages of pioneer locations.

[4] It should be acknowledged that we can look at the nature of jobs created through Burning Glass, but not the destruction of existing positions.

In the final part of the paper, we look at the generality of our results by studying the diffusion of disruptive technologies across firms, industries, and occupations. We show that similar patterns predominate. While technology-related job postings spread out over time, the original firms, industries, and occupations associated with the development and early employment in the technology retain an advantage over time. Generally, we find a faster spread across locations and firms than industries and occupations.

Our work builds on a large literature that studies the relationship between technology and labor markets. One strand of this literature studies the diffusion of technology. This literature has focused on patterns in a single specific (though important) new technology, from computers (Autor, Levy, and Murnane, 2002) to broadband (Akerman, Gaarder, and Mogstad, 2015) to robots (Acemoglu and Restrepo, 2020) and artificial intelligence (Agrawal, Gans, and Goldfarb, 2019; Webb, 2020).[5] A second strand focuses on specific innovations during important historical episodes. Examples include studies of hybrid corn (Griliches, 1959), electrification (Goldin and Katz, 1998), threshing machines (Caprettini and Voth, 2020) and encyclopedias (Squicciarini and Voigtlander, 2015). Mokyr (1992) and Gordon (2016) trace out the impact on economic development and the standard of living of a range of great inventions. Both of these classes of studies use technology- and industry-specific approaches to measure the diffusion and impact of individual technologies. A third strand examines the impact of technological progress more generally on the labor market, including inputs like research and development spending (e.g. Berman, Bound, and Griliches, 1994; Machin and Van Reenen, 1998; and Aghion *et al*., 2019) and outputs like computerization (e.g., Krueger, 1993; Autor, Katz, and Krueger, 1998; Michaels, Natraj, and Van Reenen, 2014). We contribute to this literature by providing a flexible methodology to systematically isolate those innovations that have a large impact on firms and labor markets, and to track their spread across firms, industries, occupations, and jobs requiring different skill levels. Aside from the 29 disruptive technologies we identify in this paper, variants of our approach could also be used to study the adoption and spread of some of the other specific innovations highlighted by this literature.

[5] This work is related to Comin and Hobijn (2004, 2010) and their associated work. Their 2010 paper, for instance, looks at the diffusion of 15 technologies across 166 countries, employing a variety of measures of technological utilization. The rich data that we are able to exploit allows us to analyze (albeit for one nation and a much shorter time period) the interactions between innovation and employment at the firm level on a temporal and regional basis.

A second broad literature examines regional development, in particular questions relating to the mechanisms behind the continued advantage of pioneer locations. One literature argues these patterns are driven by continued entrepreneurial activity, or localized knowledge spillovers (e.g. Jaffe, Trajtenberg, and Henderson, 1993). Gompers, Lerner, and Scharfstein (2005) posit that entrepreneurial firms in venture-rich regions are particularly likely to spawn other ventures, leading to self-reinforcing cycle of activity. These claims are supported by Glaeser, Kerr, and Kerr (2015), who highlight the extent to which cities' initial endowments affect the long-term distribution of entrepreneurship. Moretti (2019) shows that productivity of inventors is increasing with the volume of invention in the same city, field, and year. These differentials in innovative efficiency may attract a broad set of firms to locate facilities in the initial hub. We contribute to this literature by providing a systematic approach to identifying pioneer locations, characterizing their distribution across the United States, and showing there is a general relationship between successful innovation, early employment in a given new technology, and the long-term advantage that these locations preserve in high-skill employment. Moreover, we hope future work may use our granular data to study in detail the anatomy and evolution of technology hubs.

Finally, our work relates to a broader literature on the diffusion of new technologies. Since the pioneering work of Griliches (1957), the diffusion process has long been understood by economists to be a gradual one. While broader sociological and organizational literature has examined the barriers to innovation, recent work in economics has focused on understanding the importance of supply and demand factors on the speed of diffusion (e.g., Popp, 2002; Acemoglu and Linn, 2004; Greenstone, Hornbeck and Moretti, 2010; Moser, Voena, and Waldinger, 2014; Moscona, 2019; Arora, Belenzon, and Sheer, 2021). Despite this interesting work, Hall's (2006) characterization of the study of diffusion as "a somewhat neglected one in the economics of innovation" still remains a fair observation. Our contribution is to provide a first assessment of the rate at which disruptive technologies spread across locations, firms, occupations, and industries.

The remainder of this paper is structured as follows. In Section 2, we describe the construction of our data. In Section 3, we present our region-broadening and skill-broadening results. In addition, we examine the differential patterns across geographic regions. In Section 4, we examine the association between academia and technology hubs. In Section 5, we examine the diffusion across three other dimensions: industries, occupations, and firms. In Section 6, we investigate a potential mechanism for region broadening: firm rehoming towards pioneer locations. Section 7 shows a number of additional robustness checks.

2. Data Construction

In this paper, we identify a set of recent disruptive technologies. We associate with each new technology a set of business-relevant keywords, which will allow us to identify the evolution of these technologies. In particular, we seek to (a) build a firm-quarter level measure of technology exposure, (b) use this measure to pinpoint when a given technology starts affecting businesses; (c) create a measure of technology adoption at the job-posting level, and (d) aggregate the data (in various ways) to measure technology adoption at the region and firm level. This section describes our approach in more detail.

2.1. List of new technologies and associated keywords

As a first step, we want to identify influential technologies in as systematic manner as possible. We begin by examining U.S. patent filings. Patents are an attractive starting point for our analysis for two reasons. First, they are by definition novel, particularly when we focus on the most influential patents. Second, they must describe their technology and (at least some) key ways in which it is applied.[6] We focus solely on patent awards by the USPTO: because of the importance of the U.S. market, inventors worldwide typically file important discoveries with the USPTO.[7]

In order to obtain set of technical bigrams, we collect all utility patents awarded by the USPTO to either U.S. assignees or inventors between 1976 and 2016, a total of approximately three million awards. From the text of these patents (abstract, summary, claims, and background description)

[6] The reduction to practice enablement requirement refers to the requirement of 35 U.S.C. 112(a).
[7] About half of all patent applications to the USPTO are filed by residents of foreign countries (USPTO, 2020). This pattern reflects the fact that patent protection in a given nation depends critically on having a patent issued in that specific nation. Important discoveries (the focus of our analysis) are disproportionately likely to be filed in major patent offices world-wide (Lanjouw, Pakes, and Putnam, 1997).

we remove stop words (such as "of," "the," and "from") and decompose the remaining text into about 17 million unique two-word combinations ("bigrams"). We focus on bigrams because they are less ambiguous than single-word keywords: while words like "autopilot" or "cloud" could have a variety of colloquial meanings, "autonomous vehicle" and "cloud computing" are much less ambiguous (e.g., Tan, Wang, and Lee, 2002; Bekkerman and Allan, 2004). To undertake this processing, we follow the methodology in papers undertaking textual analyses of patents and earnings calls, such as Kelly *et al.* (forthcoming) and Hassan *et al.* (2019).

At the same time, many of the bigrams collected from the patents are too general for our purposes. Moreover, many of the bigrams, while related to a specific technology, are not frequently encountered in business contexts, being too scientific or technical in nature. We address these concerns in two ways.

First, we narrow the text down to "technical" bigrams by dropping "non-technical" bigrams that were in common use long before the emergence of our disruptive technologies. To this end, we select all text dating prior to 1970 from the Corpus of Historical American English (COHA), a representative sample of text constructed by linguists from prominent fiction and non-fiction sources (Davies, 2009). We then similarly decompose this historical text into bigrams and remove any bigram appearing in this source (for instance, "of the" or "equipment used") from the set of bigrams obtained from patents.

Second, to identify bigrams associated with influential inventions in the remaining list of 1.5 million bigrams, we collect citations for all the patents that mention these bigrams between 1976 and 2016. We normalize the citations to each patent by the mean within each technology class and application year.[8] We then retain bigrams that cumulatively obtain at least 1000 normalized citations. Thus, we would include a phrase that appeared in a thousand average-cited patents, or in a single patent that was one thousand times more cited than its peers.

[8]Citation rates vary considerably over time and across technology classes. Lerner and Seru (forthcoming) document this heterogeneity and the biases that can result from failing to correct properly for these differences.

After these eliminations, we have a list of 35,063 "technical" bigrams in influential patents from 1976 to 2016. Appendix Table 1 lists those bigrams that most frequently appeared in the sample, weighting the count of patents in which they appeared by the number of citations to these patents. As above, each patent count is normalized by the mean of citations within each technology class and year.

In the second step, we focus on which of our technical bigrams figured increasingly into the business discussions of firms, to gauge the extent to which each innovation changed or disrupted how firms operated. Here we use earnings conference calls from publicly listed firms.

2.2. Earnings calls data

Quarterly earnings call transcripts consist of two sections: a presentation by management (typically the chief executive and/or financial officer(s)) and then questions posed by investment analysts with answers provided by the executives. These calls have been shown to be indicators of some of the most important issues facing these organizations (Bushee, Matsumoto, and Miller, 2003; Matsumoto, Prank, and Roelofsen, 2011; Hassan *et al.*, 2019, 2020).

We tabulate the bigrams in 321,373 conference calls held by 11,905 publicly held companies and compiled by Refinitiv EIKON (formerly Thomson Reuters) between 2002 and 2019. Through this examination, we eliminate about 43% of the bigrams from the patents that are never mentioned in these calls.

We trim the remaining bigrams in two ways. First, we require that they appear in more than 100 transcripts, to focus on economically important bigrams associated with innovations that became major topics in earnings discussions. Second, we require these are increasing in their incidence in earnings calls over time, to focus on relatively recent technologies that affect a growing number of firms during our sample period. To do this, we keep bigrams which appear at least ten times as frequently in their peak year as in the first year of the earnings call data in 2002.[9] After these steps,

[9] Bigrams that do not appear at all in 2002 automatically meet this criteria.

we end up with 305 technical bigrams describing technologies which are widely used and rising in importance, which we label as disruptive technologies.

Table 1 shows the 30 technical bigrams most frequently appearing in earnings calls. It shows that our simple two-step approach of cross-referencing bigrams from influential patents with those featuring increasingly in business discussions clearly identifies some of the major disruptive technological advances of the past two decades. The first four bigrams on the list are “mobile devices,” “machine learning,” “cloud computing,” and “cloud services.” Other top-ranking bigrams on the list include “social networking” and “smart grid.”

In order to obtain a coherent set of technologies from our 305 bigrams, we take two approaches. In the first, described in detail below, we manually group the 305 bigrams into a set of technologies, recognizing the fact that, for example, “cloud computing” and “cloud services” refer to related innovations. We apply a number of further refinements, allowing us to quantify the spread of specific technologies along a variety of dimensions. This approach inevitably relies on human judgment, aided by machine learning methodologies. This “supervised” approach is the basis for the analyses presented in the main body of the paper. We describe it in detail below.

An alternative “unsupervised” approach is to use all 305 bigrams “as is”—i.e., with no human processing. We show later on that all of our main results are robust to this approach, both qualitatively and quantitatively. In this sense, all human intervention from this point on serves the purpose of measuring the spread of specific technologies and making our results more easily interpretable, but has no bearing on the validity of our main stylized facts about disruptive technologies as a whole.

In our “supervised” approach, we take four additional steps. First, we eliminate those bigrams (from the list of 305 bigrams) that, in our reading, do not clearly and unambiguously reflect specific technological advances. This approach allows us to eliminate bigrams that refer to problems but not technological solutions, such as “carbon footprint” or “power outage.” Similarly, we drop bigrams referring to older technologies, such as “smart grid,” which refers to a technology that has been available since the 1980s but is enjoying renewed interest in recent years, and “nand flash”

(flash memory), which has a surge of references when a global supply issue occurred. We also drop any bigram that is vague or refers to multiple innovations, such as “flow profile,” which may refer interchangeably to a genomic flow or firms’ cash flows, or “digital channel,” which can refer to interchangeably to digital marketing or digital transmission. At the end of our supervised approach, we keep 105 out of the original 305 bigrams.

We then cross-reference each of our remaining bigrams with Wikipedia and form 29 groups of bigrams (“technologies”) that each refer to a specific technological advance defined in this source. For example, the bigrams “mobile devices,” “smart phones,” and “mobile platform” all refer to “smart devices,” which Wikipdia defines as “an electronic device, generally connected to other devices or networks via different wireless protocols.” Appendix Table 11 lists the definition used for each of the technologies.[10]

Another concern was that the language used by executives to characterize new technologies might not appear in patent awards. To explore the possibility that there was a business-specific vocabulary, we use an embedding vector algorithm (Mikolov et. al, 2013) trained on the set of earnings calls. This algorithm provides us with a set of bigrams used in a similar context to a given technology bigram. For each bigram in a given technology grouping, the algorithm suggests a list of “proximate” other bigrams. For example, the most proximate bigrams to those in the technology grouping “artificial intelligence” are “machine learning” and “deep learning.” From this list, we then add to the bigrams forming each technology those that, in our reading, also clearly and unambiguously describe the technology in question.

At the end of the process, we wish to ensure that the shortlisted bigrams correctly captured the use or adoption of a given technology in a given job. To this end, we performed an iterative human audit where a team member went through the randomly sampled excerpts of the text from job

[10] We explored the possibility of doing this grouping using automated approaches. For instance, we did one grouping where we clustered two bigrams into a group if the average similarity from the patent and EC embedding vectors were more than 70%. This gave a similar grouping as using our human judgement. When differences arose between the automated and human approaches, we generally preferred the results using our human judgement, so we used the latter as our preferred approach. For example, in the automated approach, “virtual reality” and “augmented reality” were clustered together with “machine learning” and “neural network,” while in our human approach we split these into two technologies: “virtual reality” for the first two and “machine learning/AI” for the second two.

postings for each bigram. He or she classified the snippet into true positive and false positive categories, along with suggestions regarding new keywords discovered and how the accuracy of the existing keywords could be improved.[11] We allowed only bigrams that, according to our reading, unambiguously reflected discussion of the technology in question at least 80% of the time. For example, we find that the bigram "automated car" rarely refers to the "Autonomous Car" technology but instead to automated car washes. Appendix Table 2 shows this human audit process in detail.

Following these additions and subtractions, we obtain a list of 221 technical bigrams associated with our 29 disruptive technologies.

Table 2 lists the 29 disruptive technologies from our supervised approach and the associated number of Burning Glass job postings in which associated bigrams appear (see the discussion in the next section). In addition to the major innovations already mentioned above, they include well-known green technologies ("Solar Power," "Hybrid Vehicle"/"Electric Car") and process innovations, such as "3D Printing," "Fracking," and "Machine Learning," but also less well-known technical and medical advances (e.g., "Millimeter Wave," a novel band of radio frequency, and "Antibody Drug Conjugates," a class of drugs used for the treatment of cancer).

Taken together, they cover a broad range of new methods and consumer applications. In total, 21.7% of all patents granted by the USPTO between 1992 and 2016 mention at least one of our technologies. In this sense, our disruptive technologies cover a significant part of recent innovative activity. While we make no claim of completeness -- other methods might well yield different groupings and definitions of technologies – we show below that each of these 29 advancements had significant implications for businesses and jobs in the United States.

[11] As an example of a false positive, an ad for a truck driver asked "do you hold a current Class A or B commercial driver's license with an air brake endorsement? … do you enjoy playing video games or computer games with a joy stick? are you good at backing up in tight spaces?" The second question led the job to be (incorrectly) classified under "electronic gaming."

Table 3 indicates the most frequent bigrams associated with the "Hybrid Vehicle"/"Electric Car," "Cloud Computing," "Autonomous Car," and "3D Printing" technologies. Appendix Table 3 provides the full set of bigrams used for each technology.

2.3. Burning Glass Job Postings

Burning Glass (BG) aggregates online job postings using "spider bots" from online job boards (such as indeed.com), employer websites (such as stanford.edu), and other sources into a machine readable, de-duplicated database. From Burning Glass, we employ two datasets. The first is a standardized dataset (used recently by Hershbein and Kahn, 2018; Demming, 2020; and Atalay *et al.*, 2020) where each de-duplicated job posting is geo-coded and assigned to a Standard Occupational Classification (SOC) code, a United States government system of classifying occupations, and a North American Industry Classification (NAICS) code based on the job posting's title.[12] The second dataset has thus far received less attention by researchers. It contains the raw unprocessed text of the job postings, which we use text to assign exposure to our technology.

We use these data from BG for all available years, 2007 and 2010-2020, a total of roughly 200 million job postings. We show below that all of our main results are robust to dropping the 2007 vintage from the sample.[13]

We associate each posting with an occupation, geography, industry, and firm as follows:

- Skill level: We construct a skill level for each six-digit SOC code from BG by measuring the share of persons with a college degree, the share of persons with a PhD, the average wage, and the average years of schooling in the American Communities Survey (ACS 2015 release), using respondents reporting their occupation as in that six-digit SOC code.[14]

---

[12] We make extensive use of the former, which are available for 80% of all postings. Industry classifications are available for a more limited 41% of postings. We use these only in our calculations in Section 5. The strings with firm names are available for 66% of all postings.

[13] BG's efforts to compile job postings data were interrupted by the 2008-09 recession. We show below that none of our main results depend on including job postings from 2007 in the sample.

[14] For SOC codes in job postings where we do not find any persons surveyed in the ACS, we match them to the closest available SOC code in the ACS. For example, data for SOC Code 38-1967 was not available, so we match it to 38-1960. In total, the dataset includes 837 SOC codes.

- Location: We use the geo-coded dataset to assign job postings to a core-based statistical area (CBSA), a U.S. government-defined geographic area that consists of one or more counties (or equivalents) with an urban center of at least 10,000 people, plus adjacent counties that are socioeconomically tied to the urban center. In total, the dataset includes 917 CBSAs.
- Industry: We allocate a job posting to an industry using the four-digit NAICS code provided by BG.
- Firm: BG reports an employer string for about 60% of their job postings. In order to match these employer strings to firms, we extend the methodology of Autor et. al. (2020) as follows: We search for the employer string (lower case and only letters a-z) on Bing.com, and collect the top five search results. We identify pairs of employer strings as the same firm if they share at least two out of top five search results. We then cluster together all employer strings that have at least two results for the same firm, and associate them with that firm.

2.4. Constructing the Exposure Measures

Using these data, we then construct measures of exposure to the set of technologies for job postings, earnings calls, and patents using the following rule:

$$exposure_{i,\tau,t} = 1\{b_\tau \epsilon\, D_{i,\tau}\}, \tag{1}$$

where $D_{i,\tau}$ is the set of bigrams contained in a job posting/earnings call/patent that was posted/held/filed at time t and b_τ is a bigram associated with a technology τ. Essentially, a document is classified as exposed to a technology if it contains a bigram associated with the technology.

Though we use the same terminology to refer to exposed job postings, earnings calls, and patents, it is worth emphasizing that these three types of exposures naturally have different interpretations. Patents that mention one of our technologies are, of course, in some way related to the development of the technology. Appendix Figure 1 provides an example of a patent concerned with object recognition, which mentions the bigram "object recognition" (a keyword associated with our "Computer Vision" technology) 52 times. Similarly, firms exposed to a given technology might be involved in developing, producing, or using a given technology, but they may also compete with or be disrupted by the technology. Appendix Table 4 gives text-based examples of these different kinds of firm-level exposures measured from earnings calls.

Most importantly, the vast majority of job postings that mention a given technology advertise jobs that either develop, produce, or use a given technology. Figure 1 and Figure 2 provide examples of two illustrative Burning Glass job postings exposed to AI and solar technology, respectively. The first is for an applied research scientist and requires "knowledge of *machine learning*, *neural networks*, and *deep learning*" – all bigrams we associate with the "Artificial Intelligence" technology. The second is for a solar panel installer, and lists as part of the job's responsibilities "install the racking system and *solar panels.*" Further down, this posting also contains another, more problematic mention of the same technology in the context of the company description, not the job itself.

To investigate the context of technology exposure in job postings more systematically, one of our team members went through 100 randomly sampled job postings for each of our 29 technologies. He or she classified them into two sets of categories, whether technology exposure in the posting referred to 1) either the overall company description or the specific task of the job in the posting, and 2) either the use or the production of the technology.

Appendix Table 5 summarizes the findings from the analysis. In Panel A, we report that in 80% of the postings, the technology mentions refer specifically to the job task (as in Figures 1 and 2). These are split about half and half into the use and the production of the technology. An example of produce would be "*You will be designing the graphics module for our* ***virtual reality*** *training system*" while an example of use would be "*The role will involve assisting customers and selling tickets from your* ***smart tablet*** *in the entrance of the cinema*". During the audit, we also noted that company descriptions are usually mentioned in the beginning or towards the end of job postings. For this reason, we disregard any technology mentions in the top and bottom 50 words of each job posting. This procedure increased the rate of capturing specific job-related tasks associated with the technology to 88% in our human audit. An additional 6% of mentions were unspecific (for example, mentions of these technologies being available in the workspace), and only 6% referred to the company but not the job.

In total, we find our 221 technical bigrams mentioned in 13 million job postings, where on average each bigram appears in 59,013 postings. To put this number into perspective, it is useful to compare this frequency with the frequency of other, "non-technical," bigrams often used by investors and executives in earnings conference calls. As documented in Appendix Tables 6 and 7, we here reverse our methodology and, instead of selecting bigrams that appear in both patents and earnings calls, select those that appear in earnings calls but not in patents. We find that the top 221 most frequent non-technical bigrams from earnings calls are on average mentioned in only 142 job postings. That is, our technical bigrams are four hundred times more frequent in job postings than other language frequently used by investors and executives, already suggesting that our 29 disruptive technologies indeed had a large impact on the U.S. labor market.

Having constructed our document-level exposure measures, we next aggregate over various documents D (job postings, earnings calls, and patents) to construct measures at the occupation, firm, and geographic levels:

$$share\ exposed_{a,\tau,t} = \frac{\sum_{i \in a,t} 1\{b_\tau \in D_{i,t}\}}{\sum_{i \in a,t} 1\{D_{i,t}\}} \tag{2}$$

where a may be a firm, sector, region, or occupation, and t is time. Appendix Table 8 illustrates a list of top occupations exposed to one of our technologies, virtual reality. Appendix Table 9 provides a shorter list of the top three most exposed occupations for each technology.

3. Region and Skill Broadening

We first seek to understand the overall patterns in the diffusion of these 29 technologies. The analysis suggests that job postings referring to given technologies grows in tandem with references in earning calls; and that over time, hiring to move from a sharp focus on high-skilled jobs to a much broader intake of workers with lower skills.

Figure 3 takes a first look at the diffusion of disruptive technologies. The 29 images plot measures of activity in job postings and in earnings calls on an annual basis for each technology. The red line denotes the percentage of firms in earnings calls that mention the given technology. In some cases, such as touchscreen and RFID, the number of mentions climb and then fade, presumably reflecting the increasing ubiquity, and hence the declining competitive relevance, of the technologies for firms. In others, such as 3-D printing and artificial intelligence, there is a steady climb over time.

In each plot, we mark the year in which the technology became economically significant, which we henceforth refer to as the "emergence year." To compute this, for each of our technologies, we calculate the maximum of the "percentage of earnings calls" time series graphed in Figure 3. We define the emergence date to be the year in which the time series first attains at least 10% of this value. Appendix Table 10 lists the emergence date for each technology, along with an alternative definition using the time series of the share of patents exposed to the technology. All of our main results are unchanged when we use this alternative definition.[15] Appendix Table 11 lists each technology, its definition as discussed above, and a suggested contemporaneous event around the year of emergence of the technology.

The second series in Figure 3, denoted with gray dots, indicates the share of positions in Burning Glass that mention a given technology (the size of the dots scale with the number of jobs posted). While in some cases a given technology continues to be important in hiring even after its mentions in earning calls drop off (e.g., GPS technology), in general, the two series are quite closely correlated. The correlation coefficient between them across the figures is 0.81. The close tie between these series helps validate the reasonableness of our empirical methodology: when a technology becomes more commercially relevant for firms, it also becomes more relevant for jobs.

[15] These robustness checks are reported in Appendix Tables 16 and 17.

Consistent with this pattern, we also find that more extensive discussions of a technology in earnings calls correlate strongly with more patenting activity in that technology. Appendix Figure 2 shows the share of firms exposed to each technology (in red-solid), and the share of citation-weighted patents (normalized by the average number citations within each technology class and year) associated with each of our 29 technologies (in black-dashes). Again, the series are highly correlated: the correlation coefficient is 0.80.

Figure 3 suggests that there is an increase in the representation of these disruptive technologies over time, which of course reflects the criteria we used to identify them. Figure 4 highlights a related feature: the increase in the usage of these technologies in job announcements over time is associated with greater geographic diffusion. To show this, we compute the coefficient of variation in the years after the emergence of a technology (defined as above) measured across locations. More specifically, we create the normalized share of job postings in technology τ and year t for each CBSA-technology-year triple by calculating:

$$Normalized\ share_{cbsa,\tau,t} = \frac{share\ exposed_{cbsa,\tau,t}}{share\ exposed_{\tau,t}}, \tag{3}$$

where the numerator is defined as in (2) and the denominator is the average share of jobs exposed to technology τ across CBSAs. $Normalized\ share_{cbsa,\tau,t}$ thus measures the regional over or underrepresentation of job postings associated with each technology relative to the overall distribution. This normalization allows us to control for the facts that, for instance, Los Angeles, the largest CBSA, will have a large share of job postings of nearly every type and that different technologies may be implemented at very different scales at a given point in time. Appendix Table 12 summarizes the data used in the analysis.

