Institute for
New Economic Thinking

Working Papers
ineteconomics.org/workingpapers

The Geography of New Technologies

Nicholas Bloom, Tarek A. Hassan, Aakash Kalyani,
Josh Lerner, and Ahmed Tahoun[1]

Working Paper No. 126

June 15th, 2020

ABSTRACT

We identify novel technologies using textual analysis of earnings conference calls, newspapers, announcements, and patents. Our approach enables us to document the rollout of 20 new technologies across firms and labor markets in the U.S. Four stylized facts emerge from our data. First, as technologies develop, the number of new positions related to them grows, but the average education requirements and wage levels of the positions drop. Second, as technologies develop, their employment impact diffuses across the country: initially, technologies are concentrated in local hubs, but over time, their adoption diffuses geographically. Third, despite this diffusion, the initial hubs retain a disproportionate share of employment in the technology,

---

[1] Stanford University; Boston University; Boston University; London Business School; Harvard University. Bloom, Hassan, and Lerner are affiliated with the National Bureau of Economic Research. We thank seminar audiences at the Toulouse Network on Information Technology for helpful comments, as well as Lisa Kahn for data sharing. 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 Kauffman Foundation, Harvard Business School's Division of Research, the 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 and omissions are our own.

particularly at the high-skill end of the spectrum. Finally, technology hubs are more likely to arise in areas with universities and high skilled labor pools.

https://doi.org/10.36687/inetwp126

JEL Codes: O31, O32

Keywords: Technology, Geography, Employment, Innovation, R&D

1. Introduction

Rising inequality has focused attention on the benefits of new technologies.[2] Do these accrue primarily to inventors, early investors, and highly skilled users, or to society more widely as their adoption generates employment growth? This paper aims to trace out the impact of 20 new technologies on employment of low and high skilled workers, around the birthplace of the technology and across the US.

Policymakers are also focused on the location of technology, with a growing source of competition between nations and regions involving the attraction of nascent industries and technologies. Whether payments to attract professors specializing in artificial intelligence and biotechnology researchers, the construction of facilities to house nascent firms, or matching funds provided to angel or venture capital investors, governments have sought to establish an early presence in promising industries.[3]

In undertaking these initiatives, policymakers have been attracted by the success of technology clusters such as Silicon Valley, Herzliya, Shanghai, and Shenzhen. The success of the Boston area in biotechnology illustrates the positive dynamics that can occur. The initial wave of start-ups spinning out of Harvard and MIT in the 1980s and 1990s resulted in a number of significant enterprises, such as Biogen, Genzyme (now part of Sanofi), and Vertex Pharmaceuticals. The region has continued to attract new entrants, which raised almost $5 billion dollars in venture capital in 2018 alone. Finally, the presence of academic institutions and younger firms have attracted investments in major research facilities by established firms based elsewhere, such as Novartis, Pfizer, and Takeda. Between 2010 and 2018, the Massachusetts biopharmaceutical sector grew from 55 to 74 thousand jobs.[4]

But despite the many billions of dollars devoted annually to the attraction of infant industries, these policies have attracted relatively little systemic attention from economists (e.g., Bloom, Van Reenen, and Williams 2019). Motivating these policy interventions are two propositions. The first is that public interventions will effectively enhance the presence of an emerging industry in a given region; the second, that this initial activity will yield long-run impacts for the region. This paper examines the second of these claims.

To do so, we study the evolution of twenty new technologies in the United States from 2002 to 2020. We develop a methodology for systematically identifying two-word phrases, or "bigrams," associated with a new technology. We then look for documents linked to these new technologies, including U.S. patents, earning calls from publicly listed firms, and job announcements in the

---

[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.

[3] A few of many examples of such efforts include https://www.straitstimes.com/singapore/a-home-for-worlds-best-scientists; https://www.nytimes.com/2020/02/06/us/chinas-lavish-funds-lured-us-scientists-what-did-it-get-in-return.html; https://www.businessinsider.com/neom-what-we-know-saudi-arabia-500bn-mega-city-2019-9; and https://www.france24.com/en/20200325-france-to-support-start-ups-with-%E2%82%AC4-billion-plan-amid-coronavirus-crisis.

[4] This paragraph is taken from press accounts and Massachusetts Biotechnology Council (2019).

Burning Glass database. This evidence allows us to assess the impact of these technologies along a dimension of crucial importance to policymakers: employment. In particular, we examine the evolution of the number, location, and quality of new jobs associated with these new technologies.

The key results of this analysis are as follows:

- Earnings call mentions and hiring announcements linked to the new technologies rise in parallel over time;
- While initial hiring is focused on high-skilled jobs, over time the mean skill level in new positions associated with the technologies declines sharply, which we term a "skill-broadening" effect;
- New hiring in new technologies increases its geographic footprint over time, becoming less concentrated, which we dub "region broadening";
- The initial geographic hub retains an important advantage that persists over time. This pattern is particularly pronounced among high skill jobs; and
- Hubs are most likely to arise around universities and areas with more educated populations.

This work thus suggests a strong advantage for areas that were associated with the earliest activity in a technology. Despite the skill- and region-broadening effects alluded to above, not only does a disproportionate share of new employment continue to locate there, but especially the most desirable high-skilled positions.

A natural follow-up question relates to the mechanisms behind the continued advantage of the initial hubs. One possibility is that these patterns are driven by continued entrepreneurial activity. 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's (2015) work highlighting the extent to which cities' initial endowments affect the long-term distribution of entrepreneurship.

An alternative possibility, suggested by the evolution of the Boston biopharmaceutical sector, is that the persistence of nascent industries in their initial hubs is driven by firms based elsewhere moving into the initial hub, a phenomenon that we term "rehoming." 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 hope to better distinguish between these mechanisms in the next version of this paper, as well as to understand the characteristics of the firms that engage in rehoming.

This study is related to the substantial 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, the 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; Moscona, 2019). 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.

Another related strand of literature looks specifically at programs that finance entrepreneurial firms and the research from which they spring. These papers typically look at the design of a single program, and exploit a discontinuity in program design to understand its consequences (e.g., Bronzini and Iachini, 2014; Criscuolo et al., 2017; Howell, 2017). While the cross-industry nature of our analysis implies that we must forego the sharp identification that characterizes many of these papers, we believe it can nonetheless yield valuable insights.

The plan of this paper is as follows. Section 2 describes the construction of the data set. Section 3 presents a first look at overall changes. We examine the differential patterns across geographic regions in Section 4. Section 5 looks at the role of academic hubs. The final section concludes the paper.

2. Data Construction

In this paper, we identify a set of recent and influential 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 exposure at the job-firm-location-quarter level, and (d) aggregate in various ways to measure technology adoption at the region and firm level. This section describes our approach in more detail.

A. List of new technologies and associated keywords

We initially want to identify influential technologies in as systematic manner as possible. We start 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 (the enablement requirement refers to the requirement of 35 U.S.C. 112(a)). We focus solely on patent awards by the U.S. Patent and Trademark Office (USPTO): because of the importance of the U.S. market, inventors worldwide will file important discoveries with the USPTO.

More particularly, we identify patents with a primary or supplementary classification in the thirty three-digit technology classes with the most ultimately granted patents applied for between 1976 and 2016.[5] These thirty (out of 128 three –digit classes) account for the vast majority of activity, with roughly 88% of all patents filed here. For each class and application year, we then identified the ten most cited patents applied for in each year between 1976 and 2016. This approach yielded a total of 89,218 patents for analysis.

[5] In this analysis, we employ the Combined Patent Classification (CPC) scheme, which is modeled after the International Patent Classification scheme and is far more conducive to economic analysis than its predecessor U.S.-specific scheme (Lerner, 1995). While the U.S. only switched to employing this scheme in 2013, it has retroactively classified earlier awards using this scheme.

We then study the language in these patents. From the text, we extract the universe of two-word combinations ("bigrams"), numbering about 17 million in total. We focus on bigrams because they were less ambiguous than single-word keywords: while words like "autopilot" or "cloud" could have a variety of colloquial meanings, "autonomous vehicle" and "cloud computing" were much less ambiguous. To undertake this processing, we follow the methodology in papers undertaking textual analyses of patents and earnings calls, such as Kelly et al. (2018) and Hassan et al. (2019).

At the same time, many of the bigrams collected from the patents are not relevant for our purposes. One issue was that many of the bigrams, while related to a given 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 the "non-technical" bigrams. To this end, we construct a library of "non-technical" bigrams using corpora collected by linguists. In particular, the Corpus of Historical American English (COHA) is a decade-by-decade representative sample of text constructed by linguists from prominent fiction and non-fiction sources. Similar to the patent text processing, we decompose this text into bigrams and construct a decade-by-decade library of "non-technical" bigrams (for instance, "of the," "equipment used"). For patents applied for in a given year, we remove the bigrams that are included in the "non-technical" library in the previous decade. For example, for a patent applied for in 2007, we remove bigrams that were mentioned in the non-technical library from 1990 to 2000. Figure A-1 illustrates the methodology to obtain technical bigrams from a patent. Finally, in the remaining list of bigrams, we keep the most frequently appearing ones for each technology class and year. After these eliminations, we have a list of 15,716 "technical" bigrams in the most cited patents from 1976 to 2016.

In the second step, we focus on whether the bigrams figured into business discussion. Here we use earnings conference calls from publicly listed firms. We explain this in more detail in the section below.