Figure 4 depicts, for each technology and year since emergence, the ratio of the standard deviation and the mean of this measure across CBSAs, also known as the coefficient of variation. The analysis reveals an intriguing pattern: 28 of 29 technologies exhibit a decline in the coefficient of variation over time (the only exception being job postings associated with the "Search Engine" technology). Put another way, although job postings in a given technology are highly regionally concentrated in the early years after their emergence, the geographic distribution of adoption over time becomes more homogeneous.

Figure 4 is corroborated by Table 4, which examines these patterns using a regression framework. The table presents the results of a regression of the coefficient of variation on the years since emergence for an annual panel of technologies, with technology and year fixed effects (column 1). Observations are weighted by the square root of the number of job postings associated with a given technology in the year, in order to give more weight to coefficients of variation that are measured more accurately.[16]

Our preferred estimate in Column 1 shows that the coefficient of variation declines by 0.105 (s.e.=0.027) per year. The mean coefficient of variation across technologies and years is 4.74. Thus, this estimate implies that the regional concentration of technology job postings declines by 22.1% of the sample average in the ten years after the emergence of the technology.

The remaining columns show the same pattern, using alternative measures of concentration. Column 2 uses the ratio of the normalized share of technology jobs of the top five CBSAs relative to all CBSAs. Column 3 uses the share of CBSAs with a normalized share of technology employment of less than 1%. Both variations show concentration at the top significantly decreasing over time.

We next examine the hiring advantage of pioneer locations that excel in initial technology-related inventions. More specifically, we define pioneer locations as those which collectively accounted for 50% of the cite-weighted patent grants associated with a given technology in the ten years before its emergence year.[17] For example, the CBSAs surrounding Trenton (NJ, 21.7%), New York (NY, 11.5%), Rochester (NY, 9.9%) and Los Angeles (CA, 9.3%) are pioneer locations for OLED

[16] This weighting scheme is for accuracy of our estimates and has no impact on the qualitative results. See Appendix Table 17 for details.

[17] An alternative approach is to define pioneer locations using the regional distribution of a given technology's job postings prior to the technology's emergence year. This approach yields a very similar allocation, as can be seen from comparing the figures in Panels A and B.

Display technology because they together accounted for 52.2% of total OLED Display patenting in the U.S. Appendix Table 13 shows the top pioneer location for each of our 29 technologies.

Panel A of Figure 5 shows the geographical distribution of pioneer locations across the United States, where the size of the blue circles is proportional to the share of the 189 technology-pioneer location pairs situated in a given CBSA. Although 23 of the 50 states host at least one pioneer location, the map shows remarkable concentration in this kind of successful innovative activity. Silicon Valley (the San Jose Jose-Sunnyvale-Santa Clara CBSA) and San Francisco were each involved in the development of 23 of our disruptive technologies, followed by New York (21), Boston (18), and Los Angeles (17). Collectively, locations in California alone host a remarkable 40.2% of our pioneer locations.[18] Another cluster along the northeast corridor from Washington to Boston accounts for another 21.2%.

Consistent with this pattern, we also find that the geographic distribution of patenting related to our 29 disruptive technologies is even more skewed than that of general patenting, which as discussed by Moretti (2019), is unevenly distributed geographically. Figure 6 depicts the population-normalized share for the top 20 CBSAs of patents linked to disruptive technologies, and the population-normalized share for all patents over the same period.

These differences can also be shown through summary statistics, The coefficient of variation of the geographic distribution of overall cite-weighted patenting is 1.21, while that of patents exposed to our 29 disruptive technologies is 1.42. Similarly, for overall patenting, it takes 12 CBSAs to account for 50% of the patenting, while the top five urban regions represent 33.8% of the patenting. Looking only at disruptive patents, it takes 7 CBSAs to account for 50%, and the top five represent 42.2%. When we look at the 189 technology-pioneer location pairs discussed above, the corresponding numbers are 5 and 54.5%.

Panels B through E continue to mark pioneer locations with hollow blue circles, but now also add the location of technology job postings in the start year of the technology (the average $Normalized\ share_{i,\tau,0}$ across technologies at $t = 0$), where darker dots correspond to a higher

[18] This fact notwithstanding, all of our main results are robust to removing California from the sample.

normalized share of jobs.[19] The figure shows a remarkable alignment between innovation and early employment. Even after accounting for differences in the size of the local labor market, early employment is strongly concentrated in the same places where the technology was developed. The remaining panels (C-E) show the evolution of this relationship as the technology matures (in years 1-2, 3-4, and 5-6, respectively). Although pioneer locations retain a higher share of technology employment throughout this period, we see a gradual diffusion of technology job postings, away from the pioneer locations and spreading out across the country.

In Table 5, we explore this relationship more formally using the specification:

$$Normalized\ share_{i,\tau,t} = \alpha_0 + \beta_1\ Pioneer_{i,\tau} + \beta_2\ Pioneer_{i,\tau}(t - t_{0,\tau}) + \chi_{i,\tau} + \varepsilon_{i,\tau,t} \qquad (4)$$

where i denotes a CBSA, τ denotes one of our 29 technologies, t denotes year, and t_0 denotes year of emergence for the technology. $Pioneer_{i,\tau}$ is a dummy which denotes the pioneer status of a CBSA-technology pair. In all specifications in Table 5, we control for technology, CBSA, and year fixed effects.

In Column 1, we see that while there is diffusion over time, the initial CBSAs where the new technology was invented retain their privileged positions. More specifically, the $Normalized\ share_{i,\tau,t}$ of a technology's job postings is about 92 percentage points higher in its pioneer locations on average throughout the lifecycle of the technology. Table 5, Column 2, however, shows that the initial advantage of pioneer locations in job postings (231 percentage points at the year of emergence) decreases at a rate of about 6% per year. The initial advantage thus has a half-life of about 8.3 years. Column 3 shows that this pattern is unchanged when we add CBSA x Year fixed effects.

We next turn to examining the skill component of technology job postings over time. Figure 7 plots a measure of skill requirements of these job postings (the red circles). We compute for each

[19] To facilitate comparison between panels, we calculate this average of normalized shares only for the 13 technologies that emerge during our Burning Glass sample and for which we have at least six years of data, that is, those emerging between 2007 and 2014.

SOC code, as reported by Burning Glass, the corresponding skill level as reported in the U.S. Census Bureau's American Community Survey for 2015. When multiple SOC codes are associated with a given technology τ in year t, we compute a weighted average of the skill measure as follows:

$$Skill_t^{\tau} = \frac{\sum_o N_{o;t}^{\tau}\, \chi_{o;2015}}{\sum_o N_{o;t}^{\tau}}$$

where o is a Census SOC code, $N_{o;t}^{\tau}$ is the number of Burning Glass job postings exposed to technology τ and SOC code o at time t, and $\chi_{o;2015}$ is the average skill level for SOC o, as measured by the 2015 ACS sample. We consider four different measures of skill at the SOC level: the share of college educated persons (baseline), the share of persons with post-graduate qualifications, the average wage of persons, and the average years of schooling for persons in the SOC.

Figure 7 plots the average share of college-educated persons associated with job postings against the year since emergence on a technology-by-technology basis. The figure suggests that for the vast majority of technologies, there is a sharp decline in the skill level required for the positions associated with new technologies over time. Even in cases where demand for positions is sharply accelerating (such as AI and virtual reality), the share of skilled positions subsides over time. These results are consistent with the view that new technologies typically start with high-skill occupations and then involve larger parts of the workforce over time. The figure also shows a few notable exceptions to this general pattern: positions exposed to the Online Streaming, Cloud Computing, Search Engine, and Software Defined Radio technologies show no evidence of a declining average skill level over time (in fact, the trend for Online Streaming appears significantly positive).

We summarize this information by presenting a binned scatterplot in Figure 8. This depiction shows the relationship across all 29 technologies between time elapsed after the emergence year and the mean share of the postings for college-educated people. It shows, on average, a strong negative linear trend, implying a declining requirement for a college-trained workforce as technologies mature.

Table 6 looks at this relationship formally. The sample consists of annual observations of each technology between 2007 and 2019. Here, we use the alternative measures of the skills required in the job postings associated with a given technology: the dependent variables include the share of the weighted SOC classes that are college educated (as of 2015), the share with graduate degrees, mean wages, and the mean years of schooling. Each regression uses as the key independent variable the years since the emergence date and controls for technology and calendar year fixed effects. The specification again follows Table 4 regarding the criteria for inclusion in the analysis and weighting.

Using each measure, there is a strong negative relationship between the maturity of the technology and the reliance on a highly educated workforce. For instance, Columns 1 and 3 show that each additional year since the emergence of the technology is associated with a fall of about 0.96 percentage points in the share of job postings requiring a college education (an annual decline of -1.71%) and a decline of $1,023 in annual wages (measured in 2015 constant dollars) for the job postings associated with the technology. Similarly, the share of job postings in occupations requiring a post-graduate degree declines by a rate of 1.80% per year on average.

This skill-broadening effect sheds an interesting light on how high-skilled labor is complementary with low-skilled work. While there is an important body of work highlighting the way in which technological change has favored high-skilled occupations and contributed to wage inequality (Acemoglu, 2002; Goldin and Katz, 2009; Acemoglu and Autor, 2011 are examples), the way in which the hiring associated with new technologies can transition over time highlights the dynamics in this relationship.

We next explore the heterogeneity of our region-broadening and pioneer persistence facts across skill categories. We use the SOC codes to divide our sample of job postings into three categories using the share of college-educated people in each SOC code. Again, we use information from the 2015 ACS to determine the qualifications of individuals in various SOCs. We termed these high (job postings for occupations with at least 60% college educated), medium (with 30% to 59% college educated) and low skilled (less than 30% college educated). For instance, almost all optometrists in the ACS are college educated: thus, all job postings for optometrists are allocated to the high-skill category. We then examine how the decline in the coefficient of variation described above changes after the emergence year, and how these shifts differ across different skill levels.

Figure 9 takes a first look at these patterns. It again is a binned scatterplot of the coefficient of variation by year, but with the two extremes (low and high skill) of this three-fold division. It shows that the decline in the coefficient of variation across regions is substantially steeper for low-skilled jobs than that for high-skilled ones. While the low-skilled jobs rapidly disperse across the country, the higher-end ones remain more bunched together.

Table 7 studies these patterns in more detail, emulating the structure of the specification in Table 4, but now breaking the observations of technologies into high and low-skill buckets (omitting the medium-skill bucket) and adding an interaction between the years since emergence variable and a dummy for low-skill occupations. All specifications show a significantly larger decline in concentration for lower-skill occupations. In terms of magnitudes, the annual decline in the coefficient of variation for low-skill job postings is more than three times larger than that for high-skill jobs, so that it declines by 3.7% annually for low-skill jobs and only 1.1% for high-skill jobs. Appendix Table 14 shows this specification separately for job postings in the three skill buckets. Again, high-skill professions show a less steep decline in geographic concentration, although the coefficient of variation declines significantly for all three groupings over time.

We obtain similar results for the persistence of pioneer advantage result in Table 8. This table repeats the analysis of Table 5, column 2 separately for each bucket (low, medium, and high skill) of job postings. Rather than looking at dispersion, however, it focuses on the related concept of the persistence of the pioneer region. Consistent with the earlier results, we find the decline in initial pioneer advantage is greater in the case of low-skilled than high-skilled positions. The degradation in geographic concentration is about 6.7% for low-skill job postings, which is about twice the magnitude for high-skill job postings (3.5%). That is, pioneer locations where disruptive technologies were developed retain a long-term advantage in attracting job postings in that technology, particularly in high-skill occupations. The estimates in column 3 suggest the half-life of this high-skill advantage is 14.3 years.

4. Properties of Technology Hubs

Up until this point, we have ignored the characteristics of the original pioneer locations where the technologies were developed (and also did the bulk of their hiring at the time of the emergence date). In this section, we explore their features. In particular, we highlight that there is a strong relationship between academic centers and the pioneer locations where nascent disruptive technologies originate.

To this end, we calculate for each CBSA-technology pair the number of patents exposed to that technology ten years prior to the technology's emergence year. (Recall our definition of pioneer locations is based on this variable: a dummy that is one for locations that account for 50% of a technology's patents in that year). We normalize this number by CBSA population in the emergence year. We then regress this variable (patents in technology τ per 1000 inhabitants) on region characteristics in 2015 (using data from the ACS).

The key independent variables, which measure the presence of research universities and skilled persons in a CBSA, are the logarithm of the volume of university assets (standardized by population), the university enrollment (standardized by population), the share of the population in the CBSA that is college educated or has a post-graduate degree, and the log average wage in the CBSA.[20] Finally, this all specifications control for technology-specific fixed effects.

Panel A of Table 9 shows a strong cross-sectional pattern. Regions with a greater academic or skill presence—whether manifested by greater research university presence or a more educated

[20] We obtain university data for 642 research universities from Higher Education Research and Development Survey (HERD), and map these universities to CBSAs. Research universities are defined as "public and private nonprofit postsecondary institutions in the United States, Guam, Puerto Rico, and the U.S. Virgin Islands that granted a bachelor's degree or higher in any field; expended at least $150,000 in separately budgeted R&D in FY 2015; and were geographically separate campuses headed by a president, chancellor, or equivalent." We normalize university assets and the university enrollment by CBSA population from the ACS at the time of the year of emergence. We obtain skill level variables for a particular CBSA from the ACS, by normalizing the share of graduate and post graduate persons in a CBSA by the total number of persons in the CBSA.

workforce—were more likely to be involved in the early development of disruptive technologies. These patterns are illustrated graphically in Appendix Figure 3.

Perhaps more importantly, and consistent with our results above, Panel B shows that these same variables also account for higher per capita technology job postings in the emergence year. That is, the same variables that account for the location of innovative activity also account for early employment in that technology.

5. Diffusion across Other Dimensions

In this section, we characterize the spread of disruptive technologies across industries, occupations, and firms. First, we compare the region-broadening result against broadening across industries, occupations, and firms; second, similar to Table 5, we also study initial advantage of pioneers, separately defined across the four segments, and the degradation in this advantage over time.

To that end, we extend the definition of $Normalized\ share_{i,\tau,t}$ in Section 3 to NAICS four-digit industries, SOC six-digit occupations, and firms for each technology (τ) and time (t). While the former two variables are included in the BG data (in each case, we use the finest level of disaggregation available), the latter relies on our own matching algorithm described in Section 2.

We then measure the coefficient of variation of $Normalized\ share_{i,\tau,t}$ across the segments. Because the number of firms posting job advertisements online expands over time (with more and more small firms appearing in the BG data over time), we stratify our firm-technology-year sample by including only firms that post at least one job in each of our sample-years, before calculating the coefficient of variation.[21] This step focuses attention on 10,231 larger firms which on average post 1,628 jobs per year, effectively excluding variation coming from small and medium-sized businesses.

[21] Hershbein and Kahn (2018) discuss this fact in some detail. The general increase in coverage of the BG data over time should not affect any of our main results. We discuss robustness to various weighting schemes in detail in Section 7.

Table 10, Panel A shows the results of a regression of the coefficient of variation calculated for each technology (τ) and time (t) on year since emergence. Column 1 shows our already established results for pioneer locations for comparison.[22] We find that while there is a decline in concentration as measured by coefficient of variation for all four segments, there appears to be a larger decline across locations and firms (Columns 1 and 4) than across industries and occupations (Columns 2 and 3). While the coefficient of variation declines on average by 2.48% and 2.32% for CBSAs and firms, respectively, the corresponding declines are 1.06% and 0.81% for (four-digit NAICS) industries and (six-digit SOC) occupations, respectively. Figure 10 illustrates these patterns graphically.

While it is perhaps natural to expect disruptive technologies to spread faster across firms and space than they do across industries and occupations, any quantitative comparison of course depends on the classifications of industries and occupations used. Appendix Figures 4 through 6 shows some differences across technologies in diffusion across industries, occupations, and firms. For example, the 3D Printing, Computer Vision, and Wi-Fi technologies show a clear decrease in concentration across industries over time.

In Table 10, Panel B, we estimate specification (1) for all four dimensions to examine the initial hiring advantage of pioneer cells in the four segments. The pioneer cells, as defined before, are ones that excel in initial technology-related inventions. More specifically, we define pioneer cells as those which collectively accounted for 50% of the patent grants associated with a given technology in the ten years before its emergence.

To determine the pioneer cells, we merge various public-use datasets to assign patents to our three additional segments of industries, occupations and firms: For industries, we allocate patents to individual NAICS four-digit industries by mapping patents to Compustat firms (since patents themselves do not contain industry codes), and then from firms to industries. A total of 44% of all patents exposed to any one of our 29 technologies are owned by Compustat firms, so that this

[22] In order to avoid calculating coefficients of variation for unreasonably sparse data, we only keep technology x year observations with at least 100 postings with industry coverage. This issue arises because BG provides NAICS codes for only 41% of all postings.

procedure implicitly assumes that the distribution of patents across industries is similar for Compustat firms as for all firms[23]. To obtain the patent-to-Compustat match we use the crosswalk provided by Autor *et al*. (2020). Once patents are matched to firms, we then link to industries using the Compustat Segments dataset, which gives firms' breakdown of sales across NAICS four-digit industries. So, for example, if a patent is owned by "Amazon North America," it is matched by Bing to "Amazon Inc.," and then allocated proportionally to Amazon's NAICS four-digit industries by its sales breakdowns (exclusively to "Electronic shopping and mail-order houses" in Amazon's case).

For occupations, we further construct an industry-to-occupation crosswalk from employment data within an occupation-industry cell from the Occupational Employment Statistics. We assume that the share of patenting in an industry allocated to an occupation is same as the share of employment allocated to an occupation. We can, thus, calculate the share of patents for a particular technology allocated to an occupation.

Finally, for firms, we string match patent assignees from USPTO to firm names in job postings. Using this procedure, we are able to match 36% of all patents assigned to U.S. inventors between 1976 and 2016 to 30,123 unique firms in our sample.

Following our procedure for pioneer locations, we define pioneer industries, occupations, and firms for each technology as those with the most assigned patents in the ten years prior to the technology's emergence year that collectively account for 50% of the matched patents in a given disruptive technology. Appendix Table 15 shows the top pioneer industry and occupation for each technology. For example, the top pioneer industry for 3D Printing is "Computer and Peripheral Equipment Manufacturing" (accounting for 41.9% of early patents) and that of Fracking is "Oil and Gas Extraction" (accounting for 88.1% of early patents).

[23] In order to compare patents by Compustat and non-Compustat firms, we analyze the share of patents by Compustat-firms across technology classes. We find that for the median technology class, about 50% of patents are produced by Compustat firms, and that the distribution is quite homogenous: the 25th percentile is 39.0% and 75th percentile is 58.8%.

The analysis in Table 10, Panel B shows that pioneering cells have a strong initial advantage in job postings for all four segments. However, the advantage appears more persistent for industries, occupations, and firms than for locations. In terms of magnitudes, the degradation in the advantage ($\beta(Pioneer_{i,\tau} * (t - t_0))/\ \beta(Pioneer_{i,\tau})$) for locations is 6.2%, compared to 4.4% for firms, 4.0% for industries, and 3.4% for occupations.

Taken together, this evidence suggests disruptive technologies initially generate hiring that is highly localized by location, firm, industry and occupation. Over time, this hiring disperses, particularly across locations and across firms.

6. Firm Rehoming towards Pioneer Locations

As a final analysis, we explore one of the mechanisms behind the region-broadening results: the rehoming of firms towards pioneer locations. More specifically, we investigate whether large and established firms' shift their geographical footprint towards pioneer locations as a technology progresses in its life cycle. For now, we perform the analysis with a case study; we plan to provide systematic evidence in future drafts of the paper.

To explore this, we consider geographical footprint of Ford Motor Company and General Motor Corporation before and after the emergence year of the autonomous cars technology (2014). In Figure 11, we plot these firms' job postings in three groups of places: (a) the three autonomous car pioneer locations, San Jose (CA), San Francisco (CA), and Boston (MA) (but excluding Detroit (MI)); (b) their headquarters, Detroit (MI), and (c) all other locations. Postings in red are before the emergence year of autonomous car technology, and postings in blue are post-emergence year. Black crosses in the picture denote the share of job postings exposed to autonomous vehicles post emergence year.

The figure shows that both firms, traditionally concentrated in Detroit, shifted their geographic footprint towards the autonomous cars pioneer locations, particularly in Silicon Valley (San Jose and San Francisco). A large share of new job postings in the pioneer locations involved autonomous car technologies, accounting for 22% and 65% of Ford and GM postings respectively (compared to less than 5% in all other locations). The data thus suggest that the purpose of both firms' expanding presence in autonomous cars' pioneer locations related to this new technology.

7. Additional Robustness Checks

Before concluding, we perform a number of additional robustness checks for our primary results: "region broadening," "pioneer-location persistence," "skill broadening," and "differential region-broadening by skill level."

First, we replicate our results using our "unsupervised" approach to defining technologies. That is, we treat each of the original 305 technology bigrams we obtained from our algorithm intersecting the texts of patents and earnings calls as a separate technology, without attempting to group or otherwise audit these bigrams. The goal of this exercise is to replicate our main findings in a dataset created without any human intervention.

In Table 11, columns 1 through 4 of Panel A replicate the main specifications of Tables 4 through 8, respectively. We find that all the coefficients of interest are qualitatively and quantitatively similar to our main specification.[24]

Panel B of Table 11 replicates the results of Table 10, estimating the spread of disruptive technologies across industries, occupations, and firms. The results are again similar, although this unsupervised approach yields somewhat faster spread across occupations than in our baseline specification.

[24] Column 1 shows our region-broadening result, regressing each technology-year's coefficient of variation across locations on the number of years since the emergence of the technology. The estimated coefficient (-0.140, s.e.=0.017) implies a 2.54% reduction in concentration in technology job postings per year, comparted to 2.21% in our baseline specification (Table 4, column 1). Similarly, the estimates in column 2 imply a large advantage of pioneer locations in job postings that decreases at a rate of 6.0% per year, compared to 6.6% in Table 5 column 3. Column 3 also shows significant skill broadening over time, with a decreasing share of job postings that require a college education over the life-cycle of the technology. However, the estimate here (-0.325, s.e.=0.099) is only one third the size of that in Table 6, column 1. Finally, column 4 shows that the geographic concentration of low-skill jobs exposed to disruptive technologies decays significantly faster than that of high-skill jobs, though the coefficient of interest is again somewhat smaller (-0.108, s.e.=0.028 vs -0.167, s.e.=0.048 in our baseline specification).

We conclude that the human judgement that we exerted to enable us to measure the spread of specific technologies has no bearing on the validity of our main stylized facts about disruptive technologies as a whole.

Second, in Appendix Table 16, we check for robustness with respect to our methodology for calculating the year of emergence for our technologies, with respect to the missing years (2008 and 2009) in the BG sample, and with respect to standard errors:

- In Panel A, we find that our results are robust to calculating years of emergence exclusively from patents instead of earnings calls. To calculate this alternative measure, we use our patent data, which extends back to 1975. The year of emergence for each technology is here calculated as the year in which the share of U.S. patents exposed to the technology reach 50% of their maximum value between 1976 and 2015.
- In Panel B, we find that our results are robust to excluding 2007, the first year of availability of Burning Glass job postings and immediately before the missing BG job postings in 2008 and 2009.
- In Panel C, we check for robustness of standard errors and find that if anything the statistical significance is stronger with robust standard errors (vs. clustered standard errors in the baseline specification).

Third, we deal with a potential concern with all the analyses: that they may reflect changes in the composition of the job announcements in Burning Glass, not hiring overall. Appendix Figure 7a shows that the number of job postings began increasing sharply in the mid-2010s (the blue line), which could reflect an increase in the share of jobs posted online. We note, however, that this trend also parallels the increase in overall U.S. job openings after the 2008/09 recession, as reported by the Job Openings and Labor Turnover Survey (the red line).

A more substantive compositional concern is raised by Appendix Figure 7b. The figure shows that much of the growth in Burning Glass online job postings was driven by job postings in low-skill occupations. It is natural to speculate that many of these jobs may have previously not been posted online. Thus, the increase in BG postings shown in Appendix Figure 7a likely reflects both increasing overall U.S. hiring and a growing tendency for lower-skill job announcements to be posted online. It is thus natural to wonder whether the changing composition of BG job announcements may have impacted the results above.