B. Earnings calls data

Earnings call transcripts consist of two sections: a presentation by management (typically chief executive or financial officers) and then questions posed by investment analysts and the answers provided by the executives.

We tabulate the bigrams in 321,373 conference calls held by 11,905 publicly held companies and compiled by Thomson Reuters between 2002 and 2020. Through this examination, we eliminate about 60% of the bigrams from the patents that do not ever figure in these calls.

The final part of the data construction process focuses on the 500 technical bigrams most commonly encountered in the earnings calls. From this list, we use manual discretion to select those bigrams that appeared to clearly and unambiguously reflect specific technological advances that, in our reading, changed the way businesses operate. We then group those bigrams into 20 technologies, which form the basis of our analysis. Table 1 shows this grouping for all 50 bigrams, including, for example, "rapid prototyping," "additive manufacturing," "solar cell," and "solar module," where we group the former two bigrams to form a technology we refer to as 3D Printing, and the latter two to form the technology Solar Power.

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 trained on the set of earnings calls. This algorithm provides us a set of bigrams used in similar context to a given bigram. An embedding vector analysis employs a natural language processing algorithm to reduce a large corpus to a lower dimensional space using the contexts (neighboring words) that the words are used in, and as a byproduct provides a measure of the similarity (or the distance) between words in the corpus. We use the version of the algorithm that is an implementation of Mikolov et al. (2013). For each bigram in a given technology grouping, the algorithm suggests a list of "proximate" other bigrams. For example, the most proximate bigrams to "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 firm-level exposure to a given technology. To this end, we performed human audits where a team member went through 100 randomly sampled excerpts of the text from earnings calls 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. Finally, we incorporated the learnings from the human audit process and create a list of bigrams associated with 20 technologies.

Table 2 provides the leading bigrams associated with a subset of technologies in the sample; Table A-1 provides a complete list of bigrams associated with each of the sample technologies.

C. Burning Glass Job Postings

As described in Hershbein and Kahn (2018), Burning Glass (BG) aggregates online job postings using "spider bots" from online job boards, employer websites, etc. into a machine readable, de-duplicated database. From Burning Glass, we receive two datasets:

- The first is a standardized dataset where each de-duplicated job posting is geo-coded and assigned to a Standard Occupational Classification (SOC), a United States government system of classifying occupations.
- The second has raw unprocessed text of the job postings, which we use to assign exposure to our technology. Samples of the raw text in job postings with technology mentions are provided in Figure 1 and Figure 2.

We receive these data from BG for 2007 and 2010-2020. We use the geo-coded dataset to assign job postings to a core-based statistical area (CBSA), using the scheme based on the 2010 Census. This dataset covers 917 CBSAs and 837 SOC codes.

D. 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:

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

where $D_{i,t}$ is the set of bigrams contained in a job posting/earnings call/patent held or posted at time t and b_ζ is a bigram associated with a technology ζ. Essentially, a document is classified as exposed to a technology if it contains a bigram associated with the technology. Figure 1 and Figure 2 provide illustrative Burning Glass job postings exposed to AI technology and solar technology, respectively. Table A-2 provides excerpts from earnings calls exposed to our technologies.

We then aggregate over various documents D (job postings, earnings calls, and patents) to construct measures at occupation, firm, and geographic levels:

$$share\ exposed_{a,\zeta,t} = \frac{\sum_{i\,\epsilon\, a,t} 1\{b_\zeta \epsilon\, D_{i,t}\}}{\sum_{i\,\epsilon\, a,t} 1\{\, D_{i,t}\}}$$

where a may be a firm, sector, urban region or occupation, and t is time.

Table 3a shows the top exposed occupations (using job postings) and in table 3b the top exposed firms (using earnings calls) for one sample technology.

3. Key Employment Patterns

We first seek to understand the overall patterns in the diffusion of these 20 technologies. The analysis suggests that growth in job postings referring to given technologies grows in tandem with references in earning calls. The general effect over time is for hiring to move from a sharp focus on high-skilled jobs to a much broader intake of workers.

Figure 3 takes a first look at the adoption technologies. The 20 images plot two measures of activity on an annual basis for each technology. In each case, we depict the activity from shortly before the start date through 2019.

First, the red line denotes the share of 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. In others, such as 3-D printing and artificial intelligence, there is a steady climb over time.

In each case, we begin about the time that the technology became significant, which we henceforth refer to as the "start date." To compute this, for each of our technologies, we first take the maximum of the "share of earnings calls" time series in Figure 3. We take the start date to be the year in which the time series first attains at least 10% of this value. Table A-3 in the Appendix lists the start date for each technology.[6]

The second series, denoted with gray dots, indicates the share of positions in Burning Glass exposed to a given technology (as defined above). 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 technologies), in general the two series are quite closely correlated. The correlation coefficient between them across the figures is 0.68. The close tie between these series helps validate the reasonableness of our empirical methodology.

We next turn to examining the skill component of technology job postings over time. Figure 4 plots the same red line as in the previous table (the share of earnings calls exposed to a given technology) against a measure of skill of these job postings (the blue circles). We compute for each 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. We present here the results with the share of employees in the position who are college educated. 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:

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

where o is a Census SOC code, $N^{\tau}_{o;t}$ 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.

The figure suggests that in general 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. (LCD screens is one of the few exceptions.) These results are consistent with the view that new technologies start with high-skill occupations and then pull in larger parts of the workforce over time.[7]

We summarize this information by presenting a binned scatterplot in Figure 5. This depiction shows the relationship across all 20 technologies as time elapsed after the "start date" and the mean share of the workforce that is college educated. (We control as well for technology fixed effects.) The strong linear trend--the declining demand for a college-trained workforce--is apparent: the fitted line has an R^2 in a simple linear regression of 81.9%.

[6] We are currently assigning technologies that were above the 10% threshold in 2002 as starting in 2000. We will refine this methodology in the next version of the paper.

[7] In Figures A-2,A-3 and A-4 in the Appendix, we present similar measures, using instead wage income, average years of schooling and share of post graduate persons. The results are quite similar.

Table 4 looks at this relationship in a simple ordinary least squares (OLS) regression analysis. Here, we use five alternative measures of the skills required in the positions associated with a given technology. The sample consists of annual observations of each technology between 2007 and 2019. The independent variables include the share of the weighted SOC classes that are college educated (as of 2015), the share with Ph.Ds., the share with graduate degrees, and mean wages, and the mean years of schooling. Each regression uses as the key independent variable the years since the start date, with controls for the technology and the calendar year of the observation. 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, in Table 4 Column 1 and Column 4 show that each additional year since start of the technology is associated with a fall of about 1.2% in share of college educated people and $1200 in wages for the job postings advertised for the technology.

This skill broadening effect sheds an interesting light on how high-skilled labor is complementary to low-skilled work. While there is an important body of work highlighting the way in technological change has favored high-skilled occupations and contributed to wage inequality (Acemoglu, 2002; Goldin and Katz, 2010, Acemoglu and Autor, 2011 are examples), the way in which the hiring associated with new technologies can transition over time does not appear to have been previously documented.

4. The Geographic Evolution of Nascent Technologies

In the second part of the analysis, we examine how the evolution of job creation played out across geographies. We show that hiring in new technologies became less concentrated over time. But at the same time as we see this "region broadening," the initial geographic hub retains its advantage.

Figure 6 examines how the variation across regions (we use Core Based Statistical Areas – a division of the United States into metro and micropolitan areas) changes over time. To examine this, we create a measure entitled the Normalized Share of Technology jobs for every CBSA, technology τ, and year t:

$$Normalized\ share_{cbsa,\tau,t} = \frac{share\ jobs\ expsed_{cbsa,\tau,t}}{share\ jobs\ expsed_{\tau,t}},$$

which reflects 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 fact that, for instance, Los Angeles will have a large share of job posting of nearly every type and that different technologies may be implemented at very different scales at a given point in time.

The figure depicts, for each technology, the ratio of the standard deviation and the mean of this measure across urban areas by year, or the coefficient of variation. The analysis reveals a sharp pattern. Across technologies, there is a sharp decline in the coefficient of variation. Put another way, the distribution of activity over time become more heterogeneous.

Again, this decline can be illustrated in a binned scatterplot. Figure 7 shows the relationship across all 20 technologies between the time elapsed since the "start date" and the mean share of the workforce that is college educated. (Once again, we add technology fixed effects.) The figure reveals a decline in the coefficient of variation over time. The fitted line has an R^2 in a simple linear regression of 34.6%.

This point can be underscored in a regression analysis. Table 5 examines the change in the concentration of job postings associated with new technologies using, in addition to the coefficient of variation, several measures of concentration: the mean squared deviation from one, the mean absolute deviation from the mean, the mean, the standard deviation, and the share of the top one percent of urban areas. Again, we regress these measures on the time since the technology's start date, as well as year and technology dummies. For all measures of deviation there is a strong negative relationship. Column 6 of Table 5 suggests that an additional year since technology start is associated with a 6.2% reduction in dispersion of technology relative to the mean.

We next used the SOC codes to divide our sample of job postings into three categories. We undertake these divisions using the share of college educated in the SOC codes associated with the technologies. Again, we use information from the 2015 ACS to determine the training of individuals in various SOCs. We termed them high skilled (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 start date, and how these shifts differ across different skill levels.

Figure 8 takes a first look at these patterns. It again is a binned scatterplot with technology-specific controls. It shows that the decline in the coefficient of variation across regions is substantially steeper for low-skilled jobs than that for medium- and especially high-skilled ones. While the low-end jobs rapidly disperse across the country, the higher-end ones remain more bunched together.