After three additional analyses, we do not believe these changes affect the results in our analyses. First, it is important to note, as demonstrated in Appendix Figure 7c, that the compositional patterns documented in Appendix Figure 7b are much less pronounced among job announcements associated with our 29 technologies. Second, our entire analysis uses the normalized share of job postings (except skill broadening), and controls throughout for year fixed effects. The normalization and year controls should address many of these compositional concerns. As a final check for our skill broadening result, we reweight the occupations in our sample to match hiring in that occupation in the U.S. economy. We compute hiring in each occupation by using hiring in each industry in the LEHD and then constructing a crosswalk between industry employment and occupation employment using the OES. In Appendix Table 17, we find that our skill-broadening results are robust to this reweighting exercise.

As a final check of our broadening results, we check their sensitivity to technology selection: in other words, could the results be driven by a handful; of industries out of our 29? To do this, we exclude three technologies at a time and recalculate the degradation in coefficient of variation, this provides us with 7,308 permutation estimates. In Appendix Figure 8, we plot the 10^{th} and 90^{th} percentile of these jackknife estimates, and show that the results are robust to randomly removing a subset of technologies.

8. Conclusion

Policymakers in many parts of the world devote enormous energy to foster nascent technologies, ranging from efforts to support academic research to luring start-ups from other cities and nations. Such an infant industry strategy is predicated on the notion that early advantages in innovation and employment will yield lasting benefits for regions, particularly in the form of high-quality employment.

Using the full text of millions of patents, job postings, and earnings conference calls over the past two decades, we introduce in this paper an approach to understand which new technologies affect businesses and to trace their diffusion across regions, industries, occupations, and firms. We can then map the spread of disruptive technologies in these dimensions, focusing on the hiring associated with each important innovation.

We highlight five main conclusions. First, the locations where disruptive technologies are developed are geographically highly concentrated, with a handful of urban areas contributing the bulk of the early patenting and early activity within each technology. Second, despite this initial concentration, jobs relating to use or production of the new technologies gradually spread out geographically. Third, while initial jobs associated with a given technology are typically high-skilled, over time the mean required skill levels of the new jobs declines. Fourth, these trends towards region and skill broadening are related: low-skill jobs associated with a given technology spread out geographically significantly faster than high-skill ones. Finally, because of the slower spread of high-skill jobs, disruptive technologies continue to offer long-lasting benefits for their pioneer locations, which retain a long-term advantage in these high-quality jobs for multiple decades.

Beyond these core results of our analysis, the development and spread of disruptive technologies are key objects of interest in multiple fields of economics. We therefore hope that the data we provide as part of this paper may prove useful to address a range of additional research questions in the study of economic growth, inequality, entrepreneurship, and firm dynamics.

One additional avenue for future research relates to the microeconomic dynamics of pioneer locations. To what extent is their persistent advantage in high-skill job openings driven by re-homing of established firms as opposed to the initial developers of the technology? How much of this effect is the consequence of knowledge spillovers or feedback to universities?

Another avenue would investigate the determinants and consequences of success: Why do some regions appear to develop a disproportionate share of disruptive innovations, and how does such serial success affect the local markets for labor and housing?

A related question is around the spread of technologies across firms and locations. To what extent is this firms expanding or rehoming, and what types of firms were particularly prescient in identifying the new technologies? Is it those who saw it as an especial competitive threat? Answers to these questions will help us better explain these fascinating phenomena.

References

Acemoglu, Daron. "Directed technical change." *Review of Economic Studies* 69 (2002): 781-809.

Acemoglu, Daron, and David Autor. "Skills, tasks and technologies: Implications for employment and earnings," in Orley Ashenfelter and David Card (editors), *Handbook of Labor Economics*. New York, Elsevier, volume 4, chapter 12, pages 1043-1171 (2011).

Acemoglu, Daron, and Joshua Linn. "Market size in innovation: Theory and evidence from the pharmaceutical industry." *Quarterly Journal of Economics* 119 (2004): 1049-1090.

Acemoglu, Daron, and Pascual Restrepo. “Robots and jobs: Evidence from US labor markets.” *Journal of Political Economy* 128 (2020) 2188-2244.

Aghion, Philippe, Ufuk Akcigit, Antonin Bergeaud, Richard Blundell, and David Hemous. “Innovation and Top Income Inequality.” *Review of Economic Studies* 86 (2019): 1–45.

Agrawal, Ajay. Joshua Gans, and Avi Goldfarb, editors. *The Economics of Artificial Intelligence: An Agenda.* Chicago: University of Chicago Press (2019).

Akerman, Anders, Ingvil Gaarder, and Magne Mogstad. "The skill complementarity of broadband internet." *Quarterly Journal of Economics* 130 (2015): 1781-1824.

Arora, Ashish, Sharon Belenzon and Lia Sheer. “Knowledge spillovers and corporate investment in scientific research.” *American Economic Review* 111 (2021) 871-898.

Atalay, Enghin, Phai Phongthiengtham, Sebastian Sotelo, and Daniel Tannenbaum. "The evolution of work in the United States." *American Economic Journal: Applied Economics* 12 (2020): 1-34.

Autor, David H., David Dorn, Gordon H. Hanson, Gary Pisano, and Pian Shu. "Foreign competition and domestic innovation: Evidence from US patents." *American Economic Review: Insights* 2 (2020): 357-74.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. "Trends in US wage inequality: Revising the revisionists." *Review of Economics and Statistics* 90 (2008): 300-323.

Autor, David H., Lawrence F. Katz, and Alan Krueger. “Computing inequality: Have computers changed the labor market?” *Quarterly Journal of Economics* 113 (1998): 1169-1213.

Autor, David H., Frank Levy, and Richard J. Murnane. “The skill content of recent technological change: An empirical exploration.” *Quarterly Journal of Economics* 118 (2003): 1279-1334.

Bekkerman, Ron and James Allan. “Using bigrams in text categorization.” Technical Report IR-408, Center of Intelligent Information Retrieval, University of Massachusetts-Amherst (2003).

Berman, Eli, John Bound, and Zvi Griliches. “Changes in the Demand for Skilled Labor within U.S. manufacturing: Evidence from the Annual Survey of Manufacturers.” *Quarterly Journal of Economics* 109 (1994): 367-397.

Bushee, Brian J., Dawn A. Matsumoto, and Gregory S. Miller. "Open versus closed conference calls: The determinants and effects of broadening access to disclosure." *Journal of Accounting and Economics* 34 (2003): 149-180.

Caprettini, Bruno, and Hans-Joachim Voth. “Rage against the machines: Labor-saving technology and unrest in industrializing England.” *American Economic Review: Insights* 2 (2020) 305-320.

Comin, Diego, and Bart Hobijn. “Cross-country technology adoption: Making the theories face the facts.” *Journal of Monetary Economics* 51 (2004): 39-83.

Comin, Diego, and Bart Hobijn. “An exploration of technology diffusion.” *American Economic Review* 100 (2010): 2031–59.

Davies, Mark. "The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights." *International Journal of Corpus Linguistics* 14 (2009): 159-190.

Deming, David J., and Kadeem Noray. "Earnings dynamics, changing job skills, and STEM careers." *Quarterly Journal of Economics* 135 (2020): 1965-2005.

Glaeser, Edward L., Sari Pekkala Kerr, and William R. Kerr. "Entrepreneurship and urban growth: An empirical assessment with historical mines." *Review of Economics and Statistics* 97 (2015): 498-520.

Goldin, Claudia D., and Lawrence F. Katz. “The origins of technology-skill complementarity,” *Quarterly Journal of Economics* 113 (1998): 683-732

Goldin, Claudia D., and Lawrence F. Katz. *The Race Between Education and Technology*. Cambridge, Harvard University Press (2009).

Gompers, Paul, Josh Lerner, and David Scharfstein. "Entrepreneurial spawning: Public corporations and the genesis of new ventures, 1986 to 1999." *Journal of Finance* 60 (2005): 577-614.

Gordon, Robert J. *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War.* Princton, Princeton University Press (2016).

Greenstone, Michael, Richard Hornbeck and Enrico Moretti, “Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings.” *Journal of Political Economy*, 118 (2010): 536-598.

Griliches, Zvi. "Hybrid corn: An exploration in the economics of technological change." *Econometrica* 25 (1957): 501-522.

Hall, Bronwyn H. "Innovation and diffusion." Working paper no. 10212, National Bureau of Economic Research (2004).

Hassan, Tarek A., Stephan Hollander, Laurence van Lent, and Ahmed Tahoun. "Firm-level political risk: Measurement and effects." *Quarterly Journal of Economics* 134 (2019): 2135-2202.

Hassan, Tarek Alexander, Stephan Hollander, Laurence van Lent, and Ahmed Tahoun. "Firm-level exposure to epidemic diseases: Covid-19, SARS, and H1N1.” Working paper no. 26971, National Bureau of Economic Research (2020).

Hershbein, Brad, and Lisa B. Kahn. "Do recessions accelerate routine-biased technological change? Evidence from vacancy postings." *American Economic Review* 108 (2018): 1737-72.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. “Geographic localization of knowledge spillovers as evidenced by patent citations.” *Quarterly Journal of Economics* 108 (1993): 577–98

Katz, Lawrence F. and Kevin M. Murphy. "Changes in relative wages, 1963-1987: Supply and demand factors." *Quarterly Journal of Economics* 107 (1992): 35-78.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. "Measuring technological innovation over the long run.” *American Economic Review: Insights* (forthcoming).

Krueger, Alan B. “How computers have changed the wage structure: Evidence from microdata, 1984-1989.” *Quarterly Journal of Economics* 108 (1993): 33-60.

Lanjouw, Jean O., Ariel Pakes, Jonathan Putnam. "How to count patents and value intellectual property: The uses of patent renewal and application data." *Journal of Industrial Economics* 46 (1998): 405-432.

Lerner, Josh, and Amit Seru. "The use and misuse of patent data: Issues for corporate finance and beyond." *Review of Financial Studies* (forthcoming).

Machin, Stephen, and John van Reenen. "Technology and changes in skill structure: Evidence from seven OECD countries." *Quarterly Journal of Economics*, 113 (1998) 1215-44.

Matsumoto, Dawn, Maarten Pronk, and Erik Roelofsen. "What makes conference calls useful? The information content of managers' presentations and analysts' discussion sessions." *Accounting Review* 86 (2011): 1383-1414.

Michaels, Guy, Natraj, Ashwini and Van Reenen, John. "Has ICT polarized skill demand? Evidence from eleven countries over 25 years." *Review of Economics and Statistics* 96 (2014): 60-77

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781* (2013).

Moretti, Enrico. "The effect of high-tech clusters on the productivity of top inventors." Working paper no. 26270, National Bureau of Economic Research (2019).

Moscona, Jacob. "Environmental catastrophe and the direction of invention: Evidence from the American Dust Bowl." Unpublished working paper, Massachusetts Institute of Technology (2020).

Mokyr, Joel. *The Lever of Riches: Technological Creativity and Economic Progress*. New York: Oxford University Press (1992).

Moser, Petra, Alessandra Voena, and Fabian Waldinger. "German Jewish Émigrés and US Invention." *American Economic Review* 104 (2014): 3222–55.

Piketty, Thomas, and Emmanuel Saez. "Income inequality in the United States, 1913–1998." *Quarterly Journal of Economics* 118 (2003): 1-41.

Popp, David. "Induced innovation and energy prices." *American Economic Review* 92 (2002): 160-180.

Song, Jae, David J. Price, Faith Guvenen, Nicholas Bloom, and Till Von Wachter. "Firming up inequality." *Quarterly Journal of Economics* 134 (2019): 1-50.

Squicciarini, Mara P., and Nico Voigtländer. "Human capital and industrialization: Evidence from the age of enlightenment." *Quarterly Journal of Economics*. 130 (2015): 1825-1883.

Tan, Chade-Meng, Yuan-Fang Wang, and Chan-Do Lee. "The use of bigrams to enhance text categorization." *Information Processing & Management* 38 (2002): 529–546.

United States Patent and Trademark Office. "Performance and Accountability Report." (2020).

Webb, Michael. "The Impact of Artificial Intelligence on the Labor Market." Unpublished working paper, Stanford University (2020).

Tables and Figures

Table 1 - Top bigrams from patenting and ECs

Bigram	# transcripts	Technology group
mobile devices	6597	Smart devices
machine learning	2860	Machine learning/AI
cloud computing	2781	Cloud computing
cloud services	2450	Cloud computing
quality metrics	2029	NA
flow profile	1966	NA
smart phones	1957	Smart devices
mobile platform	1605	Smart devices
public cloud	1569	Cloud computing
social networking	1548	Social networks
smart grid	1441	NA
cloud service	1393	Cloud computing
connected devices	1304	Smart devices
cloud infrastructure	1136	Cloud computing
carbon footprint	1071	NA
nand flash	1002	NA
virtual reality	903	Virtual reality
digital channel	896	NA
delivery network	887	NA
social networks	883	Social networks
autonomous driving	839	Autonomous Cars
smart devices	765	Smart devices
active user	735	Social networks
augmented reality	730	Virtual reality
mobile payment	717	Mobile payment
cloud environment	668	Cloud computing
production site	664	NA
ethanol production	662	NA
power outage	643	NA
multiple segments	595	NA

Notes: This table list top 30 out of the total initial 305 bigrams (in Column 1) that appear frequently in patents and earnings calls, and increase in their mentions between 2002 and 2019. The bigrams are sorted by the number of earnings calls that they are mentioned in (Column 2). The table also reports the technology group that they are classified in (Column 3). Note that some of the bigrams are not classified in any technology group because they do not refer to a recent disruptive technology.

Table 2 - Technologies by total job postings

Technology	Postings
Cloud computing	3684901
Social Networking	3457390
Smart devices	2376510
Machine Learning/AI	679776
Search Engine	535784
Online streaming	487731
Wi-Fi	388844
Electronic gaming	247201
Solar Power	201296
Injection molding	190538
Hybrid vehicle/Electric car	118550
Touch Screen	109538
Rfid	80894
Computer vision	76350
GPS	65922
Mobile payment	65482
Virtual Reality	61102
3d printing	57904
Autonomous Cars	52974
Lane departure warning	32107
Lithium battery	16926
Software defined radio	14187
Drug conjugates	10603
Fracking	8966
Millimeter wave	6161
Oled display	5528
Bispecific monoclonal antibody	2702
Inkjet printing	2583
Wireless charging	1649
Stent graft	1270
Fingerprint sensor	711

Notes: This table lists our 29 technologies (in Column 1) and the number of job postings that they appear in (Column 2) Burning Glass during 2007-2019.

Table 3 - Top keywords for sample technologies by number of online job postings

Hybrid vehicle/Electric car		Cloud computing	
keyword	postings	Keyword	postings
electric vehicles	87948	Saas	961142
electric vehicle	11647	cloud based	663357
vehicle charging	11402	enterprise applications	558611
hybrid electric	10284	cloud computing	485333
electric car	8219	cloud services	276906
hybrid vehicle	7926	cloud platform	241376
electrical vehicles	3875	paas	220732
electric buses	782	cloud infrastructure	216271
electric motorcycle	125	cloud environments	190832
plugin hybrids	40	iaas	187695

Autonomous cars		3dprinting	
keyword	postings	word	postings
autonomous vehicles	22099	3d printing	33398
self-driving car	17533	additive manufacturing	17008
autonomous driving	12992	3d printer	14962
automated driving	6564	3d printed	2481
autonomous cars	1489		
driverless car	1215		
robot car	1060		
driverless truck	129		
selfdriving car	19		

Notes: The table lists top bigrams (in Column 1 in each section) by the number of online job postings that they are mentioned in (Column 2), 2007-19, for a sample set of technologies.

Table 4 - Region broadening and persistence

	Coefficient of Variation	$\frac{Mean(NS)_{Top\ 5}}{Mean(NS)_{All}}$	Share CBSAs with ($NS < 1\%$)
	(1)	(2)	(3)
$Years\ since\ emergence_{\tau,t}$	-0.105***	-0.028***	-1.078***
	(0.027)	(0.006)	(0.338)
R2	0.861	0.927	0.776
N	287	287	287
Tech FE	YES	YES	YES
Year FE	YES	YES	YES
Mean	4.74	0.67	53.33
% Mean per year	2.21%	4.17%	2.02%

Notes: This table reports results from a regression of three separate measures of geographic concentration of technology hiring, calculated over $Normalized\ share_{i,\tau,t}(NS) = \frac{shar\quad jobs\ expsed_{i,\tau,t}}{shar\quad jobs\ expsed_{\tau,t}}$, where i is a location (CBSA), technology τ, and time t. The three measures are: coefficient of variation, normalized share of hiring at top 5 CBSAs relative to bottom 5 CBSAs, and share of CBSAs with normalized share more than 1%. The regression is weighted by square root of total technology postings in a year. Normalized share is capped at 99th percentile of non-zero observations. Standard errors are clustered by technology. Standard errors are clustered by technology. The last row specifies the magnitude of the coefficient of $Years\ since\ emergence_{\tau,t}$ as a percentage of the sample mean per year.

Table 5 - Persistence of pioneer locations

	Normalized Share		
	(1)	(2)	(3)
$Pioneer_{i,\tau}$	0.918***	2.313***	2.474***
	(0.285)	(0.580)	(0.699)
$Pioneer_{i,\tau} * Years\ since\ emergence_{\tau,t}$		-0.146***	-0.163***
		(0.042)	(0.057)
R2	0.074	0.075	0.104
N	266,467	266,467	266,467
$\beta(Pioneer_{i,\tau} * (t - t_0)) / \beta(Pioneer_{i,\tau})$		-0.063***	-0.066***
		(0.007)	(0.009)
Tech FE	YES	YES	YES
Year FE	YES	YES	YES
CBSA FE	YES	YES	YES
CBSA x Year FE	NO	NO	YES

Notes: This table reports results from a regression of the $Normalized\ share_{i,\tau,t}$ (for each CBSA, technology, and year) on pioneer status of the CBSA and its interaction with year since technology emergence. Normalized share is capped at 99th percentile of non-zero observations. Standard errors are clustered by technology. Standard errors are clustered by CBSA.

Table 6 - Skill measure of technology job postings and years since emergence

	(1)	(2)	(3)	(4)
	Share of college educated * 100	Share of post graduate * 100	Average wage	Average schooling
Years since emergence	-0.954***	-0.361***	-1,022.929***	-0.050***
	(0.260)	(0.121)	(241.521)	(0.014)
R2	0.847	0.878	0.845	0.859
N	287	287	287	287
Tech FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Mean	55.90	19.95	64,463	15.07
%Mean/year	-1.71%	-1.80%	-1.59%	-.33%

Notes: This table reports the results from a regression of approximate skill composition of technology jobs as the dependent variable, $Skill_t^{\varsigma} = \frac{\sum_o N_{o;t}^{\tau} \chi_{o;2015}}{\sum_o N_{o;t}^{\tau}}$ where $\chi_{o;2015}$ is the skill measure of interest from ACS 2015 at the occupation level, on the years since inception of the technology as the independent variable. Occupation in the sample is at the six-digit SOC code. These results exclude observations before the start year of a technology. The regression is weighted by square root of technology job postings in a year. Standard errors are clustered by technology.

Table 7 – Concentration across locations during the life cycle - High vs low skill

	Coefficient of Variation	$\frac{Mean(NS)_{Top\ 5}}{Mean(NS)_{All}}$	Share CBSAs with $(NS < 1\%)$
	(1)	(2)	(3)
(Years since t0) * 1{Low skill}	-0.167***	-2.218***	-0.022***
	(0.048)	(0.621)	(0.006)
(Year since t0)	-0.074**	-0.657	-0.017***
	(0.036)	(0.464)	(0.005)
R2	0.773	0.653	0.827
N	567	567	567
Skill FE	YES	YES	YES
Tech FE	YES	YES	YES
Year FE	YES	YES	YES
Mean	6.53	16.85	0.78
% Mean/year	2.56%	13.16%	2.82%

Notes: This table reports the results from regressions of coefficient of variation during lifecycle of a technology by skill (high skill and low skill occupations) on year since technology emergence. The interaction term (year since t0) * 1{low skill} tests for differential concentration trends between low and high skill technology jobs. To calculate the coefficient of variation by skill x technology x year, we aggregate the job postings data over occupation, CBSA and year, and then separately for high skill occupations (with share of college educated people > 60 %), and low skill occupations (with share of college educated people < 30 %). Finally, coefficient of variation is calculated over $Normalized\ share_{i,\tau,t,skill} = \frac{share\ jobs\ exposed_{i,\tau,t,skill}}{share\ jobs\ exposed_{\tau,t,skill}}$ across CBSAs by skill group, technology and time. These results exclude observations before the start year of a technology, and they are weighted by technology hiring in the skill-technology-year observation. Regression controls for skill, technology, and year fixed effects. Standard errors are clustered by technology.

Table 8 - Differential hiring for locations by skill

	$Normalized\ share_{i,\tau,t}$		
	(1) Low Skill	(2) Medium Skill	(3) High Skill
$Pioneer_{i,\tau}$	1.607***	1.193***	1.108**
	(0.403)	(0.453)	(0.484)
$Pioneer_{i,\tau} * (t - t_0)$	-0.108***	-0.057**	-0.039*
	(0.030)	(0.025)	(0.020)
R2	0.053	0.044	0.049
N	181,598	181,598	181,598
$\beta(Pioneer_{i,\tau} * (t - t_0)) / \beta(Pioneer_{i,\tau})$	-0.067***	-0.048**	-0.035**
	(0.007)	(0.016)	(0.014)

Notes: This table reports the results from regressions of $Normalized\ share_{i,\tau,t,skill}$ on pioneer status dummy for the CBSA i, separately for low skill (Column 1), medium skill (Column 2) and high skill (Column 3). To construct the sample at skill x CBSA x year level. We aggregate the job postings data over occupation, CBSA and year, and then separately for high skill occupations (with share of college educated people > 60%), medium skill occupations (with share of college educated people > 30%), and low skill occupations (with share of college educated people < 30%). These results exclude observations before the start year of a technology. Standard errors are clustered by CBSA. Standard errors for $\beta(\mathrm{Pioneer}_{i,\tau} * (t - t_0)) / \beta(\mathrm{Pioneer}_{i,\tau})$ are calculated using delta method.

Table 9 - Technology patenting before technology emergence versus skill composition

Panel A:	(1)	(2)	(3)	(4)	(5)
	$Patents\ per\ 1000\ people_{i,T,0}$				
log(1 + university assets (in $1,000 per capita))	0.129***				
	(0.022)				
University enrollment per capita		0.346***			
		(0.085)			
Share College Educated (in pct.)			0.0178***		
			(0.0017)		
Share post graduate (in pct.)				0.0421***	
				(0.0041)	
Log(wage)					1.004***
					(0.117)
Observations	24,731	24,731	24,731	24,731	24,731
R-squared	0.107	0.093	0.158	0.162	0.133

Panel B:	(1)	(2)	(3)	(4)	(5)
	$Postings\ per\ 1000\ people_{i,\tau,0}$				
log(1 + university assets (in $1,000 per capita))	0.0595***				
	(0.0075)				
university enrollment per capita		0.217***			
		(0.0313)			
Share College Educated			0.00657***		
			(0.00063)		
Share post graduate				0.0149***	
				(0.0015)	
Log(wage)					0.426***
					(0.045)
Observations	24,759	24,759	24,759	24,759	24,759
R-squared	0.179	0.172	0.197	0.196	0.192
Tech FE	YES	YES	YES	YES	YES

Notes: The table presents results from a regression of patents per capita (in Panel A) and postings per capita (in panel B) in a CBSA corresponding to a technology (during 10 years before year of emergence for the technology) on repeated values of various measures of skill and income for the CBSA. University measures in row 1 and row 2 are calculated by aggregating university assets and enrollment over all universities in a CBSA, and share of college educated/post graduate in row 3 and row 4 are calculated as the share of people holding a college/postgraduate degree in a CBSA. Income measure in row 5 is log of wage, where wage for a CBSA is calculated as the average yearly income of a working person. The university data is from US Dept. of Education, and the education and income data is from American Communities Survey 2015. All specifications control for technology fixed effects. Standard errors are clustered by CBSA.

Table 10 – Dispersion and pioneer persistence: Comparison across different dimensions

Panel A:	Coefficient of Variation			
	(1) Locations	(2) Industries	(3) Occupations	(4) Firms
Years since emergence	-0.092***	-0.052	-0.054	-0.360***
	(0.026)	(0.037)	(0.049)	(0.093)
R2	0.888	0.904	0.806	0.917
N	249	249	249	249
Mean	3.71	4.89	6.65	15.48
% Mean/year	-2.48%	-1.06%	-0.81%	-2.32%
Tech FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES

Panel B:	$Normalized\ share_{i,\tau,t}$			
	(1) Locations	(2) Industries	(3) Occupations	(4) Firms
$Pioneer_{i,\tau}$	2.393***	13.550***	10.746**	142.036***
	(0.528)	(3.204)	(4.675)	(35.866)
$Pioneer_{i,\tau} * Years\ since\ emergence$	-0.149***	-0.547**	-0.367	-6.215**
	(0.039)	(0.224)	(0.269)	(2.990)
R2	0.076	0.137	0.033	0.026
N	266,467	26,883	204,041	38,990,627
$\beta(Pioneer_{i,\tau} * (t - t_0))/\ \beta(Pioneer_{i,\tau})$	-0.062***	-0.040***	-0.034***	-0.044***
	(0.007)	(0.011)	(0.013)	(0.013)
Tech FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Cell FE	YES	YES	YES	YES

Notes: This table reports: 1) in Panel A, the results from a regression of Coefficient of Variation calculated across $Normalized\ share_{i,\tau,t} = \frac{share\ jobs\ expsed_{i,\tau,t}}{share\ jobs\ expsed_{\tau,t}}$ (where i is a location (Column 1), industry (Column 2), occupation (Column 3) or firm (Column 4)) for each technology τ and time t. Location refers to a CBSA, industry is at NAICS 4-digit level, and occupation is at SOC 6-digit level. Coefficient of variation in Column (4) is calculated across 10,231 firms which have at least 1 job posting in each of the 11 years of Burning Glass. The results only include observations at the time of and after the start year of a technology, and observations with more than 100 technology jobs which have an industry associated with them. The regression is weighted by square root of total technology postings in a year. Normalized share is capped at 99th percentile of non-zero observations. Standard errors are clustered by technology. 2) In Panel B, results from regression of $Normalized\ share_{i,\tau,t} = \frac{share\ jobs\ exposed_{i,\tau,t}}{share\ jobs\ exposed_{\tau,t}}$ on pioneer status of each cell for a panel of technologies (τ) across time t. Pioneer status is given to cells which account for more than 50% of patents in the 10 years before the year of emergence for the technology. These results exclude observations before the start year of a technology. Normalized share is capped at 99th percentile of non-zero observations. Standard errors are clustered by cell. $\beta(Pioneer_{i,\tau} * (t - t_0))/\ \beta(Pioneer_{i,\tau})$ denotes the decrease in advantage for a pioneer cell every year, and we calculate its standard error using the delta method.

Table 11 - Robustness - Dispersion with individual bigrams as technologies

Panel A

Dependent Variable:	CV	Normalized Share	Share College Educated	CV
Result:	Region Broadening (1)	Pioneer Persistence (2)	Skill Broadening (3)	Region Broadening by Skill (4)
Years since emergence	-0.140***		-0.325***	-0.244***
	(0.017)		(0.099)	(0.024)
Pioneer location		1.454***		
		(0.342)		
Pioneer location * (Years since emergence)		-0.087***		
		(0.027)		
(Years since emergence) * 1{Low skill}				-0.108***
				(0.028)
R2	0.833	0.022	0.868	0.724
N	2,185	2,797,245	2,185	5,807
Estimate (per year)	-2.54%	-5.98%	-0.60%	-0.98%

Panel B

	Coefficient of Variation			
	(1) Locations	(2) Industries	(3) Occupations	(4) Firms
Years since emergence	-0.140***	-0.018	-0.110***	-0.377***
	(0.017)	(0.018)	(0.019)	(0.066)
R2	0.833	0.890	0.725	0.908
N	2,185	2,185	2,185	2,185
Mean(CV)	5.51	5.94	6.79	25.82
% Mean(CV)/year	-2.54%	-0.30%	-1.62%	-1.46%

Notes: This table reports our primary results replicated by treating each bigram as a separate technology. In panel A, we replicate our primary results in table 4 column 1, table 5 column 2, table 6 column 1 and table 7 column 3. In panel B, we replicate results from table 10 panel A, again treating each of the 305 bigrams as a technology. The regression is weighted by square root of total technology postings in a year. Normalized share is capped at 99th percentile of non-zero observations. Standard errors are clustered by technology for col 1, col 3 and col4 in panel A, and for all columns in Panel B. Standard errors are clustered by CBSA for col 2 in Panel A.

Figure 1 – Sample job for "AI Technology"

1.0.1 JobTitle

Applied Research Scientist, Video Understanding

1.0.2 JobText

4.4 Facebook New York, NY Glassdoor Estimated Salary: 112k159k
Applied Research Scientist, Video Understanding
Facebook

Facebooks mission is to give people the power to build community and bring the world closer together. Through our family of apps and services, were building a different kind of company that connects billions of people around the world, gives them ways to share what matters most to them, and helps bring people closer together. Whether were creating new products or helping a small business expand its reach, people at Facebook are builders at heart. Our global teams are constantly iterating, solving problems, and working together to empower people around the world to build community and connect in meaningful ways. Together, we can help people build stronger communities were just getting started.Every day, massive amounts of video are uploaded into Facebooks services. In order to serve our communities better, it is critical that we can understand this content think about being able to answer questions like This person will like this video because.... or This person will find this video inappropriate because... Our goals broadly encompass content understanding, including the ability to produce video summaries, categorize content according to topic and purpose, identify audio events, find keyframes, and do keyword spotting. To achieve these goals, we are building a Video Understanding team in New York City, that will engage in a multidisciplinary effort combining speech recognition, natural language processing, and image processing. We view video as inherently multimodal content, and seek to develop methods that use all the information available. We are looking for researchers in machine learning and AI with strong software engineering skills, and a desire to build systems that will ship to billions of people. The Video Understanding Team is part of the Applied Machine Learning organization. The team carries out applied research in MLAI and designs, develops and deploys state of the art MLAI algorithms to the rest of Facebook. Our algorithms are used for ranking, improving content integrity, keeping communities safe, and power multiple product experiences across Facebook, Messenger, Instagram, WhatsApp and Oculus.
Responsibilities:
Develop highly scalable algorithms based on stateoftheart machine learning and neural network methodologies Conduct research to advance the stateoftheart, and publish work in relevant speech, NLP, and machine learning conferences and journals Apply expert coding skills to projects in partnership with other engineers across research, product, and infrastructure teams Adapt machine learning and neural network algorithms for training competitive, stateoftheart models which make the best use of modern parallel environments e.g. distributed clusters, GPU Mininum Qualifications:
MS degree in Computer Science or related quantitative field with 5 years of work experience, or Ph.D. degree in Computer Science or related quantitative field Knowledge of **machine learning, neural networks, and deep learning** Experience building systems based on machine learning andor **deep learning** methods, especially in the areas of speech recognition, natural language processing, image processing, or other machineperception tasks Experience developing and debugging in CC andor Python Preferred Qualifications:
Overview
Website www.facebook.com
Headquarters Menlo Park, CA, United States
Size 10000 employees
Founded 2004
Type Company Public FB
Industry Information Technology

Notes: The picture is a sample job posting, which mentions AI technology related keywords, with a standardized job title, processed by Burning Glass, and the text of the job advertisement posted online on glassdoor.com.

Figure 2– Sample job for "Solar Technology"

1.0.1 Job Title

Solar Panel Installer

1.0.2 JobDomain

www.glassdoor.co.in

1.0.3 JobDate

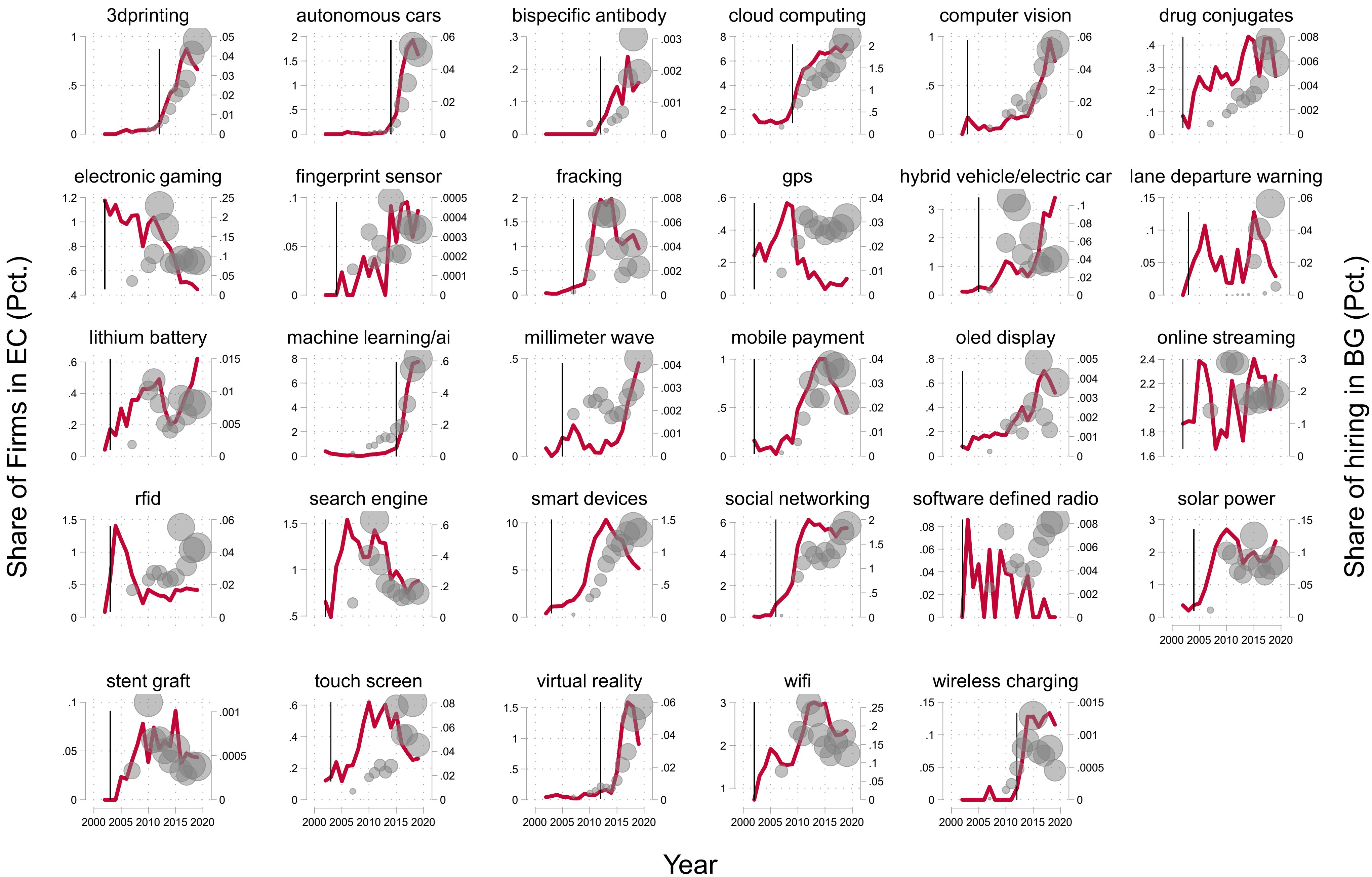
20190315

1.0.4 JobText

3.4 Vivint Solar Corp Baltimore, MD
Solar Panel Installer
Vivint Solar Corp
Job Description: Right now, we are seeking a Solar Installer for our Rosedale MD office, who will be responsible for ensuring that our products are installed properly and ontime.
Must be have a valid drivers license Must be able to pass preemployment drug screen Must be able to pass criminal background
Responsibilities: Work with a team to install the racking system and solar panels on residential roofs Service the solar system as needed
Required: Working knowledge of solar installation, construction andor roofing 1 to 2 years of relevant experience Ability to be comfortable being and working on roofs Valid drivers license Employees of Vivint Solar must submit to a criminal history check, motor vehicles check, drug screening, and obtain clearance from the state based upon the state requirements.
We do not accept resumes from headhunters, placement agencies, or other suppliers that have not signed a formal agreement with us.
Vivint Solar is a proud promoter of employment opportunities to our Military and Veterans. We, an equal opportunity employer, do not consider any protected traits e.g. race, creed, color, religion, gender, national origin, nonjobrelated disability, age, or any other protected trait when hiringunder federal, state and local laws Company Description Vivint Solar is a leading fullservice residential solar provider in the United States. With Vivint Solar, customers can power their homes with clean, renewable energy and typically achieve significant financial savings. Offering integrated residential solar solutions for the entire customer lifecycle, Vivint Solar designs, installs, monitors and services the solar energy systems for its customers. In addition to being able to purchase a solar energy system outright, customers may benefit from Vivint Solars affordable, flexible financing options or power purchase agreements. For more information, visit www.vivintsolar.com or follow VivintSolar on Twitter.
Overview
Website www.vivintsolar.com
Headquarters Lehi, UT, United States
Size 1001 to 5000 employees
Founded Unknown
Type Company Public VSLR
Industry Oil, Gas, Energy Utilities
Revenue 5 to 10 billion INR per year
Competitors Unknown
Vivint Solar Photos
Vivint Solar photo of: CEO, David Bywater, recognizing overachieving employees and thanking them for

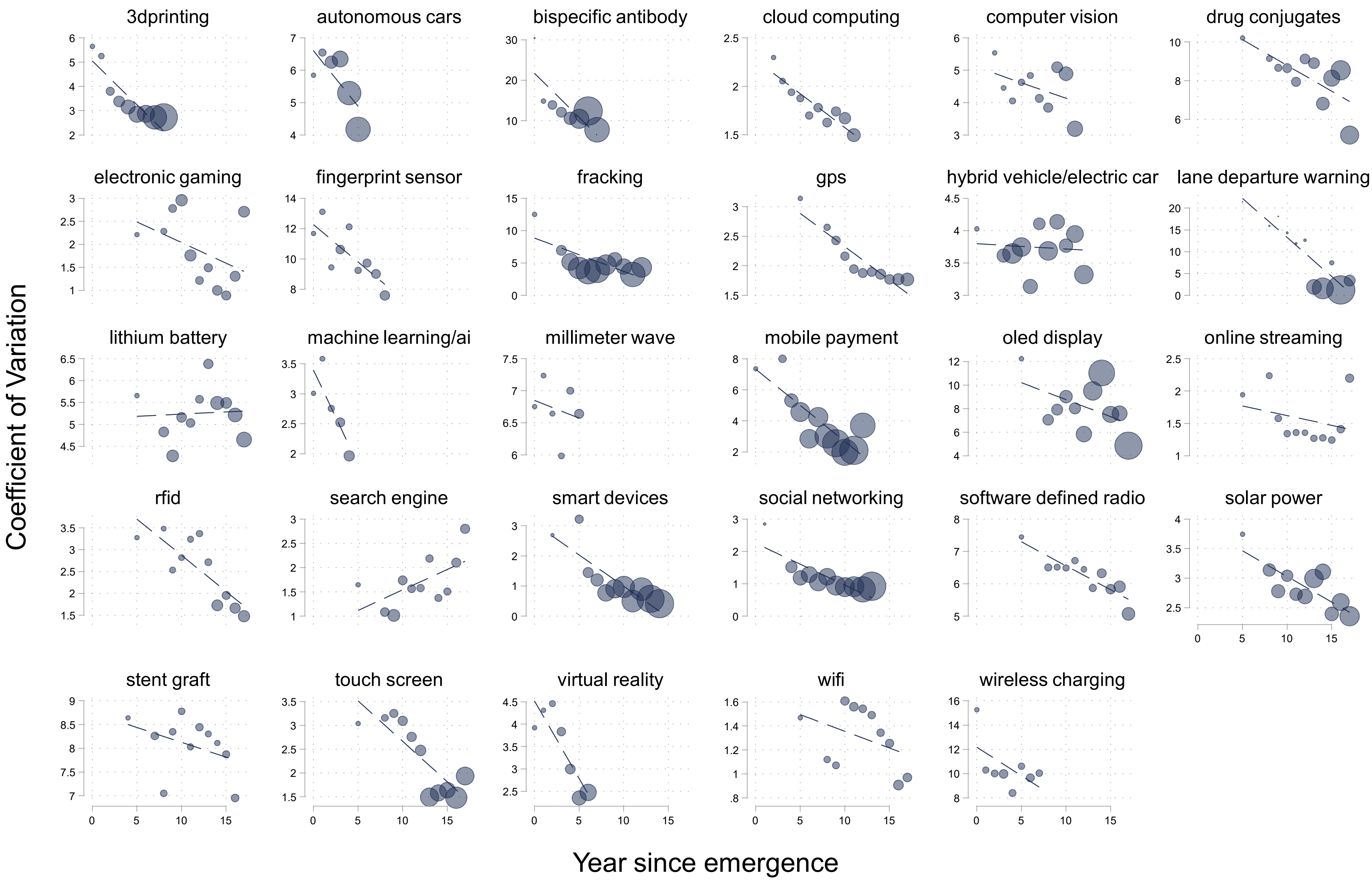
Notes: The picture is a sample job posting, which mentions solar technology related keywords, with a standardized job title, processed by Burning Glass, and the text of the job advertisement posted online on glassdoor.com.

Figure 3 – Technology exposure in earnings calls and jobs postings – Time series



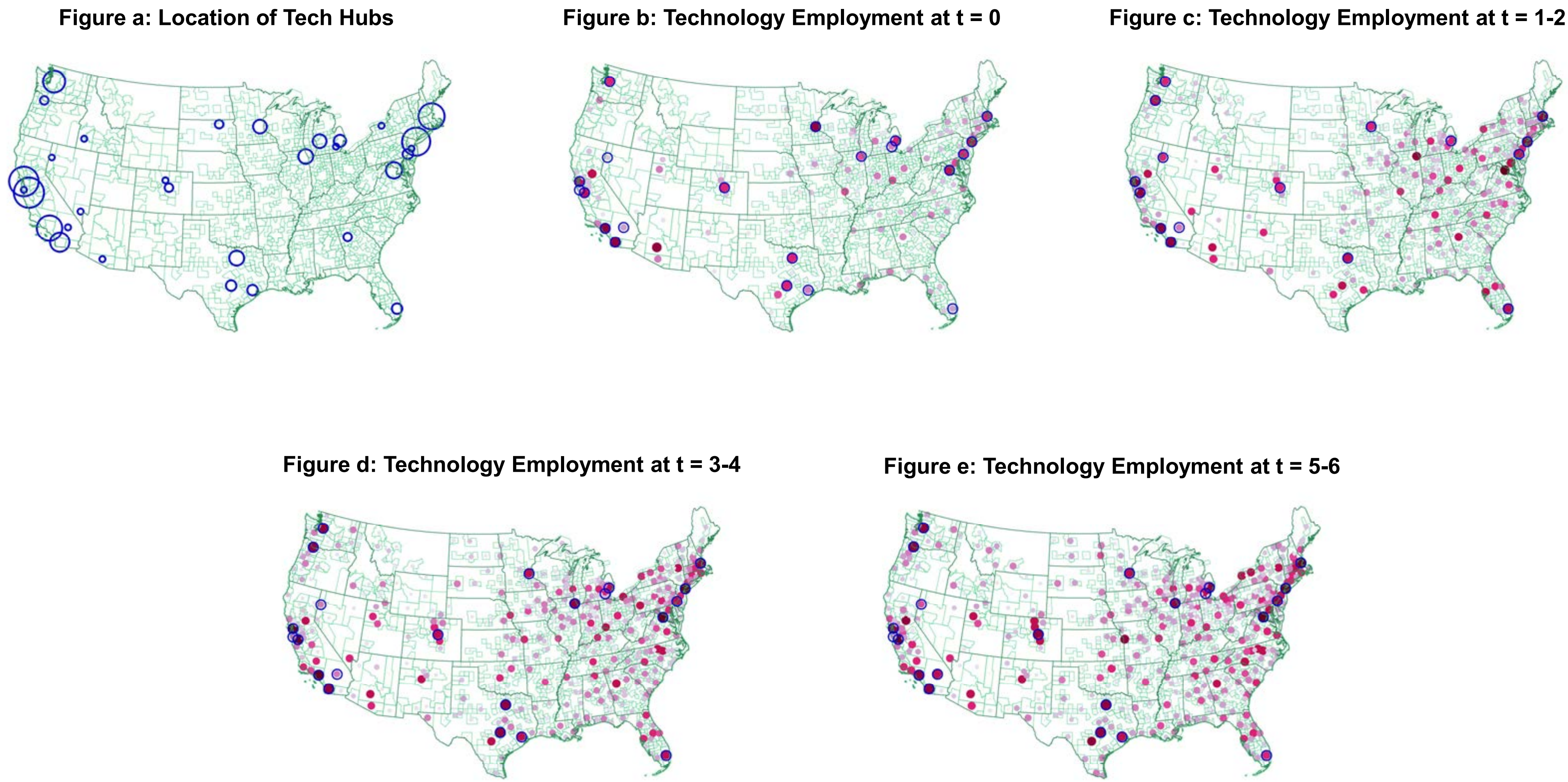
Notes: The pictures plot (year by year) the percentage of firms (red line) which mention technology-related keywords in earnings calls, and the percentage of job postings (grey circles) in Burning Glass which mention technology related keywords. Size of the grey circles denotes the level of hiring for the technology x year observation. The vertical grey line highlights the year of emergence of the technology, which is defined as the year in which firms time series (red line) attains at least 10% of sample max. The overall correlation between these two time series is 81\%.

Figure 4 – Coefficient of Variation across locations by year since emergence



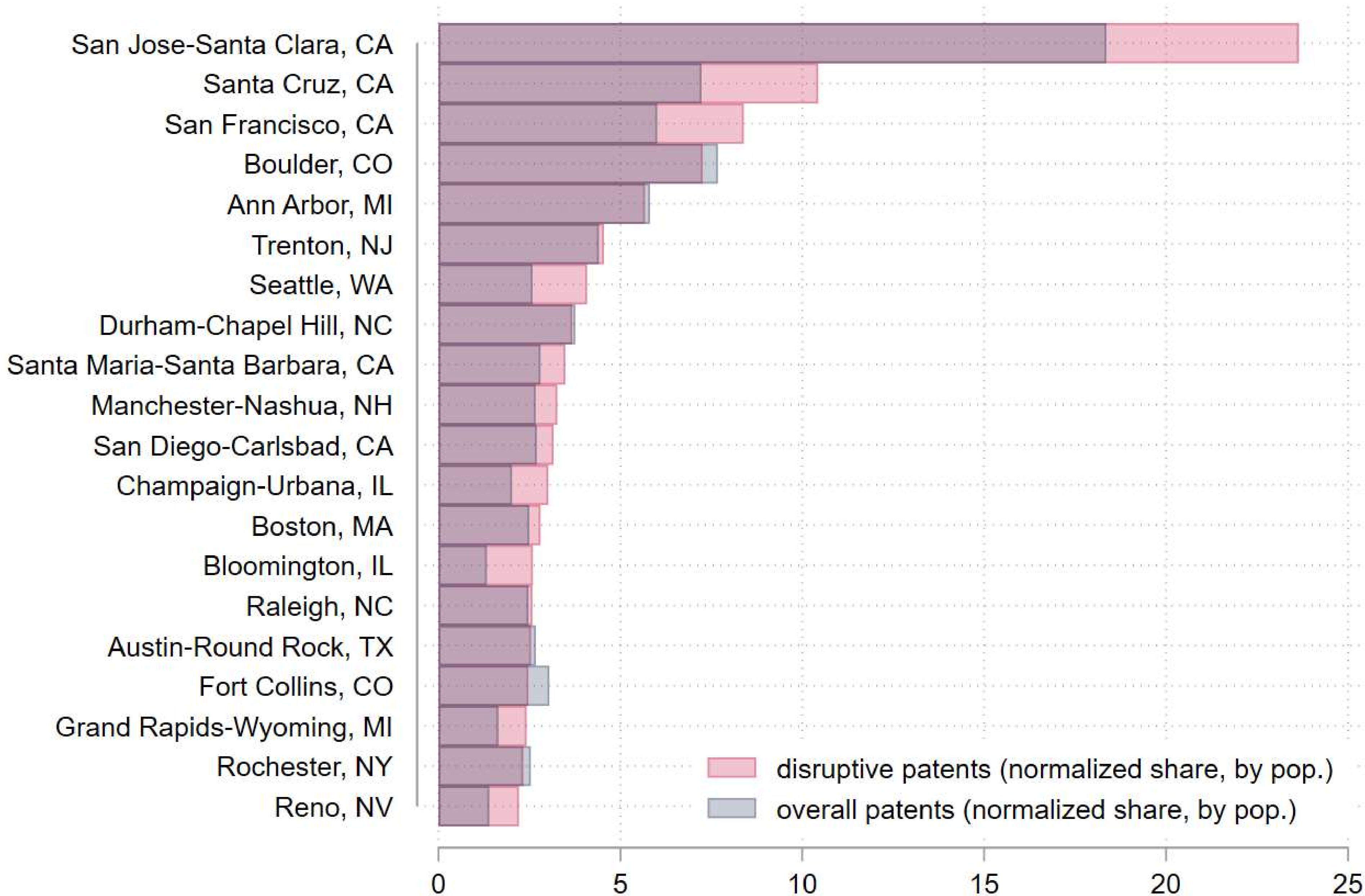
Notes: The figure plots coefficient of variation measured as coefficient of variation of normalized share of technology jobs for each of 29 technologies by year from 2007-2019 against the years since emergence of the technology, where $Normalized\ share_{i,\tau,t} = \frac{share\ jobs\ expsed_{i,\tau,t}}{share\ jobs\ expsed_{\tau,t}}$, where i is a CBSA. Only observations post year of emergence are included.