Table 6 takes another look at these patterns, emulating the structure of Table 5, but now breaking the observations of technologies into the three skill buckets. Across the five measures of dispersion reported, the duration since the start date has a consistently negative coefficient for the low-skilled group. By the time we reach the high-skilled group, only one coefficient is significantly negative. Many of the others have positive (if statistically insignificant) signs. Consistent with Figure 8, it appears that the decline in the concentration of job postings is much sharper in positions for lower-skilled workers.

A follow-on analysis, Table 7, examines pooled regressions across the technologies. The unit of observation is at the technology-year level, with the measures of dispersion used in Table 6 as the dependent variables. Again controlling for year and technology fixed effects, we include as before the time since the start date as an independent variable; but now we also include an indicator indicating if the observation is for a low-skilled occupation, and an interaction between this dummy and the duration measure. In all the regressions for measures of dispersion, the interaction term is significantly negative. The analysis suggests that jobs associated with new technologies in general become less concentrated over time, but that the decline of concentration is indeed faster in low-skilled occupations. For example, if we look in column (4), the coefficient on time is -0.093 suggesting that the standard-deviation (our measure of dispersion) of technology jobs falls by 4.6%

each year (given the sample mean of 2.74). The coefficient for the interaction with low-skill is -0.091, so that for low skilled jobs this dispersion falls at a rate of -0.184 (adding these coefficients) almost double the rate for the excluded category of medium and high skilled jobs.

The last two tables in this section look at the normalized number of technology job postings over time as the dependent variable. Table 8 presents an analysis at the technology-year-CBSA level.

The first two columns look at the persistent advantage of technology hubs. We characterize the hubs in two manners. The first panel uses as the key independent variable the number of job postings associated with a given technology in a CBSA at the start date; the second a dummy designating whether the region was a technology hub. (We define the hubs for a given technology as the CBSAs in the starting year or 2007—whichever is later—that collectively accounted for 80% of the job postings in the nascent technology.[8]) Each regression controls for the region, the year, the duration since the start date, and the technology.

In general, we see a strong pattern of continuity in where the new jobs locate. The number of jobs posted in a given technology and urban area in the starting year has strong predictive power for the location of jobs posted in future years. While there is a modest dissipation of this effect in the regressions, it would take many decades—in some specifications, more than a century—for the initial advantage to dissipate based on the predicted decay rates.

The last two columns of Table 8 look at the impact of being proximate to a technology hub. Rather than looking at the initial employment in the CBSA itself, we look at that in nearby CBSAs. (We exclude hubs themselves from the analysis.) In particular, we compute the inverse distance weighted average for each non-hub CBSA. This measure is defined as:

$$IDW(tech\ hub)_{c,\varsigma} = \frac{\Sigma_c \, {}^{1}\!/_{dist_{c,c'}} \, 1\{tech\ hub\}_{c'}}{\Sigma_c \, {}^{1}\!/_{dist_{c,c'}}},$$

where $dist_{c,c'}$ is the distance between CBSA c and CBSA c', and $1\{tech\ hub\}_{c'}$ is an indicator for a technology hub. Each regression again controls for the urban region, the year, the duration since the start date, and the technology.

The results suggest a strong positive impact of proximity. Those urban areas that are near technology hubs also see far higher levels of technology jobs, consistent with the strong persistence of hiring in hubs documented above. The interaction with time in column (4) is negative but not significant, suggesting little (if any) diminution in the benefits of being located near a technology hub.

In Table 9, we repeat the analysis in Panel B in Table 8 (that is, using the technology hub designation) but further dividing the job postings into low-, medium-, and high-skilled jobs.

[8] All regressions take inverse hyperbolic sines of the key independent variables, which is a generalization of the logarithmic transformation. The one exception is the first two regressions in Panel B, where the key independent variables are 0/1 dummies.

Consistent with the results in Table 7, we see that there is a strong differential in the concentration of jobs. There is much greater concentration over time in high-skilled jobs. In general, the decay effects over time are modest, as in Table 8, with lower skilled jobs seeing faster decay than medium skilled jobs and particularly highly skilled jobs.

5. Academia and Technology Hubs

Up until this point, we have ignored the characteristics of the original locations where the technologies did the bulk of their hiring at the time of the start date. In the final section, we explore their features. As we saw in Section 4, these initial location decisions have enormous implications for the subsequent locations of jobs, particularly high-skilled ones. In particular, we highlight that there is a strong relationship between academic centers and where nascent industries begin.