Figure 5 – Technology diffusion from hubs



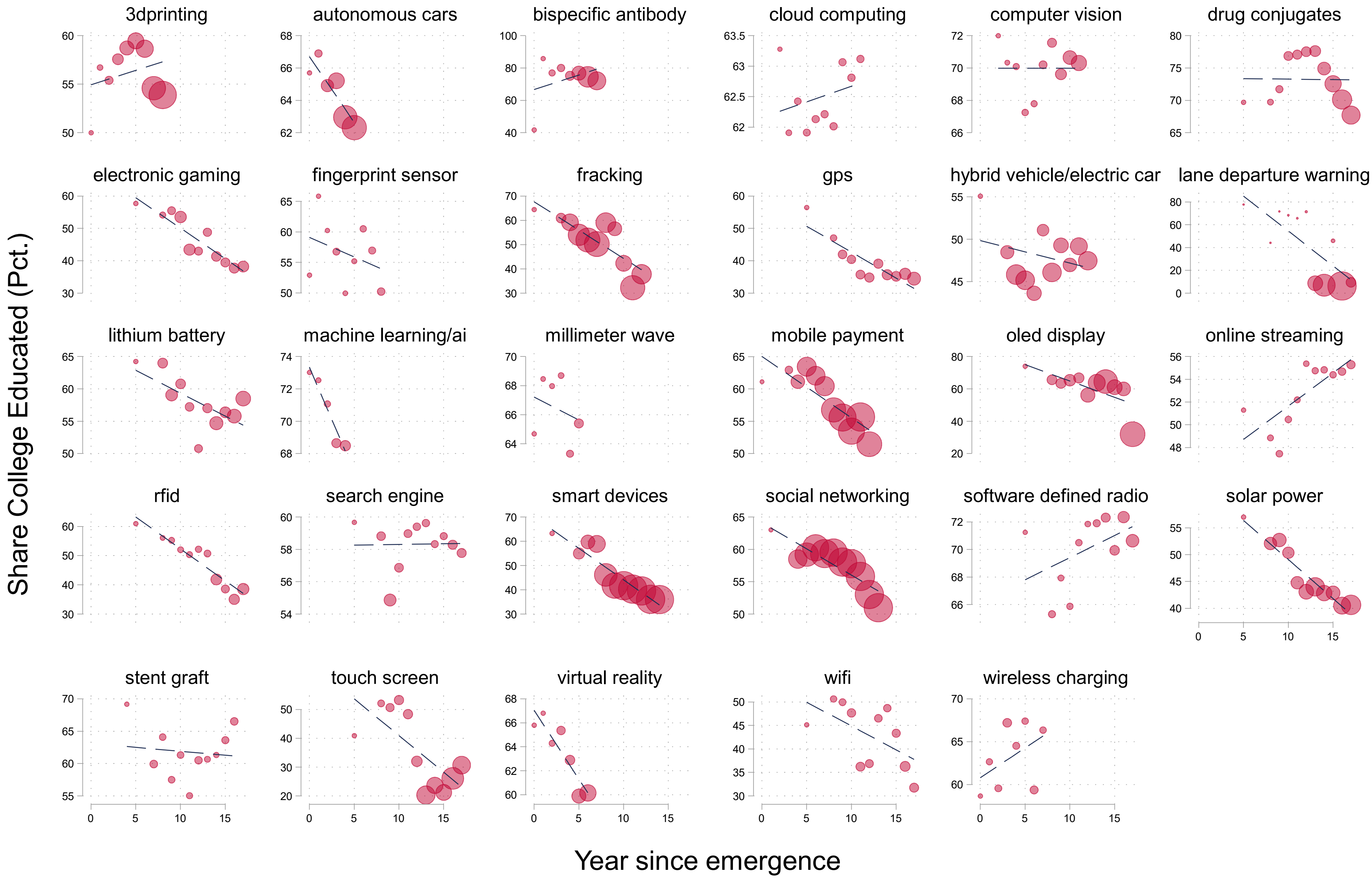
Notes: This figures shows: 1) In Figure (a) the Core based Statistical Areas (CBSAs) which are technology hubs for at least one technology, where the size of the circles is proportional to the share of technologies for which the CBSA is a hub; 2) In Figure (b) we show the share of technologies for which the $Normalized\ share_{i,\tau,t}$ of technology hiring at the CBSA in the year of emergence is greater than 1% (t=0). In Figure (c), (d) and (e), we repeat the mapping in (b) for years since emergence 1-2, 3-4, and 5-6, respectively. We plot these pictures only for 13 out of our 29 technologies which have year of emergence after 2007 for a complete panel of each technology.

Figure 6 – Disruptive vs. overall patents, by top CBSAs.



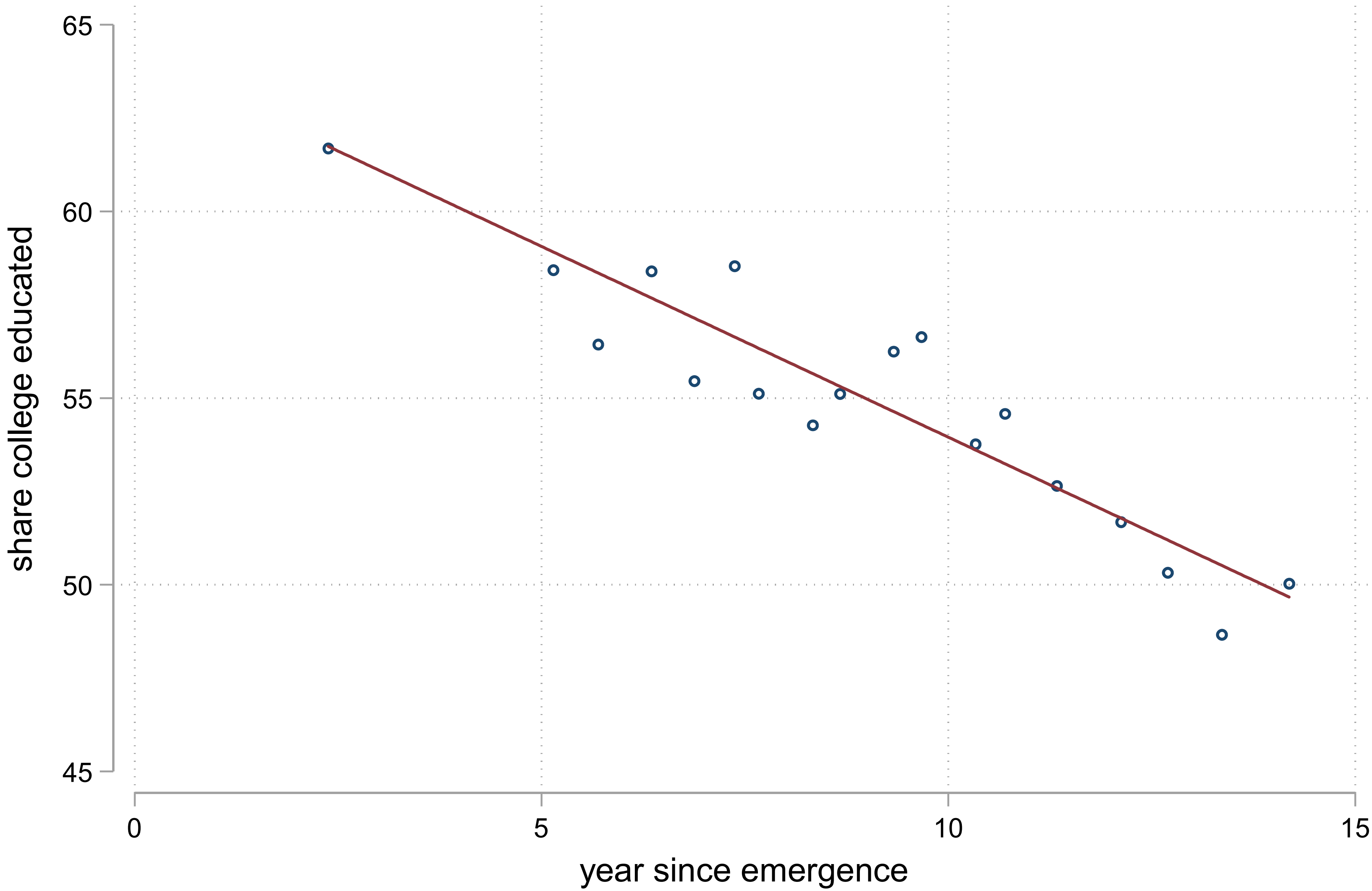
Notes: The figure shows, for disruptive patents (in red) and overall patents (in blue), the normalized share of patenting for top 20 CBSAs. Normalized share of patents for a CBSA is defined as the share of total patents filed by US inventors in the CBSA (between 1992 and 2016) divided by the share of U.S. population in the CBSA (as of 2015). The figure is sorted by largest to smallest normalized share of disruptive patenting.

Figure 7 - Share of college educated by year since emergence



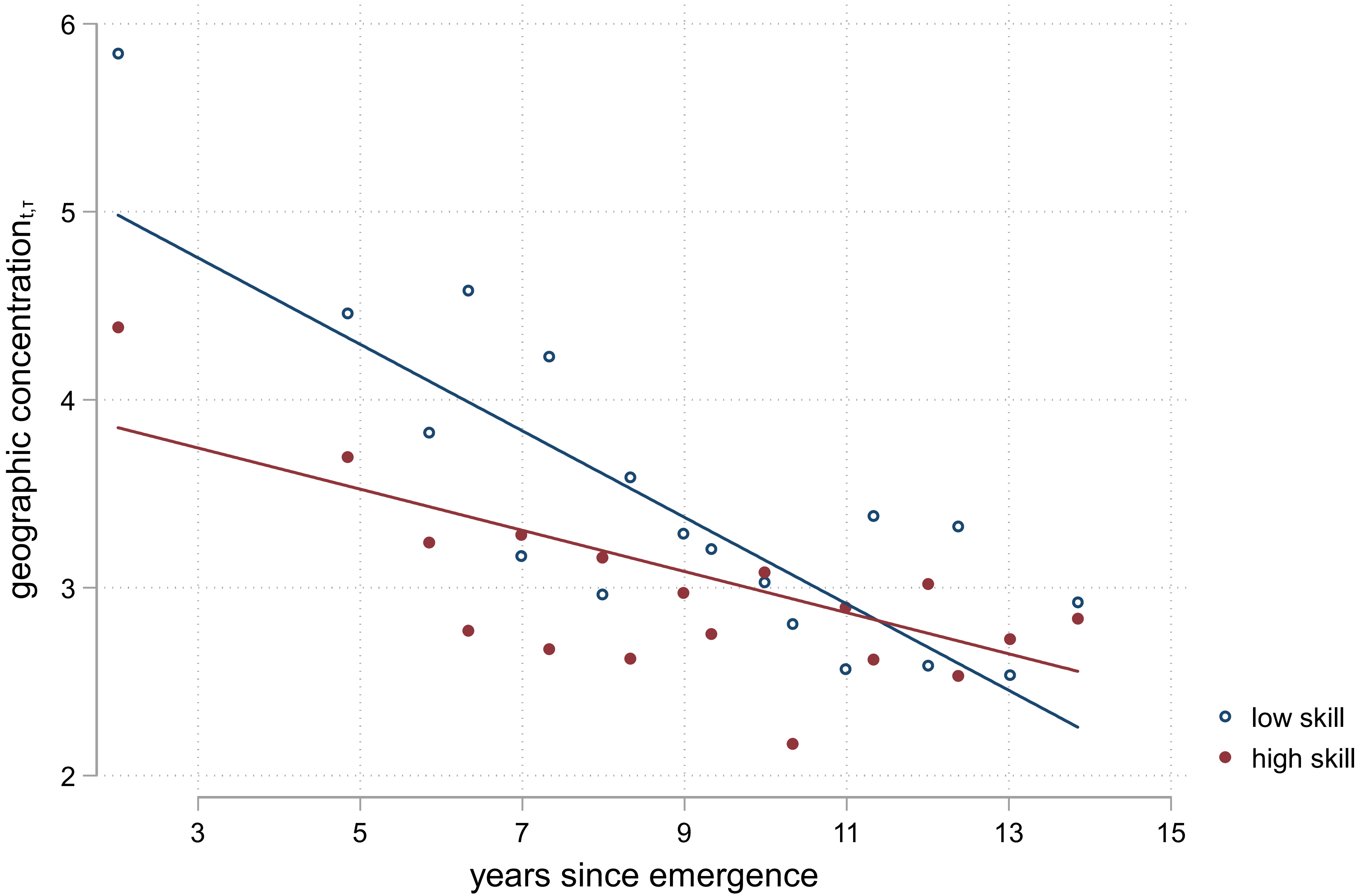
Notes: The figure plots approximate share of technology job postings which require college-educated people (red circles, where the size of the circle represents the total number of technology job postings) against year since emergence for the technology. Approximate share of college-educated people is calculated using $Skill_t^{\varsigma} = \frac{\sum_o N_{o;t}^{\tau} \chi_{o,2015}}{\sum_o N_{o;t}^{\tau}}$, where $\chi_{o,2015}$ is the share of college-educated people in an occupation in ACS 2015 and $N_{o;t}^{\tau}$ is the number of technology job postings in technology τ. Only observations post year of emergence are included.

Figure 8 - Share of college educated by year since emergence



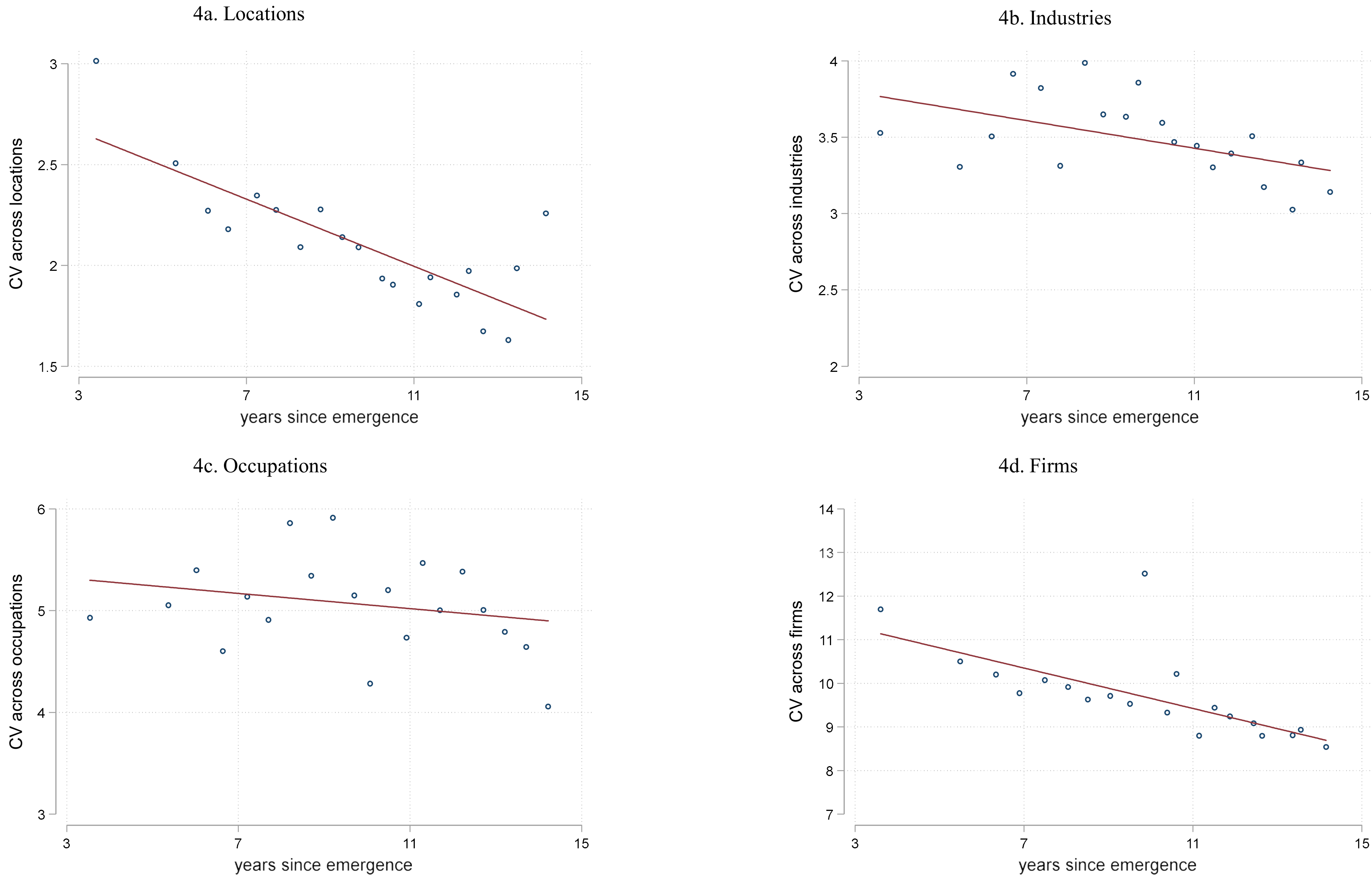
Notes: The figure plots a bin scatter of the approximate share of college-educated technology job postings (in percentage) for each technology and calendar year against the years since emergence of the technology. We weight observations by square root of hiring in that technology x year. The approximate share of college-educated job postings for a technology is measured as $Skill_t^{\varsigma} = \frac{\sum_o N_{o;t}^{\tau} \chi_{o,2015}}{\sum_o N_{o;t}^{\tau}}$, where $\chi_{o,2015}$ is the share of college-educated people in an occupation in ACS 2015 and $N_{o;t}^{\tau}$ is the number of technology job postings in technology τ. Only observations post year of emergence are included. The figure controls for technology fixed effects.

Figure 9 – Coefficient of Variation by year since emergence of technology



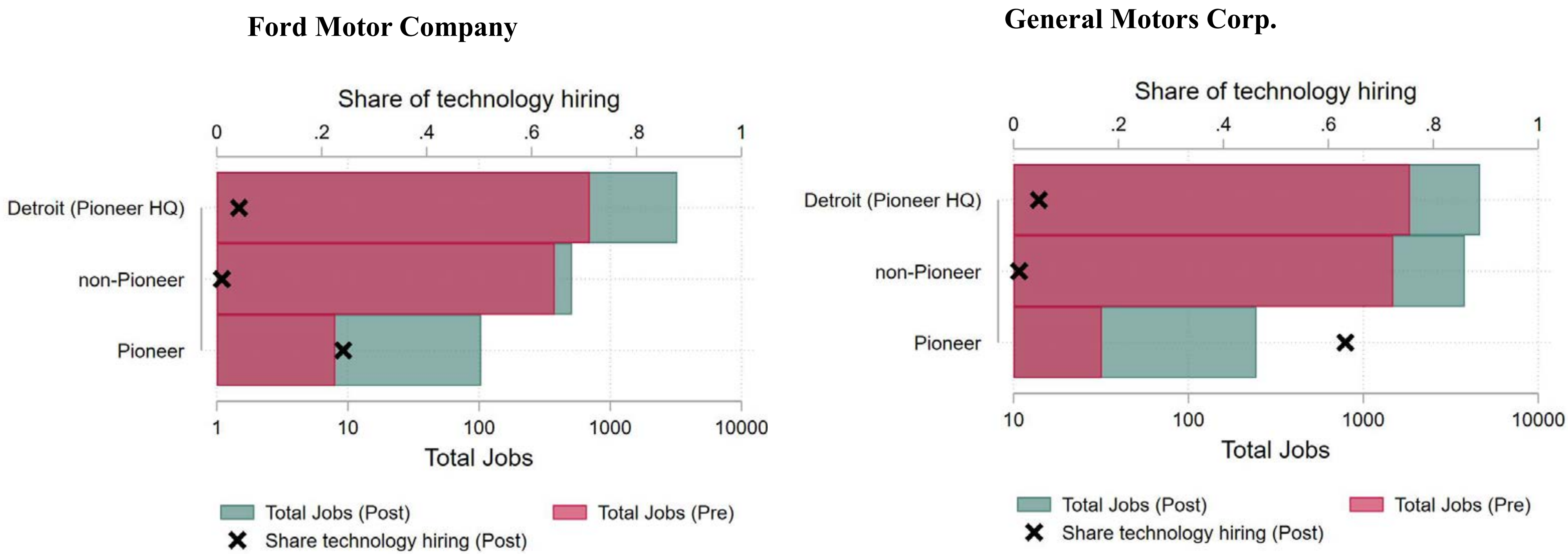
Notes: This figure plots a binned scatter plot with 30 bins of coefficient of variation by technology and time against year since the emergence of the technology for high skill and low skill occupations. The picture controls for technology fixed effects, and we weigh observations by square root of hiring in technology x year observations. To calculate the coefficient of variation by skill, we aggregate the job postings data over occupation, CBSA and year, and then separately for high skill occupations (with share of college-educated people > 60%), medium skill occupations (with share of college-educated people > 30% and < 60%), and low skill occupations (with share of college-educated people < 30%). Coefficient of variation is calculated over the normalized share of technology jobs by skill group, technology, and time.

Figure 10 – Coefficient of Variation across Locations, Industries, Occupations, Firms



Notes: The figure is a binned scatter plot of the coefficient of variation against year since emergence for our panel of 29 technologies. Coefficient of Variation is calculated across $Normalized\ share_{i,\tau,t} = \frac{share\ jobs\ expsed_{i,\tau,t}}{share\ jobs\ expsed_{\tau,t}}$ (where i is a location (4a), industry (4b), occupation (4c), or firm (4d)) for each technology τ and time t. Location refers to a CBSA, industry is at NAICS 4-digit level, and occupation is at SOC 6-digit level. Coefficient of variation in column (4) is calculated across 10,231 firms which have at least one job posting in each of the 11 years of Burning Glass. The results only include observations at the time of and after the emergence year of a technology, and observations with more than 100 technology jobs which have an industry associated with them. The binscatter is weighted by square root of total technology postings in a year. Normalized share is capped at 99th percentile of non-zero observations.

Figure 11 – Rehoming of firms to Technology hubs



Notes: The above figure plots decomposition of footprint between technology pioneers (HQ – when HQ coincide with pioneer), technology pioneers, and other locations. This decomposition is done for Ford Motor Company (on the left) and General Motor Corporation (on the right). There are 4 hubs for Autonomus car technology: San Jose-Sunnyvale-Santa Clara (CA), San Francisco-Oakland-Hayward (CA), Boston-Cambridge-Newton (MA-NH), Detroit-Warren-Dearborn (MI). Headquarter for Ford and GM are located in Detroit-Warren-Dearborn (MI), which is labeled as Hub-HQ. Pre/post-period means period before/after year of emergence calculated for Autonomous Cars.

Appendix Tables and Figures

Appendix Table 1 - Top technical bigrams in patents

Bigram	Citations (stand.)	Bigram	Citations (stand.)
readable medium	204379	vapor deposition	61255
user interface	197444	image data	59499
readable storage	131350	fiber optic	59395
fluid communication	117896	personal computer	59216
storage media	97478	volatile memory	58589
electrically conductive	96715	computer executable	58501
transitory computer	85426	acceptable carrier	57709
readable media	82035	disposed adjacent	55760
conductive material	80232	computing system	55655
machine readable	80079	optical fiber	54593
user input	76641	disk drive	54282
polyethylene glycol	75889	plan view	53676
data stored	72884	digital data	52337
proc natl	71562	acceptable salts	52291
natl acad	71523	graphical user	52256
computer implemented	71392	electrically coupled	51310
acid sequence	69521	dielectric layer	50642
pharmaceutical compositions	69082	temperature sensor	50612
positioned adjacent	69018	polymeric material	50360
pharmaceutical composition	68895	acceptable salt	48690
data structures	68401	data stream	48320
service provider	67424	network interface	47779
output signals	66494	support surface	47497
data structure	62563	acid sequences	47484

Notes: This table provides top 50 out of 35,063 bigrams by their citation score in patents. Citation score for a bigram is calculated as $\sum_{p\,|\,i\varepsilon p} Citations_p$, which is the sum of citations attributed to all patents which contain the bigram i and are applied for between 1975 and 2015. Citations to the patent p are standardized by the average citations in its primary technology class.

Appendix Table 2 - Example of human audit – “Autonomous car” keywords

Bigrams	True Positive Rate	Comments	Status
autonomous vehicle*	100%		Keep
autonomous vehicles*	100%		Keep
autonomous driving*	100%		Keep
self-driving car	90%		Keep
automated car	0%	-automated car washes.	Drop
robotic car	0%	- robotic car wash, - "robotics, car," - Shelley the robotic car from a video	Drop
robot car	100%		Keep
driverless car	90%		Keep
driverless truck	100%		Keep
autonomous car	100%		Keep
driver assistance	0%	- [role of a] senior living team members who is performing in a **driver assistance** role, spotter, or resident care.	Drop
automated driving	100%		Keep
autonomous cars	100%		Keep

Notes: The table presents results for human audit on keywords for “autonomous car” technology. For the audit, we go through each of the shortlisted bigrams (in Column 1) and randomly sample 10 job postings from BG between 2007 and 2019. Column 2 presents the true positive rate and Column 3 shows comments from auditor. Column 4 shows whether we keep or drop the bigram for the final list. * Bigrams are ones which we originally obtained from the intersection of patent corpus with earnings calls.