Table 10 takes a first look at this question, looking at the job postings in the starting year in each urban area across technologies. We take the start date to be 2007 for technologies that started before the coverage of the BG data. We use as a dependent variable the normalized share of jobs in a given technology, defined as before and winsorized at the 1% level. The key independent variables are the number of research universities in the urban area, the volume of university R&D spending in that year, and the share of the population that is college educated or has a post-graduate degree. The number of research universities is calculated using the number of institutions which grant a bachelor's or higher degree and had at least $150,000 allocated separately for R&D expenditures during 2013 fiscal year. We determine this information from the higher education R&D survey and ACS 2015. Finally, this analysis controls for technology specific fixed effects.

The analysis suggests a strong cross-sectional pattern. Urban regions with a greater academic presence—whether manifested by greater research university presence or a more educated workforce—were more likely to enjoy a surge of initial hiring as a new industry emerges.

Figure 9 looks at this pattern graphically. The strong positive relationship between the normalized share of jobs for each technology and the number of research universities is evident.

We can estimate a similar regression, now looking at specific technologies. Table 11 presents the results of this analysis when use the number of research universities in an urban area and their R&D spending as dependent variables. The t-statistics on the academic variables are almost entirely positive and statistically significant, suggesting that this pattern is not driven by any single technology or a handful of ones. The other academic-related variables produce similar results.

We undertake a similar analysis, now pooling all the observations of the different technologies. We then can estimate the overall sensitivity of the number of job postings in a given urban area, technology, and year to the presence of research universities. Figure 10 presents the results of such an analysis.

Again, we see a strong positive relationship between the share of job postings in each technology, year, and region and the amount of academic activity. For instance, in the upper-left panel, which examines the number of research universities, as we move from one to three universities (from zero to a little more than one when the logarithm is taken), the predicted coefficient moves from about 0.35 to about twice this level.

6. Conclusion

Policymakers have devoted enormous energy in recent years to seeking to develop capabilities in nascent technologies. Such an infant industry strategy is predicated on the notion that early advantages will yield lasting benefits for regions, particularly in the form of high-quality employment.

This analysis suggests a nuanced view. On the one hand, the skill-broadening and region-broadening effects documented here suggest that it is naïve to equate new technologies exclusively with high-skilled jobs. Similarly, employment from new technologies diffuses over time geographically. The complementarities over time between high- and low-skilled positions associated with new technologies are also intriguing.

On the other hand, the disproportionate share of the employment activity in the initial hub appears to be very persistent. This effect is most pronounced among the high-skilled positions, which may be particularly desirable. The strong association of academia with the location of the initial hubs is also striking.

The analysis leaves several questions open, which we hope to explore in the next draft of the paper. The first relates to the drivers of the persistence hiring activity in initial technology hubs. To what extent is it driven by hiring by the initial entrants or the start-ups that they spawn? Or is it instead largely a consequence of rehoming by incumbent firms based elsewhere? A related question is what types of firms choose to rehome. Is it the firms that were particularly prescient in identifying the new technology? Those who saw it as it as an especial competitive threat? Answers to these questions will help us better explain these fascinating phenomena.

References:

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

—. "What does human capital do? A review of Goldin and Katz's The race between education and technology." *Journal of Economic Literature* (2012): 426-63.

Acemoglu, Daron, and Joshua Linn. "Market size in innovation: theory and evidence from the pharmaceutical industry." *The Quarterly journal of economics* (2004): 1049-1090.

Autor, David H and Katz, Lawrence F and Kearney, Melissa S. "Trends in US wage inequality: Revising the revisionists." *The Review of economics and statistics* (2008): 300--323. (MIT Press).

Bloom, Nicholas and Van Reenen, John and Williams, Heidi. "A toolkit of policies to promote innovation." *Journal of Economic Perspectives* (2019): 163--84.

Bronzini, Raffaello, and Eleonora Iachini. "Are incentives for R&D effective? Evidence from a regression discontinuity approach." *American Economic Journal: Economic Policy* (2014): 100-134.

Criscuolo, Chiara, Ralf Martin, Henry Overman, and John Van Reenen. "The causal effects of an industrial policy." *American Economic Review, 109* (2019): 48-85.

Davies, Mark. "The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights." *International journal of corpus linguistics 14.2* (2009): 159-190.

Fagerberg, Jan, David C. Mowery, and Richard R. Nelson. *The Oxford handbook of innovation.* Oxford University Press, 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* (2015): 498-520.

Goldin, Claudia Dale, and Lawrence F. Katz. *The race between education and technology*. Harvard University Press, 2009.

Gompers, Paul, Josh Lerner, and David Scharfstein. "Entrepreneurial spawning: Public corporations and the genesis of new ventures, 1986 to 1999." *The journal of Finance 60.2* (2005): 577-614.

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

Hall, Bronwyn H. "Innovation and diffusion." *National Bureau of Economic Research* (2004): No. w10212.