Appendix Table 3 - Keywords by technology

Technology	Keywords
d printing	d printer; d printing; additive manufacturing; d printed
Autonomous Cars	Self-driving car; robot car; autonomous vehicles; autonomous car; autonomous cars; automated driving; driverless car; autonomous driving; autonomous vehicle; driverless truck
Bispecific monoclonal antibody	bispecific monoclonal; the bispecific; bispecific antibody
Cloud computing	paas; cloud infrastructure; distributed cloud; cloud provider; cloud offerings; cloud service; cloud applications; community cloud; private cloud; public cloud; cloud deployments; cloud environments; cloud management; cloud services; cloud security; enterprise class; iaas; hybrid cloud; cloud platform; cloud providers; cloud hosting; personal cloud; enterprise network; cloud computing; cloud based; saas; cloud storage; enterprise applications; cloud solution; enterprise cloud; cloud solutions; cloud deployment
Computer vision	pose estimation; motion estimation; visual servoing; facial recognition; gesture recognition; computer vision; image recognition; sensor fusion; object recognition
Drug conjugates	kinase inhibitor; drug conjugate; antibody drug; drug conjugates
Electronic gaming	social game; video games; social games; video game; game content; electronic gaming; gaming products
Millimeter wave	millimeter wave
Fingerprint sensor	fingerprint sensor; fingerprint scanner
Fracking	fraccing; hydrofracking; hydrofracturing; hydraulic fracturing; fracking
GPS	gps systems; global positioning; navigation devices
Hybrid vehicle electric car	hybrid vehicle; electric vehicle; electric motorcycle; vehicle charging; hybrid electric; plugin hybrids; electric buses; electrical vehicles; electric car; electric vehicles
Lane departure warning	lane departure; departure warning
Lithium battery	ion battery; lithium ion battery; lithium ion batteries; lithium batteries; ion batteries; lithium polymer; lithium ion; lithium battery
Machine Learning AI	neural network; deep learning; language processing; machine learning; machine intelligence; natural language; artificial intelligence; ai technology; supervised learning; learning algorithms; unsupervised learning; reinforcement learning; ai machine
Mobile payment	mobile transfer; mobile commerce; mobile payment; mobile wallet; mobile money
Oled display	Oled
Online streaming	streaming content; music streaming; interactive tv; live stream; digital video; video conferencing; online streaming; online video; mobile video; streaming services; streaming media; live video; video ondemand; live streaming; video ad; internet radio; video streaming; streaming video
Rfid tags	frequency identification; keyless entry; rfid tags; rfid
Search Engine	search engine; search engines
Smart devices	mobile devices; tablet computers; wearable devices; tablet pcs; smartphone tablet; android phones; media devices; smart phones; smart devices; smart tvs; smart speaker; smart watch; smart car; smart phone; iphone ipad; portable media; smart tablets; connected

	devices; smartphones tablets; android smartphones; phones tablets; android devices; smart refrigerator; smartcar; smartphone; smart tv; smart band
Social Networking	user generated; user generated content; social platforms; networking sites; social channels; social media; social networking; social networks; social network
Software defined radio	defined radio
Solar Power	solar wafer; rooftop solar; solar modules; solar cells; crystalline silicon; silicon solar; solar panel; solar power; solar wafers; solar energy; solar applications; solar module; solar cell; solar pv; solar grade; solar panels; photovoltaic; solar thermal
Stent graft	stent graft
Touch screen	touch controller; touch panel; capacitive touch; touchscreen; touch screens; touch sensor
Virtual Reality	virtual reality; augmented reality; mixed reality; extended reality
Wifi	wifi hotspots; wifi network; wifi; broadband connectivity; wireless networks
Wireless charging	wireless charging; inductive charging

Notes: This table shows, for each of our 29 technologies (in Column 1), the full set of 221 final keywords used to associate earnings calls, patents and job postings with the technology.

Appendix Table 4 - Technology Excerpts from Earnings Calls

Company	EC month	Excerpt	Category
Ambarella Inc	4/2018	results that are many times higher in terms of processing performance per watt In March we successfully demonstrated to customer and investors our fully| AUTONOMOUS VEHICLE or embedded vehicle autonomy on Silicon Valley Road EVA navigated various traffic scenarios presented by Silicon Valleys challenging urban environment The fully autonomous	Development
General Motors Co	7/2017	safely deploy our selfdriving electric vehicles in commercial ridesharing networks Last month GM became the first company to use mass production methods to build| AUTONOMOUS VEHICLES |growing our test fleet to We plan to deploy these vehicles in the challenging driving environment of San Francisco as well as Scottsdale Arizona	Production
Agenus Inc	10/2019	differentiated proofofmechanism of our potentially first or bestinclass agents These discoveries include Our nextgeneration CTLA AGEN our differentiated CD agonist AGEN our firstinclass Tregdepleting| BISPECIFIC ANTIBODY |AGEN and of course GS a bifunctional molecule now exclusively licensed to Gilead and being developed by them In summary this year we generated	Development
Cloudera Inc	4/2019	combined company road map which we rolled out in March of this year During this period of uncertainty we saw increased competition from the| PUBLIC CLOUD |vendors Second the announcement in March of Cloudera Data Platform our new hybrid and multicloud offering created significant excitement within our customer base CDP	Competition
NVIDIA Corp	7/2015	lot of very exciting development And were working with a lot of them because we have a platform that was really designed to fuse| COMPUTER VISION |cameras from all around the car As well as radars and LIDARS and sonars and be able to do path planning and all of	Development
Proto Labs Inc	1/2015	orders in addition we added capacity to our manufacturing facility in europe in we completed our first acquisition purchasing fineline an| ADDITIVE MANUFACTURING |or| D PRINTING |company based in raleigh north carolina the acquisition was completed last april and is highly complementary to proto labs roughly of our customers use	Acquisition

Cellectar Biosciences Inc	10/2017	collaboration with Acunova Therapeutics each provide these types of strategic benefits Avicenna provides us with the unique opportunity to collaborate with experts in the antibody| DRUG CONJUGATE |or ADC field Not only does this provide the opportunity to work with a very promising small molecule payload but it also allows	Development
L-3 Communications Holdings Inc	10/2002	metal detectors where they always make you take your shoes off This is a passive scanner as I told some of you It uses| MILLIMETER WAVE |It is nonintrusive and causes no harm or disease It will guarantee you won't have a weapon on you of any kind or be	Use
Oasis Petroleum Inc	1/2011	tell you is that the build in the backlog is really a function of the weather that we experienced and it is always difficult| FRACKING |wells in the winter but this year was particularly brutal So I think the build in the backlog was largely around the weather And then	Production
InvenSense Inc	7/2016	as they strive to enable improved locationbased services and mapping user experience A significant opportunity for increasing our mobile content is UltraPrint our ultrasonic| FINGERPRINT SENSOR |I am very pleased to report that we are on track with the development of this gamechanging technology and have successfully passed several technology	Development
Tesla Inc	4/2011	with our store opening in Santana Row in San Jose in April The goal here is really to engage and inform potential customers about| ELECTRIC VEHICLES |in general and the advantages of Tesla in particular and really to try to catch people before they have actually made a buying decision	Production
SunPower Corp	10/2006	then be able to participate in the global electricity market which is measured in the form of trillion We have direct control over the| SOLAR CELL |and| SOLAR PANEL |portions of the value chain the technology core of the value chain that represents to of total installed costs In these	Production
Vocus Inc	1/2011	content distribution along with our expansion into| SOCIAL MEDIA |Vocus is uniquely positioned to help organizations of all sizes reach and influence buyers across| SOCIAL NETWORKS |online and through the media While PR will remain a core element of the Vocus product suite we believe there is a new and	Use
Donnelley Financial Solutions Inc	4/2018	speed and improve both the quality and consistency of business results for our clients In capital markets through the introduction of| MACHINE LEARNING |and| ARTIFICIAL INTELLIGENCE |we will improve the efficiency of XBRL tagging	Use

		and align with the efforts at the SEC to move from documents to data This investment	
Millennial Media LLC	4/2013	how We recently released our new Software Development Kit or SDK which is designed to enhance monetization of apps across| SMARTPHONES TABLETS |and other| CONNECTED DEVICES |SDK enhances our video advertising and rich media capabilities while adding new functionality like interactive voice ads and integration with iOSs Passbook for coupon	Use

Notes: This tables presents 15 earning calls excerpts (in column 3) with 25 words before and after technology keyword mentions, with the firm (in column 1), the date of the earnings call (in column 3). In column 4, we manually categorize each excerpt into 4 categories: 1) Use, 2) Produce, 3) Development, and 4) Acquisition.

Appendix Table 5 - Human audit of job postings

Panel A: Audit Results			
Audit	Use	Produce	Total
Describes company	6%	10%	16%
Describes Task	46%	34%	80%
Neither	NA	NA	4%
Panel B: Audit Results after clipping top 50 and bottom 50 words			
Audit	Use	Produce	Total
Describes company	0%	5%	6%
Describes Task	30%	58%	88%
Neither	NA	NA	6%

Panel C: Examples Excerpts	
Describes company - Produce	"[Company's] systems offer a unique combination of technology linking **RFID tags** and sensors with displays which permit users to track locate and observe movement of equipment and people in real time currently locates millions of patient's staff visitors and assets in healthcare facilities all over the world."
Describes task - Use	"passion for learning about new technology including low power RF technologies voice command systems motion control and **capacitive touch** ability to learn other non-electrical related topics mechanical and design considerations"
Neither	"our super cool office space which doesn't feel like an office is designed with our employees in mind techy surroundings a great outdoor space with Wi-Fi hookups for your laptop plus Bluetooth capabilities for **music streaming** we enjoy cultivating a supportive and all around positive culture that keeps our employees happy this will be a place you will want to come to everyday"

Notes: This tables presents results from a human audit of Burning Glass technology job postings. As a part of the human audit, we classify each of randomly sampled 100 job postings into two types of categories 1) whether the technology keyword describes the company in the job posting or the task content of the job posting, 2) whether the job describes use or production of the technology. See text for details. In Panel A, we perform the audit on original text of job postings. In Panel B, we clip the text of job postings by 50 words at the top and bottom, resample 100 postings, and then repeat the audit.

Appendix Table 6 - Posting summary statistics for technical and non-technical bigrams

Statistic	Supervised bigrams	Non-technical bigrams (top 221)	Technical bigrams (Unsupervised)	Non-technical bigrams* (top 305)	Non-technical bigrams (ext)* (top 4000)
# bigrams	221	221	305	305	4000
Avg. postings/bigram	59,013	142	49,677	157	474
Bigrams w/ more than 100 postings	88.3%	10.0%	92.4%	9.2%	8.1%

Notes: The table presents summary statistics (number of bigrams, average job postings per bigram, and bigrams with more than 100 job postings) for our list of supervised bigrams for 29 technologies (in column 2), top 221 non-technical bigrams (in column 3), unsupervised technical bigrams (in column 4), top 305 non-technical bigrams (in Columns 5) and top 4000 non-technical bigrams (in column 6). Technical bigrams are as described in section 2; we get to the list by intersecting bigrams in patents with bigrams in earnings calls. Non-technical bigrams are ones in earnings calls but not in patents. For both sets of bigrams, we restrict to the sample to bigrams for which share of firms increases in earnings calls (2002-2019).

*Through the aforementioned process, we obtained many more non-technical (104,627) bigrams than supervised bigrams (221) and technical bigrams (305). We restrict the sample to top (by frequency in earnings calls) 221 (in column 3), 305 non-technical bigrams (in Column 5) and 4,000 non-technical bigrams (in Column 6).

Appendix Table 7 - Top technical and non-technical bigrams

Top technical bigrams			Top non-technical bigrams		
bigram	# earnings	# job postings	bigram	# earnings	# job postings
mobile devices	6597	1078049	bofa merrill	34490	221
machine learning	2860	525286	stifel nicolaus	28877	256
cloud computing	2781	485333	division associate	12472	4237
cloud services	2450	380980	keefe bruyette	11682	16
quality metrics	2029	196497	bruyette woods	11498	14

Notes: The table presents the top five technical and non-technical bigrams. Technical bigrams are as described in section 2; we get the list by intersecting bigrams in patents with bigrams in earnings calls. Non-technical bigrams are ones in earnings calls but not in patents. For both sets of bigrams, we restrict to the sample to bigrams for which share of firms increases in earnings calls (2002-2019).

Appendix Table 8 - Top exposed occupations to virtual reality

Occupation	Total Postings	Exposed Postings	Pct. Exposed
Computer Hardware Engineers	100329	1000	1
Fine Artists, Including Painters, Sculptors, and Illustrators	67574	658	0.97
Multimedia Artists and Animators	75492	607	0.8
Computer and Information Research Scientists	233763	1630	0.7
Art Directors	84990	422	0.5
Sound Engineering Technicians	29187	140	0.48
Interior Designers	92453	382	0.41
Producers and Directors	152199	576	0.38
Astronomers	11905	45	0.38
Computer Science Teachers, Postsecondary	36470	134	0.37
Social Science Research Assistants	56496	207	0.37
Biomedical Engineers	18654	65	0.35
Film and Video Editors	16458	56	0.34
Instructional Coordinators	187871	587	0.31
Commercial and Industrial Designers	205700	632	0.31
Communications Teachers, Postsecondary	20412	62	0.3
Natural Sciences Managers	349157	1027	0.29
Helpers-Electricians	20492	60	0.29
Designers, All Other	226587	575	0.25
Library Technicians	31440	70	0.22
Atmospheric and Space Scientists	11806	26	0.22
Software Developers, Applications	8330098	18225	0.22
Computer and Information Systems Managers	228442	470	0.21
Web Developers	1819140	3625	0.2
Models	22230	44	0.2
Operations Research Analysts	983408	1943	0.2
Aerospace Engineers	69175	136	0.2
Graphic Designers	409096	795	0.19
Physicists	26118	46	0.18
Postsecondary Teachers, All Other	817030	1419	0.17

Notes: This table lists the top exposed occupations (Column 1), their overall postings in Burning Glass (Column 2), and the overall and percentage of jobs exposed to “virtual reality” (Column 3 and 4).

Appendix Table 9 - Top Exposed Occupations by Technology

Technology	Top Exposed Occupations
3d printing	Materials Engineers (1.11); Engineering Teachers, Postsecondary (0.58); Commercial and Industrial Designers (0.57);
Autonomous Cars	Computer Hardware Engineers (0.87); Computer and Information Research Scientists (0.43); Health and Safety Engineers, Except Mining Safety Engineers and Inspectors (0.41);
Bispecific monoclonal antibody	Biological Scientists, All Other (0.22); Biological Technicians (0.11); Chemists (0.08);
Cloud computing	Sales Engineers (13.15); Computer and Information Systems Managers (11.33); Information Security Analysts (11.10);
Extremely high frequency	Electronics Engineers, Except Computer (0.70); Radio, Cellular, and Tower Equipment Installers and Repairs (0.23); Physicists (0.14);
GPS	Surveyors (3.06); Surveying and Mapping Technicians (2.70); Biological Technicians (1.97);
Hybrid vehicle electric car	Wind Turbine Service Technicians (3.11); Power Plant Operators (2.30); Control and Valve Installers and Repairers, Except Mechanical Door (1.53);
Machine Learning AI	Computer and Information Research Scientists (40.67); Computer Science Teachers, Postsecondary (3.29); Computer Hardware Engineers (2.83);
Online streaming	Electronic Equipment Installers and Repairers, Motor Vehicles (31.38); Audio and Video Equipment Technicians (14.40); Film and Video Editors (8.03);
Search Engine	Writers and Authors (4.05); Advertising and Promotions Managers (3.09); Market Research Analysts and Marketing Specialists (2.85);
Smart devices	Automotive Glass Installers and Repairers (17.30); Electronic Equipment Installers and Repairers, Motor Vehicles (12.75); Merchandise Displayers and Window Trimmers (11.79);
Social Networking	Reporters and Correspondents (29.17); Public Relations Specialists (21.44); Market Research Analysts and Marketing Specialists (20.77);
Solar Power	Solar Photovoltaic Installers (45.34); Wind Turbine Service Technicians (5.87); Door-to-Door Sales Workers, News and Street Vendors, and Related Workers (4.08);
Virtual Reality	Fine Artists, Including Painters, Sculptors, and Illustrators (0.57); Computer Hardware Engineers (0.55); Computer and Information Research Scientists (0.54);
Wifi	Electronic Home Entertainment Equipment Installers and Repairers (31.73); Electronics Engineers, Except Computer (3.76); Computer Network Architects (3.38);

Wireless charging	Engineering Teachers, Postsecondary (0.08); Computer Hardware Engineers (0.07); Electronics Engineers, Except Computer (0.05);
computer vision	Computer Hardware Engineers (2.82); Computer and Information Research Scientists (2.82); Computer Science Teachers, Postsecondary (0.56);
drug conjugates	Chemical Technicians (0.69); Biological Scientists, All Other (0.65); Biochemists and Biophysicists (0.50);
electronic gaming	Demonstrators and Product Promoters (6.01); Fine Artists, Including Painters, Sculptors, and Illustrators (4.86); Gaming Service Workers, All Other (4.41);
fingerprint sensor	Physical Scientists, All Other (0.02); Computer Hardware Engineers (0.02); Electrical and Electronics Drafters (0.01);
fracking	Petroleum Engineers (1.98); Geoscientists, Except Hydrologists and Geographers (0.19); Community and Social Service Specialists, All Other (0.18);
injection molding	Tire Builders (35.31); Molders, Shapers, and Casters, Except Metal and Plastic (23.72); Molding, Coremaking, and Casting Machine Setters, Operators, and Tenders, Metal and Plastic (11.12);
inkjet printing	Printing Press Operators (0.07); Photographic Process Workers and Processing Machine Operators (0.05); Roustabouts, Oil and Gas (0.04);
lane departure warning	Heavy and Tractor-Trailer Truck Drivers (0.36); Mobile Heavy Equipment Mechanics, Except Engines (0.06); Traffic Technicians (0.04);
lithium battery	Materials Engineers (0.39); Life Scientists, All Other (0.25); Meter Readers, Utilities (0.21);
mobile payment	Food Scientists and Technologists (0.74); Maintenance Workers, Machinery (0.23); Sales Representatives, Wholesale and Manufacturing, Except Technical and Scientific Products (0.22);
oled display	Chemical Equipment Operators and Tenders (0.12); Home Appliance Repairers (0.11); Computer Hardware Engineers (0.05);
rfid tags	Locksmiths and Safe Repairers (0.75); Electronics Engineers, Except Computer (0.52); Purchasing Managers (0.44);
software defined radio	Electronics Engineers, Except Computer (0.67); Computer Hardware Engineers (0.45); Electrical Engineers (0.21);
stent graft	Health Diagnosing and Treating Practitioners, All Other (0.03); Sales Representatives, Wholesale and Manufacturing, Technical and Scientific Products (0.02); Cardiovascular Technologists and Technicians (0.02);
touch screen	English Language and Literature Teachers, Postsecondary (1.66); Audio and Video Equipment Technicians (1.04); Survey Researchers (0.52);

Notes: This table lists the top exposed occupations (in Column 2) for each of our 29 technologies (in column 1), and the share of online postings exposed to the technology (in brackets alongside each occupation).

Appendix Table 10 - Year of emergence by technology

	Year of Emergence	
Technology	EC (baseline)	Patents
3dprinting	2011	2013
autonomous cars	2014	2012
bispecific antibody	2012	1999
cloud computing	2008	2011
computer vision	2008	2006
drug conjugates	2002*	2002
electronic gaming	2002*	1995
millimeter wave	2014	2012
fingerprint sensor	2011	2005
Fracking	2007	2005
Gps	2002*	1999
hybrid vehicle/electric car	2007	2006
lane departure warning	2002*	2004
lithium battery	2002*	1994
machine learning/ai	2015	2005
mobile payment	2007	2007
oled display	2002*	2005
online streaming	2002*	1997
Rfid	2002*	2004
search engine	2002*	1997
smart devices	2005	2010
social networking	2006	2009
software defined radio	2002*	2005
solar power	2002*	1975
stent graft	2003	1995
touch screen	2002*	2010
virtual reality	2013	2012
Wifi	2002*	2007
wireless charging	2012	2012

Notes: The table provides our set of technologies (in Column 1), their year of emergence from earnings calls (Column 2), and their emergence year from patents (Column 3). The year of emergence is calculated as the first year that the share of firms mentioning technology in earnings calls reaches 10% of its maximum between 2002 and 2019. Years of emergence marked with * denotes technologies which reach more than 10% of their max share of firms in earnings calls in 2002. For these technologies, we impute the years to be 2002. In Column 3, the year of emergence as the year in which the share of U.S. patents for a technology reach 50% of their maximum value between 1976 and 2015.

Appendix Table 11 - Technology descriptions and contemporaneous events around emergence

Technology	Description	Emergence year	Contemporaneous Event
Smart devices	A smart device is an electronic device, generally connected to other devices or networks via different wireless protocols such as Bluetooth, Zigbee, NFC, Wi-Fi, LiFi, 5G, etc., that can operate to some extent interactively and autonomously.	2005	Apple announces first iPad. – Apple (2005)
Cloud computing	Cloud computing is the on-demand availability of computer system resources, especially data storage and computing power, without direct active management by the user.	2008	Microsoft and Google announced their cloud platforms. – Google and Microsoft blogs (2008)
Social Networking	The use of dedicated websites and applications to interact with other users, or to find people with similar interests to oneself.	2006	Mark Zuckerberg leaves Harvard. – Harvard Crimson (2005) Facebook receives $25 mill venture funding, and valued at half a billion. – Market Watch (2006)
Machine Learning/AI	Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves.	2015	Tesla's Elon Musk and venture capitalist Peter Thiel dedicated $1 billion to found Open AI, a non-profit for artificial intelligence research. – USA Today (2015)
Solar Power	Solar power is the conversion of energy from sunlight into electricity, either directly using photovoltaics (PV), indirectly using concentrated solar power, or a combination.	2002	Gov. Arnold Schwarzenegger announces plans for solar power subsidies. – Sacramento Bee (2005)
Autonomous Cars	A self-driving car, also known as an autonomous vehicle (AV), connected and autonomous vehicle (CAV), full self-driving car or driverless car, or robo-car or robotic car, (automated vehicles and fully automated vehicles in the European Union) is a vehicle that is capable of sensing its environment and moving safely with little or no human input.	2014	Google unveiled its first "fully functional" prototype for a self-driving car Monday and plans to test it on Bay Area public roads in the new year. – Mercury News, The (2014)
Virtual Reality	Virtual reality (VR) refers to a computer-generated simulation in which a person can interact within an artificial three-dimensional environment using electronic devices, such as special goggles with a screen or gloves fitted with sensors.	2013	Oculus raises $16 million in venture funding for virtual reality headset. – The Verge (2013)
Search Engine	A search engine is a software system that is designed to carry out web searches (Internet searches), which means to search the World Wide Web in a systematic way for particular information specified in a textual web search query.	2002	Sausalito start-up Groxis released a new search tool that categorizes search results in a more visually friendly way. - Mercury News, The (2003)
Hybrid vehicle/electric car	Any land-based autonomotive which uses electricity as one of the power sources.	2007	Toyota announces its plans for a plug-in hybrid car. – New York Times (2008). The Obama Administration lends Tesla Motors $465 million to build an electric sedan and the battery packs needed to propel it. – Wired (2009)
Wireless charging	Inductive charging (also known as wireless charging or cordless charging) is a type of wireless power transfer. It uses	2012	General Motors invest $5 million in wireless charging start-up Powermat. – Reuters (2012)

	electromagnetic induction to provide electricity to portable devices.		
touch screen	The touchscreen enables the user to interact directly with what is displayed, rather than using a mouse, touchpad, or other such devices (other than a stylus, which is optional for most modern touchscreens).	2003	Santa Clara county uses touch machines for voting. San Jose Mercury News (2003)
drug conjugates	Antibody-drug conjugates or ADCs are a class of biopharmaceutical drugs designed as a targeted therapy for treating cancer.	2002	Seattle Genetics signed a licensing deal granting MedImmune rights to use its antibody-drug-linking technology in research against a single biological marker of cancer. – Seattle Times, The (2005)
fracking	Hydraulic fracturing, also called fracking, fracing, hydrofracking, fraccing, frac'ing, and hydrofracturing, is a well stimulation technique involving the fracturing of bedrock formations by a pressurized liquid.	2007	Congress signs fracking as an exemption from the Safe Drinking Water Act. - Denver Post, The (CO) (2003)
software defined radio	Software-defined radio (SDR) is a radio communication system where components that have been traditionally implemented in hardware (e.g. mixers, filters, amplifiers, modulators/demodulators, detectors, etc.) are instead implemented by means of software on a personal computer or embedded system.	2002	Boeing was awarded a $220 million subcontract to Northrop Grumman's Radio Systems business in San Diego to expand development of the communications, navigation and identification system specializing in software-defined radios for the Army's Comanche helicopter. San Diego Union-Tribune, The (2003)
Wifi	Wi-Fi is a family of wireless network protocols, based on the IEEE 802.11 family of standards, which are commonly used for local area networking of devices and Internet access.	2002	San Francisco officials invited responses from 17 companies - including Google - that are interested in bringing affordable wireless Internet connections to the entire city. – Mercury News, The (2005)
3d printing	3D printing, or additive manufacturing, is the construction of a three-dimensional object from a CAD model or a digital 3D model.	2011	Federal government released plans to spend $45 million to help fund a planned additive manufacturing institute. - USA Today (2012)
Millimeter Wave	Extremely high frequency (EHF) or Millimeter Wave is the International Telecommunication Union (ITU) designation for the band of radio frequencies in the electromagnetic spectrum from 30 to 300 gigahertz (GHz).	2014	Facebook develops millimeter-wave networks for Internet.org. – The Verge (2016)
GPS	The Global Positioning System (GPS), originally NAVSTAR GPS,[1] is a satellite-based radionavigation system owned by the United States government and operated by the United States Space Force.[2]	2002	The Clinton administration removes "Selective Availability" of civilian GPS in order to make it more useful worldwide. – GPS.gov (2000)
Lithium-ion Battery	A lithium-ion battery or Li-ion battery is a type of rechargeable battery.	2002	Sion Power Corp. starts production of a new lithium-sulfur battery that can last twice as long as the previous model commonly used in laptops, cell phones and digital cameras. - Arizona Daily Star, The (2004)

OLED display	An organic light-emitting diode (OLED or organic LED), also known as organic electroluminescent (organic EL) diode, is a light-emitting diode (LED) in which the emissive electroluminescent layer is a film of organic compound that emits light in response to an electric current.	2002	Kodak announced the first consumer product to include a full-color, active-matrix organic light-emitting diode (OLED) display on the Kodak EasyShare LS633 digital camera. - Mercury News, The (2003)
Stent Graft	In medicine, a stent is a metal or plastic tube inserted into the lumen of an anatomic vessel or duct to keep the passageway open, and stenting is the placement of a stent.	2003	A stent graft system designed to correct life-threatening thoracic aortic aneurysms is fast track approved by the Food and Drug Administration. - Houston Chronicle (2003)
RFID	Radio-frequency identification (RFID) uses electromagnetic fields to automatically identify and track tags attached to objects. An RFID tag consists of a tiny radio transponder; a radio receiver and transmitter.	2002	Wal-Mart Stores ordered its 100 top suppliers to begin using RFID tags on shipments beginning in January 2005. - Mercury News, The (2003)
Electronic Gaming	An electronic game is a game that employs electronics to create an interactive system with which a player can play.	2002	Sony launches PlayStation 2 capable of playing video games from DVDs. Gamespy.com (1999) Microsoft launches Xbox, first mainstream device with online capabilities. Xbox.com (2000)
Computer Vision	Computer vision is an interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images or videos.	2003	The state of Illinois processes 10 million driver's license images using facial recognition. Chicago Sun-Times (2002)
Lane Departure Warning	In road-transport terminology, a lane departure warning system (LDWS) is a mechanism designed to warn the driver when the vehicle begins to move out of its lane (unless a turn signal is on in that direction) on freeways and arterial roads.	2002	Iteris and DaimlerChrysler develop a first Lane Departure Warning System. The device is mounted on a truck's windshield. It houses a tiny camera, computer and software that tracks the difference between the road and visible lane markings. Seattle Times, The (2003)
Bispecific monoclonal antibody	A bispecific monoclonal antibody (BsMAb, BsAb) is an artificial protein that can simultaneously bind to two different types of antigen. BsMabs can be manufactured in several structural formats, and current applications have been explored for cancer immunotherapy and drug delivery.	2012	Novartis Pays Genmab $2M to research into Bispecific Antibody Technology - Genetic Engineering and Biotechnology News (June 2012)
Fingerprint sensor	A fingerprint sensor is an electronic device used to capture a digital image of the fingerprint pattern.	2011	Apple buys fingerprint sensor firm AuthenTec for $356 million. - ZDNet (July 2012)
Mobile payment	Mobile payment (also referred to as mobile money, mobile money transfer, and mobile wallet) generally refer to payment services operated under financial regulation and performed from or via a mobile device.	2007	Bank of America, Citibank, Wachovia, Washington Mutual, Wells Fargo, and ING Direct announce mobile banking services, including mobile payment services. - CNBC (June 2007)
Online streaming	Streaming media is multimedia that is constantly received by and presented to an end-user while being delivered by a provider.	2002	Apple invested $12.5 million in Akamai, a content delivery company, with the aim to develop video streaming services for QuickTime TV. Akamai.com (1999)

Notes: The table above lists our 29 technologies (in Column 1), descriptions for each technology taken from Wikipedia (Column 2), emergence year (Column 3), and a suggested contemporaneous event around the year of emergence (Column 4).

Appendix Table 12- Summary statistics

	Mean	SD	p25	p50	p75	N
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Location						
Normalized Share	0.533	3.801	0	0	0.301	266467
University assets per capita	5670.065	12314.42	411.181	2234.962	5455.663	917
University enrollment per capita	0.141	0.178	0.024	0.1	0.171	917
Share of college educated	0.198	0.065	0.148	0.182	0.237	917
Share of post-graduate	0.068	0.028	0.049	0.06	0.081	917
Coefficient of Variation	3.716	2.471	1.729	3.033	5.181	249
Panel B: Industry						
Normalized Share	1.4	18.267	0	0	0.253	88490
Coefficient of Variation	4.885	2.315	3.433	4.429	5.867	249
Panel C: Occupation						
Normalized Share	0.704	3.851	0	0	0.075	238777
Share College Educated	54.774	13.125	47.462	56.785	63.325	249
Share Post Graduate	19.098	7.054	14.825	18.18	22.309	249
Wage	63044.66	11143.04	57776.79	64014.05	71611.97	249
Years of Schooling	14.993	0.769	14.579	15.064	15.403	249
Coefficient of Variation	6.654	3.32	3.914	5.853	8.909	249
Panel D: Firm						
Normalized Share	0.591	14.258	0	0	0	38990627
Coefficient of Variation	34.515	25.759	16.342	29.253	44.908	249

Notes: This tables shows summary statistics for variables used in the analysis of the paper. Summary statistics (columns 2-6) are shown for the pooled sample of technologies after year since emergence. In our sample, location is one of 917 Core-based statistical areas (CBSA), industry is one of 311 4-digit North American Industry Classification System (NAICS) codes, occupation is one of 836 six-digit Standard Occupation Classification Codes, and a firm is one of over 300K string clusters in Burning Glass Job Postings data. Normalized Share of technology jobs in all panels is calculated as $Normalized\ share_{i,\tau,t} = \frac{share\ jobs\ expsed_{i,T,t}}{share\ jobs\ expsed_{T,t}}$, where i is a location, industry, occupation, or firm (cell). Coefficient of Variation in all panels is calculated over normalized share of technology job postings over cells for technology x year observation. Location variables (in Panel A, row 2-5) are reported in the table for cross section of 917 CBSAs and calculated as following: university assets per capita is calculated as the total assets reported by unis in a CBSA in the Higher Education Research and Development Survey (HERD) and normalized by the population of the CBSA; enrollment per capital is calculated as the total enrollment reported by universities in a CBSA in HERD and normalized by the population of the CBSA. Skill level variables (in Panel C row 2-5) are calculated using $Skill_t^{\varsigma} = \frac{\sum_o N_{o;t}^{\tau} \chi_{o;2015}}{\sum_o N_{o;t}^{\tau}}$, where $\chi_{o;2015}$ is the share of college-educated people in an occupation in ACS 2015 and $N_{o;t}^{\tau}$ is the number of technology job postings in technology τ.

Appendix Table 13 - Top pioneer location by technology

Technology	Top CBSA pioneer	State
3dprinting	Boston-Cambridge-Newton	MA-NH
autonomous cars	San Jose-Sunnyvale-Santa Clara	CA
bispecific antibody	San Francisco-Oakland-Hayward	CA
cloud computing	San Jose-Sunnyvale-Santa Clara	CA
computer vision	San Jose-Sunnyvale-Santa Clara	CA
drug conjugates	Boston-Cambridge-Newton	MA-NH
electronic gaming	San Jose-Sunnyvale-Santa Clara	CA
millimeter wave	New York-Newark-Jersey City	NY-NJ-PA
fingerprint sensor	San Jose-Sunnyvale-Santa Clara	CA
fracking	Houston-The Woodlands-Sugar Land	TX
gps	San Jose-Sunnyvale-Santa Clara	CA
hybrid vehicle/electric car	Detroit-Warren-Dearborn	MI
lane departure warning	Grand Rapids-Wyoming	MI
lithium battery	Los Angeles-Long Beach-Anaheim	CA
machine learning/ai	San Jose-Sunnyvale-Santa Clara	CA
mobile payment	San Francisco-Oakland-Hayward	CA
oled display	Trenton	NJ
online streaming	San Jose-Sunnyvale-Santa Clara	CA
rfid	Grand Rapids-Wyoming	MI
search engine	San Jose-Sunnyvale-Santa Clara	CA
smart devices	San Jose-Sunnyvale-Santa Clara	CA
social networking	San Jose-Sunnyvale-Santa Clara	CA
software defined radio	Boulder	CO
solar power	San Jose-Sunnyvale-Santa Clara	CA
stent graft	San Francisco-Oakland-Hayward	CA
touch screen	San Jose-Sunnyvale-Santa Clara	CA
virtual reality	San Jose-Sunnyvale-Santa Clara	CA
wifi	New York-Newark-Jersey City	NY-NJ-PA
wireless charging	Boston-Cambridge-Newton	MA-NH

Notes: This table shows the top location hub (in Column 2) for each of our 29 technologies (Column 1), and their state (Column 3). We define pioneer locations as those which collectively accounted for 50% of the patent grants associated with a given technology applied for within ten years before its emergence. Top Pioneer location is the one with most patents.

Appendix Table 14 - Concentration during the life cycle - By skill levl

	Coefficient of Variation across Locations		
	(1)	(2)	(3)
	Low Skill	Medium Skill	High Skill
Years since	-0.154***	-0.169***	-0.097***
emergence	(0.049)	(0.048)	(0.033)
R2	0.841	0.851	0.916
N	231	231	231

Notes: This table reports the results from regressions of coefficient of variation during lifecycle of a technology on year since inception of the technology, separately for low skill occupations (Column 1), medium skill occupations (Column 2), and high skill occupations (Column 3). To calculate the coefficient of variation by skill, we aggregate the job postings data over occupation, CBSA and year, and then separately for high skill occupations (with share of college educated people > 60%), medium skill occupations (with share of college educated people > 30% and <60%), and low skill occupations (with share of college educated people < 30%). Finally, the coefficient of variation is calculated over $Normalized\ share_{cbsa,\tau,t,skill}$ across CBSAs by skill group, technology, and time. Standard errors are clustered by technology.

Appendix Table 15 - Pioneer Occupations and Industries by Technology

Technology	Top Pioneer Occupation (share of jobs at t0)	Top Pioneer Industry (share of jobs at t0)
3d printing	Mechanical Engineers (0.140)	Computer and Peripheral Equipment Manufacturing (0.419)
Autonomous cars	Computer Occupations All Other (0.186)	Motor Vehicle Manufacturing (0.370)
Bispecific monoclonal antibody	Operations Research Analysts (0.375)	Pharmaceutical and Medicine Manufacturing (0.946)
Cloud computing	Software Developers Applications (0.228)	Software Publishers (0.300)
Computer vision	Software Developers Applications (0.295)	Semiconductor and Other Electronic Component Manufacturing (0.174)
Drug conjugates	Natural Sciences Managers (0.135)	Pharmaceutical and Medicine Manufacturing (0.910)
Electronic gaming	Software Developers Applications (0.202)	Software Publishers (0.202)
Millimeter Wave	Electronics Engineers Except Computer (0.169)	Semiconductor and Other Electronic Component Manufacturing (0.371)
Fingerprint sensor	Software Developers Applications (0.203)	Semiconductor and Other Electronic Component Manufacturing (0.215)
Fracking	Geoscientists Except Hydrologists and Geographers (0.286)	Oil and Gas Extraction (0.881)
Gps	Computer Occupations All Other (0.173)	Communications Equipment Manufacturing (0.187)
Hybrid vehicle/electric car	Mechanical Engineers (0.151)	Motor Vehicle Manufacturing (0.681)
Lane departure warning	Mechanical Engineers (0.500)	Motor Vehicle Manufacturing (0.393)
Lithium battery	Electrical Engineers (0.188)	Commercial and Service Industry Machinery Manufacturing (0.115)
Machine learning ai	Software Developers Applications (0.251)	Other Information Services (0.225)
Mobile payment	Marketing Managers (0.154)	Semiconductor and Other Electronic Component Manufacturing (0.227)
Oled display	Engineers All Other (0.400)	Commercial and Service Industry Machinery Manufacturing (0.320)
Online streaming	Sales Representatives (0.095)	Semiconductor and Other Electronic Component Manufacturing (0.188)
Rfid tags	Architectural and Engineering Managers (0.098)	Computer and Peripheral Equipment Manufacturing (0.191)
Search engine	Marketing Managers (0.124)	Other Information Services (0.264)
Smart devices	Software Developers Applications (0.229)	Software Publishers (0.243)
Social networking	Marketing Managers (0.128)	Other Information Services (0.299)
Software defined radio	Software Developers Applications (0.489)	Communications Equipment Manufacturing (0.353)
Solar power	Mechanical Engineers (0.099)	Semiconductor and Other Electronic Component Manufacturing (0.243)
Stent graft	Physicians and Surgeons All Other (0.375)	Medical Equipment and Supplies Manufacturing (0.628)
Touch screen	Sales Representatives Wholesale (0.134)	Commercial and Service Industry Machinery Manufacturing (0.211)
Virtual reality	Software Developers Applications (0.198)	Semiconductor and Other Electronic Component Manufacturing (0.214)
Wifi	Retail Salespersons (0.255)	Communications Equipment Manufacturing (0.314)
Wireless charging	Computer Occupations All Other (0.222)	Semiconductor and Other Electronic Component Manufacturing (0.412)

Notes: The table shows the top occupation pioneer (Column 2) and top industry pioneer (Column 3) for each of our 29 technologies (in Column 1). A pioneer is defined as the set of occupations/industries/locations/firms that account for more than 50% of patents associated with the technology during the 10 years before year of emergence of the technology. The top pioneer is the one with most patents.

Appendix Table 16 - Robustness: Primary results

	(1) Region Broadening	(2) Pioneer Persistence	(3) Skill Broadening	(4) Region Broadening by Skill
Panel A: Patent Emergence Year				
Years since emergence (patents)	-0.070***		-0.727***	-0.121***
	(0.020)		(0.226)	(0.040)
Pioneer		1.369***		
		(0.410)		
Pioneer * Years since emergence (patents)		-0.033**		
		(0.014)		
Years since emergence (patents) * {skill = low}				-0.046*
				(0.026)
R2	0.893	0.077	0.880	0.750
N	255	275,751	255	510
Panel B: Without 2007				
	(1)	(2)	(3)	(4)
Years since emergence (patents)	-0.100***		-0.877***	-0.089*
	(0.036)		(0.272)	(0.046)
Pioneer		2.475***		
		(0.643)		
Pioneer * Years since emergence		-0.157***		
		(0.048)		
Years since emergence * {skill = low}				-0.184***
				(0.044)
R2	0.891	0.079	0.880	0.780
N	236	248,873	236	504
Panel C: Robust Standard Errors				
	(1)	(2)	(3)	(4)
Years since emergence (patents)	-0.092***		-0.919***	-0.110***
	(0.023)		(0.224)	(0.027)
Pioneer		2.313***		
		(0.202)		
Pioneer * Years since Emergence		-0.146***		
		(0.016)		
Years since emergence * {skill = low}				-0.195***
				(0.023)
R2	0.888	0.075	0.873	0.772
N	249	266,467	249	538

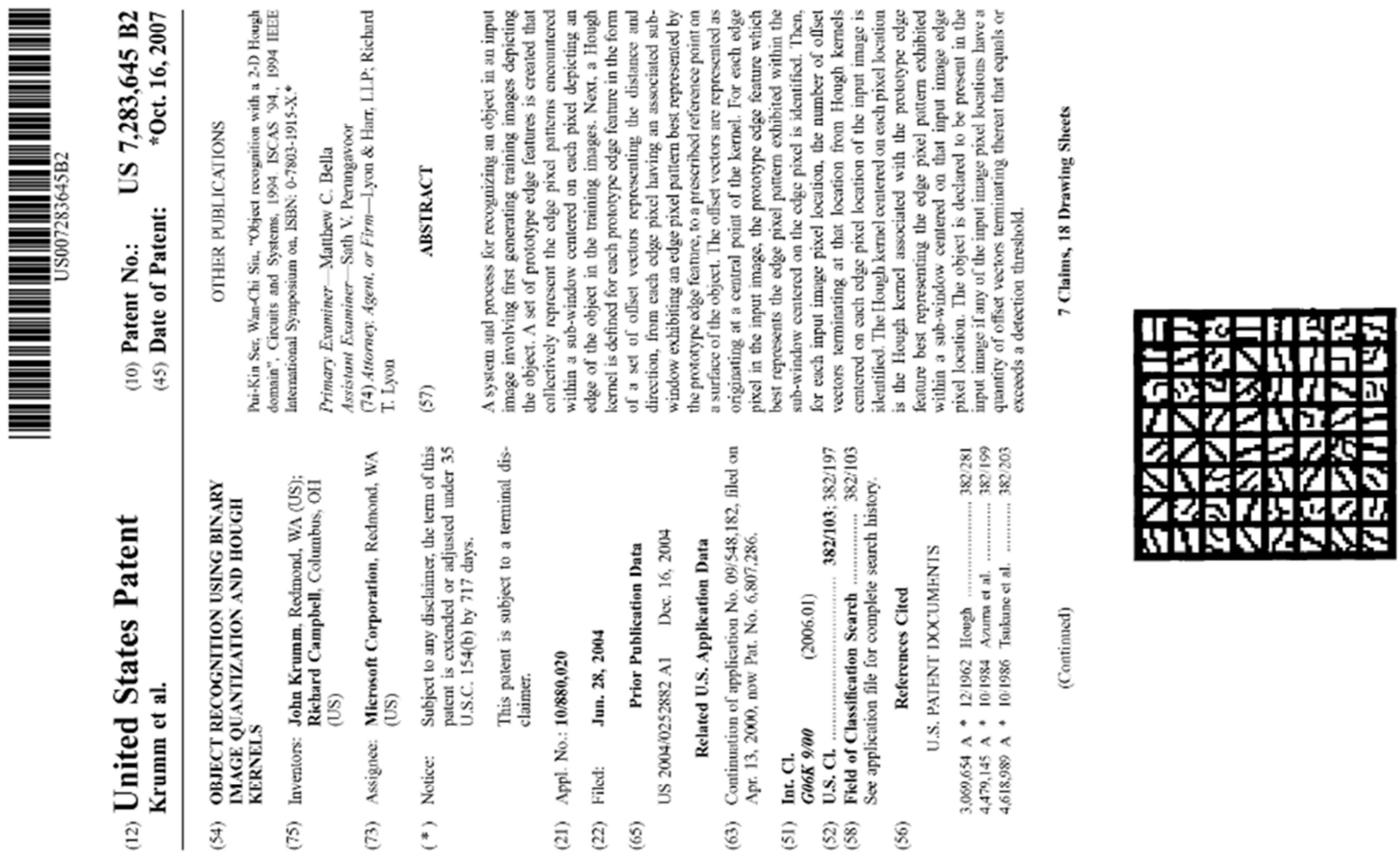
Notes: This table reports results for robustness checks for broadening results in Table 5. The results are from a regression of Coefficient of Variation calculated across $Normalized\ share_{i,\tau,t} = \frac{share\ jobs\ expsed_{i,\tau,t}}{share\ jobs\ expsed_{\tau,t}}$ (where i is a location (Column 1), industry (Column 2), occupation (Column 3) or firm (Column 4)) for each technology τ and time t on years since emergence for a technology. In Panel A, we calculate the year of emergence as the year in which the share of US patents for a technology reaches 50% of their maximum value between 1976 and 2015; in Panel B, we exclude the year 2007; in Panel C, we use robust standard errors instead of clustered in the baseline specification. These results exclude observations before the start year of a technology. Standard errors are clustered in Panel A and B; they are robust in Panel C.

Appendix Table 17 - Robustness: Skill broadening with sample reweighted to U.S. employment

	(1) Share of college educated * 100	(2) Share of post graduate * 100	(3) Avg. wage	(4) Avg. schooling
Year since emergence	-0.593**	-0.180	-627.958**	-0.035**
	(0.272)	(0.118)	(241.790)	(0.015)
R2	0.902	0.915	0.907	0.905
N	249	249	249	249

Notes: This table reports the results from robustness of our skill broadening result. We regress approximate skill composition of technology jobs $\text{Skill}_t^{\varsigma} = \frac{\sum_o N_{o;t}^{\tau} \chi_{o;2015}}{\sum_o N_{o;t}^{\tau}}$ on the left-hand side, where $\chi_{o;2015}$ is the skill measure of interest from ACS 2015 at the occupation level), on the years since inception of the technology on the right-hand side. In this case, the technology jobs in an occupation are reweighted according to hiring in the U.S. economy for each two-digit occupation. Hiring in a two-digit occupation is calculated using hiring in an industry in LEHD and then constructing a crosswalk between industry employment and occupation employment. These results exclude observations before the start year of a technology. Regressions are weighted by square root of technology job postings in a year. Standard errors are clustered by technology.

Appendix Figure 1 – Microsoft Computer Vision Patent



US007283645B2

(12) **United States Patent**
Krumm et al.

(10) **Patent No.: US 7,283,645 B2**
(45) **Date of Patent: *Oct. 16, 2007**

(54) **OBJECT RECOGNITION USING BINARY IMAGE QUANTIZATION AND HOUGH KERNELS**

(75) Inventors: **John Krumm**, Redmond, WA (US); **Richard Campbell**, Columbus, OH (US)

(73) Assignee: **Microsoft Corporation**, Redmond, WA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 717 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **10/880,020**

(22) Filed: **Jun. 28, 2004**

(65) **Prior Publication Data**

US 2004/0252882 A1 Dec. 16, 2004

Related U.S. Application Data

(63) Continuation of application No. 09/548,182, filed on Apr. 13, 2000, now Pat. No. 6,807,286.