Hassan, Tarek A., Stephan Hollander, Laurence van Lent, and Ahmed Tahoun. "Firm-level political risk: Measurement and effects." *The Quarterly Journal of Economics 134, no. 4* (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. No. w26971." *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 108.7* (2018): 1737-72.

Howell, Sabrina T. "Financing innovation: Evidence from R&D grants." *American Economic Review 107.4* (2017): 1136-64.

Katz, Lawrence F and Murphy, Kevin M. "Changes in relative wages, 1963--1987: supply and demand factors}." *The quarterly journal of economics* (1992): 35--78. (MIT Press).

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. "Measuring technological innovation over the long run. No. w25266." *National Bureau of Economic Research* (2018).

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." *National Bureau of Economic Research* (2019).

Moscona, Jacob. "Environmental Catastrophe and the Direction of Invention: Evidence from the American Dust Bowl." (2020).

Piketty, Thomas, and Emmanuel Saez. "Income inequality in the United States, 1913–1998." *The Quarterly journal of economics 118.1* (2003): 1-41.

Popp, David. "Induced innovation and energy prices." *American Economic Review 92.1* (2002): 160-180.

Snapshot, Massachussetts Biotechnology Council - Industry. *https://www.massbio.org/wp-content/uploads/2020/03/2019-INDUSTRY-SNAPSHOT-ELECTRONIC-FINAL_SEPT2019.pdf.* 2019.

Song, Jae and Price, David J and Guvenen, Fatih and Bloom, Nicholas and Von Wachter, Till. "Firming up inequality." *The Quarterly journal of economics* (2019): 1--50. (Oxford University Press).

Table 1 - Top selected bigrams from Patenting and EC process

Bigram	# EC	Technology group
flat panel	1864	lcds
fiber optic	1527	optical fiber
disk drive	1421	disk drive
monoclonal antibody	1296	monoclonal antibody
fiber network	1280	optical fiber
disk drives	1012	disk drive
mobile computing	965	cloud
virtual reality	928	virtual reality
solar cell	786	solar
panel display	770	lcds
monoclonal antibodies	749	monoclonal antibody
fiber optics	736	optical fiber
touch screen	720	touch screen
coalbed methane	555	fracking
smart card	555	smart card
solar panel	533	solar
hydraulic fracturing	516	fracking
hybrid electric	511	electric cars
optical fiber	497	optical fiber
additive manufacturing	463	3dprinting
fiber networks	441	optical fiber
machine vision	363	machine vision
cellular phone	355	smartphone
unmanned aerial	337	drone
solar modules	329	solar
global positioning	328	gps
solar module	317	solar
autonomous vehicle	310	driverless
hybrid vehicle	243	electric cars
rapid prototyping	228	3dprinting
oled display	224	oled
vehicle charging	211	electric cars
unmanned aircraft	196	drone
disc drive	194	disk drive
touch panel	171	touch screen
fiber cable	153	optical fiber

Notes: The table lists (in Column 1) top shortlisted bigrams and their technology groups (in Column 3) from the process detailed in section 3, by the number of earnings calls that the bigrams were mentioned in (in Column 2). We use these bigrams to select 20 prominent recent technologies.

Table 2 - Top Keywords for sample technologies by number of online job postings

electric cars	
keyword	jobs
electric vehicles	90348
electric vehicle	57993
hybrid electric	14670
electric drive	9960
electric cars	8313
electric car	6574
hybrid vehicle	6366
electric hybrid	5106
hybrid vehicles	3992
electrical vehicle	2861

AI	
keyword	jobs
machine learning	1212056
artificial intelligence	283341
predictive analytics	166832
deep learning	140793
language processing	120323
neural networks	48776

fracking	
keyword	jobs
hydraulic fracturing	10519
shale gas	9258
horizontal drilling	6303
fracking	6300
unconventional gas	5526
shale plays	3684
shale oil	2254
gas shale	1393
unconventional shale	1097
oil shale	911

3dprinting	
keyword	jobs
rapid prototyping	105885
d printing	68132
additive manufacturing	39123
d printers	17304
d printer	9021
d printed	4448
d print	3886
binder jetting	124

Notes: The table lists top bigrams (in Column 1 for each technology) by the number of online job postings that they are mentioned in (Column 2) 2007-19, for 4 out of 20sample technologies. For a full list of all bigrams associated with the set of 20 technologies, refer to Appendix Table 1.

Table 3a - Top Exposed Occupations to Virtual Reality

Standard Industry Occupation Name	Total Jobs	Exposed Jobs	% 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

Notes: The table above lists top exposed occupations (in column 1) by percentage of job postings (in Column 4) exposed to virtual reality technology. Table only shows occupations with at least 10,000 job postings in Burning Glass between 2007 and 2019.