(51) **Int. Cl.**
G06K 9/00 (2006.01)
(52) **U.S. Cl.** **382/103**; 382/197
(58) **Field of Classification Search** 382/103
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,069,654 A * 12/1962 Hough 382/281
4,479,145 A * 10/1984 Azuma et al. 382/199
4,618,989 A * 10/1986 Tsukune et al. 382/203

(Continued)

OTHER PUBLICATIONS

Pui-Kin Ser, Wan-Chi Siu, "Object recognition with a 2-D Hough domain", Circuits and Systems, 1994. ISCAS '94., 1994 IEEE International Symposium on, ISBN: 0-7803-1915-X.*

Primary Examiner—Matthew C. Bella
Assistant Examiner—Sath V. Perungavoor
(74) *Attorney, Agent, or Firm*—Lyon & Harr, LLP; Richard T. Lyon

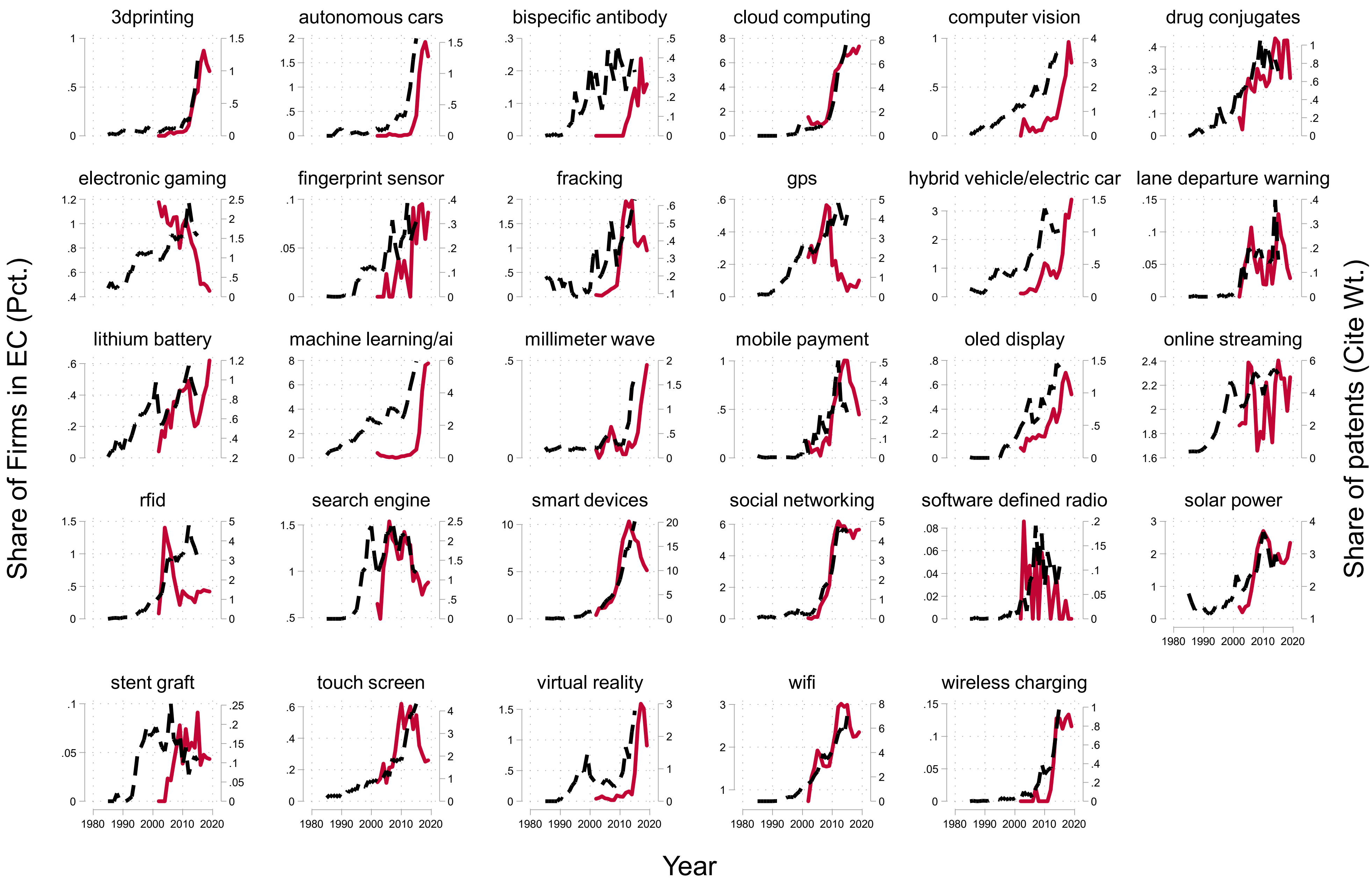
(57) **ABSTRACT**

A system and process for recognizing an object in an input image involving first generating training images depicting the object. A set of prototype edge features is created that collectively represent the edge pixel patterns encountered within a sub-window centered on each pixel depicting an edge of the object in the training images. Next, a Hough kernel is defined for each prototype edge feature in the form of a set of offset vectors representing the distance and direction, from each edge pixel having an associated sub-window exhibiting an edge pixel pattern best represented by the prototype edge feature, to a prescribed reference point on a surface of the object. The offset vectors are represented as originating at a central point of the kernel. For each edge pixel in the input image, the prototype edge feature which best represents the edge pixel pattern exhibited within the sub-window centered on the edge pixel is identified. Then, for each input image pixel location, the number of offset vectors terminating at that location from Hough kernels centered on each edge pixel location of the input image is identified. The Hough kernel centered on each pixel location is the Hough kernel associated with the prototype edge feature best representing the edge pixel pattern exhibited within a sub-window centered on that input image edge pixel location. The object is declared to be present in the input image if any of the input image pixel locations have a quantity of offset vectors terminating thereat that equals or exceeds a detection threshold.

7 Claims, 18 Drawing Sheets

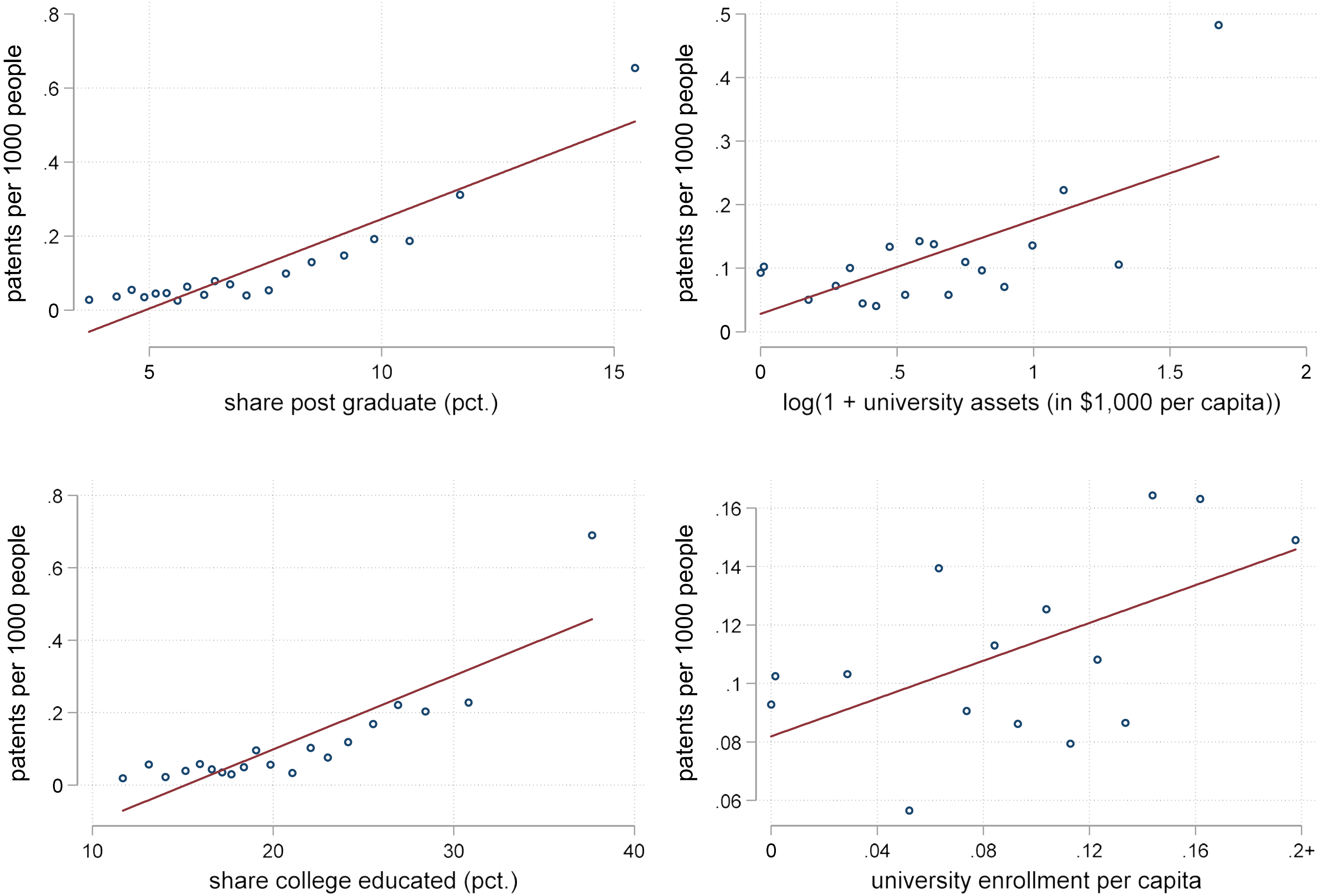
Notes: This figure illustrates a Microsoft patent applied for in 2004 and is exposed to the technology Computer Vision.

Appendix Figure 2- Patent (in black-dashed) and Earnings Call (in red-solid) Time Series



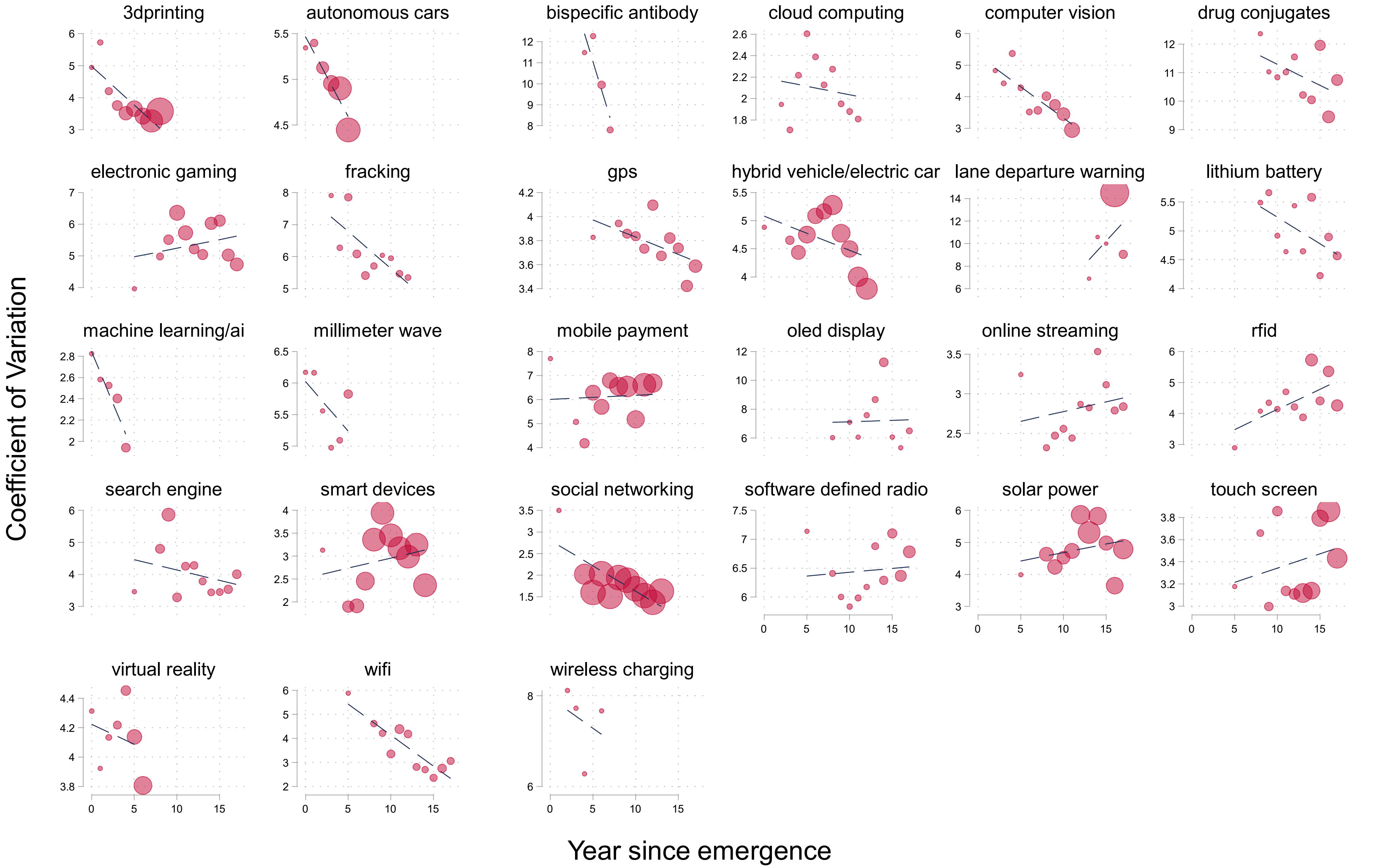
Notes: The figure plots share of firms in earnings calls exposed to each technology (in red-solid), and the share of cite weighted patents associated with each of our 29 technologies in (black-dashed). The sample of earnings calls (2002-19) and of patents (1985-2015). The overall correlation between the two time series is 80.26%.

Appendix Figure 3 - Technology innovation vs local skill composition



Notes: The figure plots a binned scatter plot of patents associated with a technology per 1000 people in a CBSA for each of our 29 technologies against repeated values of skill/university presence in the CBSA. Patents associated with a technology are calculated 10 years before the year of emergence of the technology. The figure controls for technology fixed effects.

Appendix Figure 4 – Coefficient of Variation across industries by year since emergence

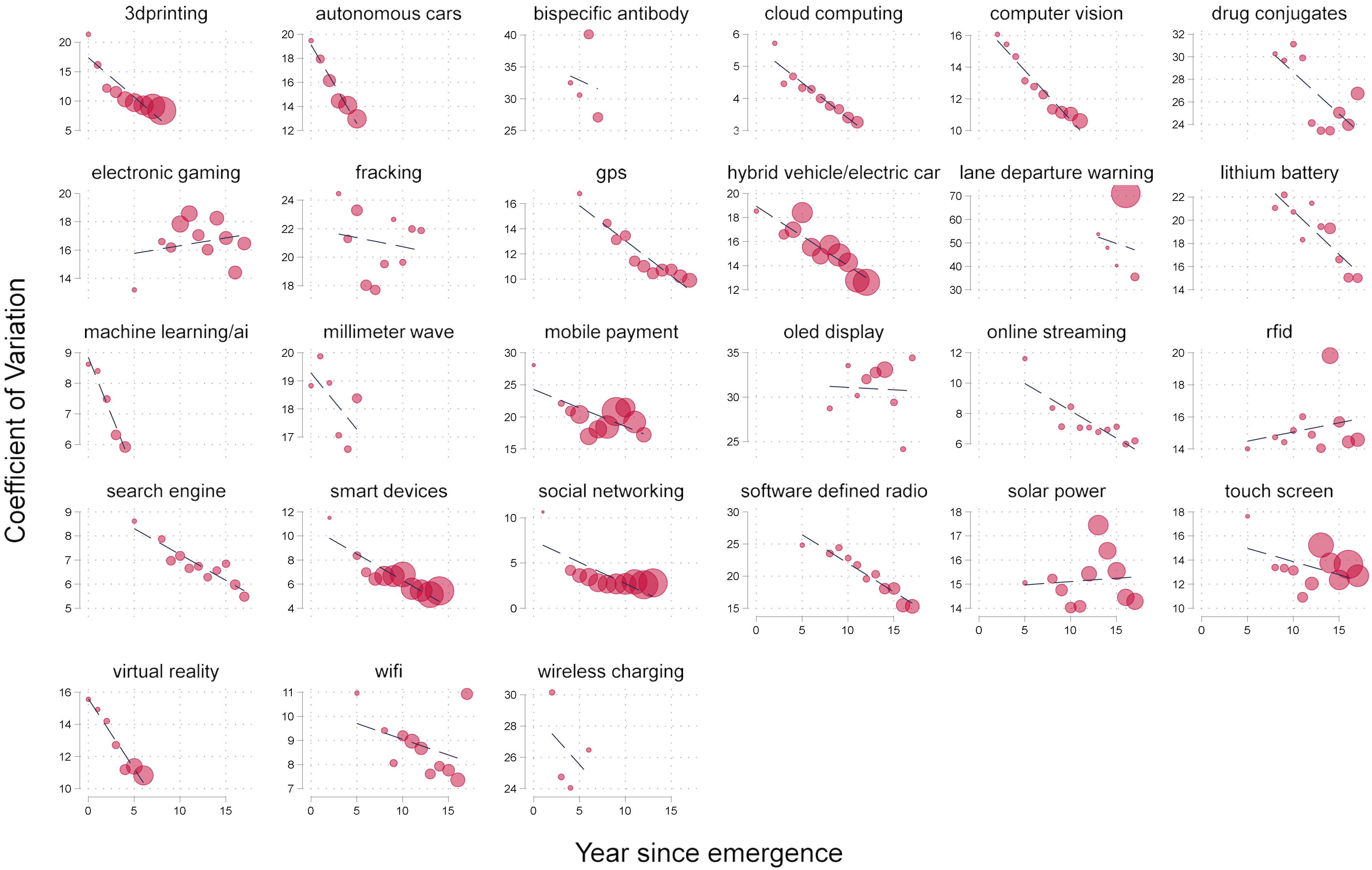


Notes: The figure plots coefficient of variation measured as coefficient of variation of normalized share of technology jobs for each of 29 technologies by year from 2007-2019 against the years since emergence of the technology, where $Normalized\ share_{i,\tau,t} = \frac{share\ jobs\ expsed_{i,\tau,t}}{share\ jobs\ expsed_{\tau,t}}$, where i is an industry.

Appendix Figure 5 – Coefficient of Variation across occupations by year since emergence

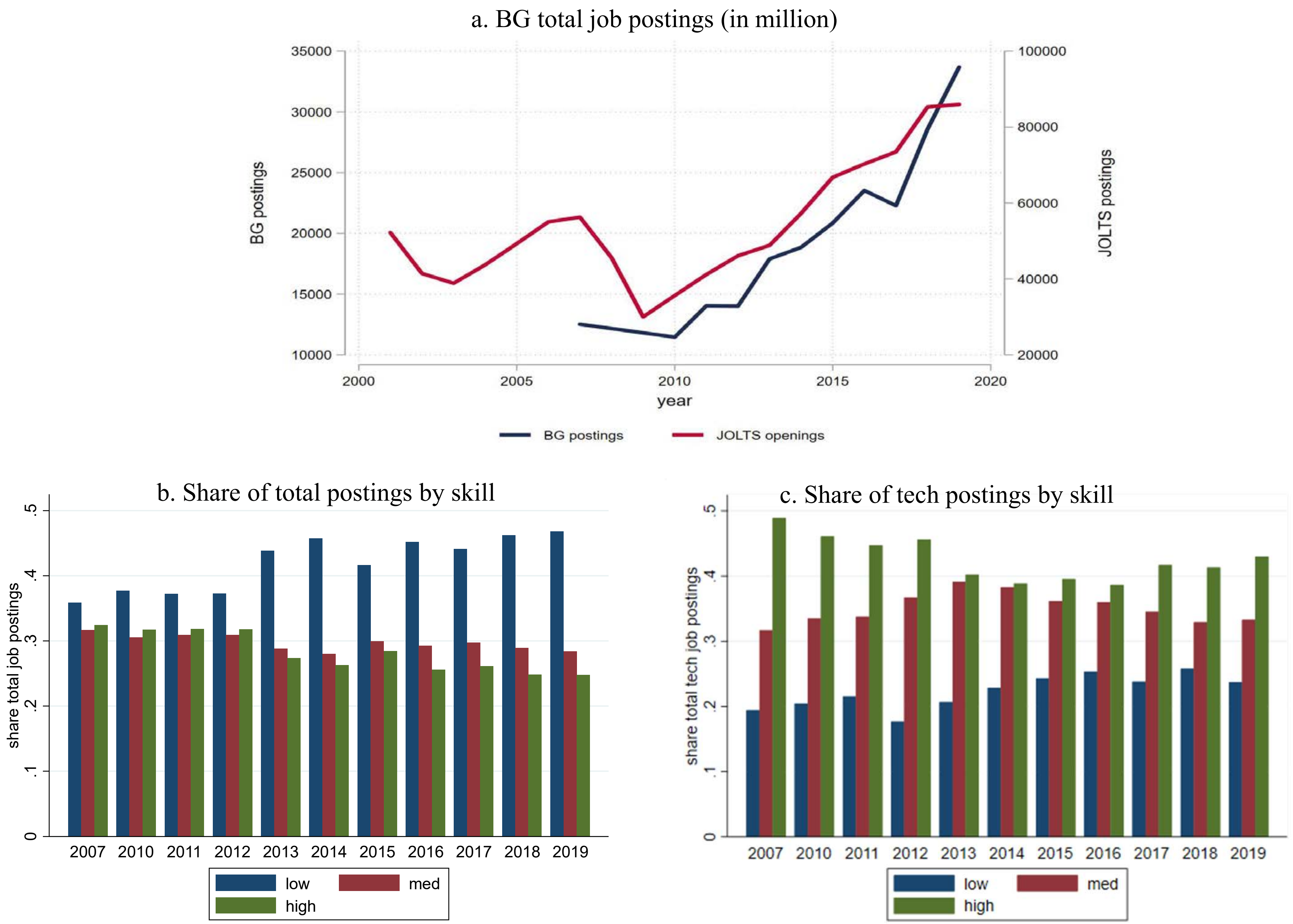
Notes: The figure plots coefficient of variation measured as coefficient of variation of normalized share of technology jobs for each of 29 technologies by year from 2007-2019 against the years since emergence of the technology, where $Normalized\ share_{i,\tau,t} = \frac{share\ jobs\ expsed_{i,\tau,t}}{share\ jobs\ expsed_{\tau,t}}$, where i is an occupation.

Appendix Figure 6 – Coefficient of Variation across firms by year since emergence



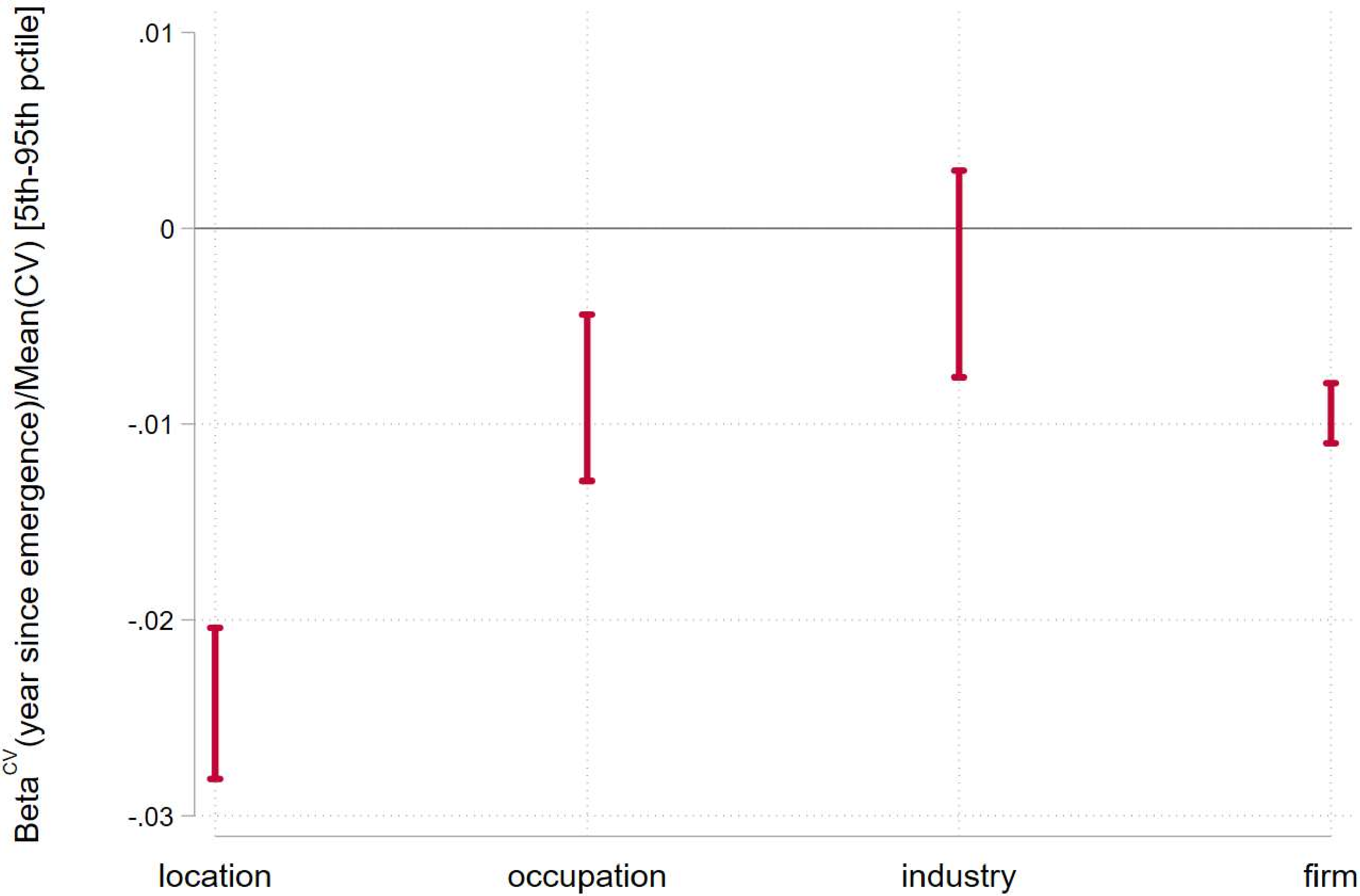
Notes: The figure plots coefficient of variation measured as coefficient of variation of normalized share of technology jobs for each of 29 technologies by year from 2007-2019 against the years since emergence of the technology, where $Normalized\ share_{i,\tau,t} = \frac{share\ jobs\ expsed_{i,\tau,t}}{share\ jobs\ expsed_{\tau,t}}$, where i is a firm.

Appendix Figure 7 – Overall Patterns of Burning Glass Job Postings



Notes: In this figure, we show aggregate patterns of Burning Glass (BG) online job postings. Figure a shows the total number of job postings (in millions) by year in BG. Figure b and c share of total job postings by skill and share of total technology job postings (aggregated over 29 selected technologies) by skill, respectively. To calculate skill level for job postings, we aggregate the data over occupation, and then use share of college-educated workforce from the ACS (2015) to assign them to high skill occupations (with share of college-educated people > 60%), medium skill occupations (with share of college-educated people > 30% and < 60%), and low skill occupations (with share of college-educated people < 30%).

Appendix Figure 8 – Coefficient of Variation across firms by year since emergence



Notes: This figure plots results from a jackknife estimate of regressions of coefficient of variation of normalized share of technology jobs calculated across locations, occupations, industries and firms. For our jackknife estimates, we exclude three technologies at a time and recalculate the degradation in coefficient of variation, this provides us with 7,308 permutation estimates. In this figure, we plot we plot the 10th and 90th percentile of these jackknife estimates