Table 3b - Top Exposed Firms to Virtual Reality

Company	SIC	Tech Exposure (in pct.)	# ECs
Snap Inc	7370	100%	12
Gannett Co Inc	2711	59%	17
Facebook Inc	7370	52%	31
Dave & Buster's Entertainment Inc	5810	50%	20
eMagin Corp	3679	48%	64
Lumentum Holdings Inc	3576	47%	19
DIRTT Environmental Solutions	2522	41%	17
Adobe Inc	7372	40%	30
Shopify Inc	7370	37%	19
Wayfair Inc	5961	36%	22
GoPro Inc	3861	35%	23
Immersion Corp	7372	31%	65
InvenSense Inc	3674	30%	20
Shanda Interactive Entertainment	7370	30%	27
GSV Capital Corp	6797	28%	29
MicroVision Inc	3679	27%	71
Technicolor SA	7812	27%	15
Charter Communications Inc	4841	27%	15
AVEVA Group PLC	7372	27%	15

Notes: The table above lists top exposed firms (in column 1) by percentage of earnings calls (in column 3) mentioning virtual reality technology. Table only shows firms with at least 10 earnings calls in the Thomson Reuters database between 2002 and 2019.

Table 4 - Skill measure of technology job postings against year since start

	(1) share college educated * 100	(2) share phd * 100	(3) share post graduate * 100	(4) avg. wage	(5) avg. schooling
years since start	-1.243***	-0.0698***	-0.493***	-1,248***	-0.0654***
	(0.245)	(0.0167)	(0.0963)	(248.7)	(0.0130)
Observations	203	203	203	203	203
R-squared	0.833	0.971	0.892	0.835	0.854
Tech FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES

Notes: This table reports the results from a regression of approximate skill composition of technology jobs $Skill_t^{\varsigma} = \frac{\sum_o N_{o;t}^{\tau} \chi_{o;2015}}{\sum_o N_{o;t}^{\tau}}$ on the left hand side, where $\chi_{o;2015}$ is the skill measure of interest from ACS 2015 at the occupation level), against the years since start of the technology on the right hand side. The regression uses online Burning Glass job postings data between 2007 and 2019. Standard errors are clustered by technology.

Table 5 - Concentration during the technology life cycle

	(1) Mean sq. dev. from 1	(2) mean abs. dev. from mean	(3) Normalized share top 1%	(5) SD	(6) CV
years since start	-0.0505***	-0.0120***	-0.150***	-0.0245***	-0.0625**
	(0.0167)	(0.00415)	(0.0466)	(0.00503)	(0.0219)
Observations	203	203	203	203	203
R-squared	0.669	0.915	0.835	0.878	0.929
Tech FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES

Notes: This table reports the results from regressions of various measures of concentration during lifecycle of a technology regressed on years since inception of the technology. The dependent variables are aggregations over $Normalized\ share_{cbsa,\tau,t}$ for every technology and year, where $Normalized\ share_{cbsa,\tau,t} = \frac{share\ jobs\ expsed_{cbsa,\tau,t}}{share\ jobs\ expsed_{\tau,t}}$. Mean sq dev from 1 is calculated as the average over square deviations of normalized share of technology jobs for every CBSA from 1, Mean abs dev from mean is calculated as the average over absolute deviations of normalized share across CBSAs over its average, Mean Top 1% share is calculated as the average normalized share of tech jobs for top 1% of CBSAs, and similarly others. $Normalized\ share_{cbsa,\tau,t}$ is winsorized by technology and year. SD is standard deviation of normalized shares across CBSA and CV is the coefficient of variation of this. Standard errors for all specifications are clustered by technology.

Table 6 - Concentration during the life cycle - By Skill

Panel A: Low Skill	(1) Mean sq. deviation from 1	(2) Mean abs. dev. from mean	(3) Normalized share top 1%	(4) SD	(5) CV
years since start	-0.185***	-0.0320***	-0.520***	-0.0727***	-0.106***
	(0.0370)	(0.00551)	(0.0732)	(0.00701)	(0.0264)
Observations	202	202	202	202	200
R-squared	0.859	0.800	0.832	0.831	0.891
Panel B: Medium Skill	(1)	(2)	(3)	(5)	(6)
years since start	-0.124*	-0.0133	-0.268**	-0.0365**	-0.0924***
	(0.0674)	(0.0133)	(0.107)	(0.0178)	(0.0279)
Observations	203	203	203	203	202
R-squared	0.568	0.849	0.807	0.826	0.934
Panel C: High Skill	(1)	(2)	(3)	(5)	(6)
years since start	-0.0837**	-0.0107	-0.107	-0.0192*	-0.0532*
	(0.0373)	(0.00984)	(0.0777)	(0.0107)	(0.0274)
Observations	203	203	203	203	202
R-squared	0.733	0.908	0.866	0.890	0.945
Tech FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES

Notes: This table reports the results from regressions of various measures of concentration during lifecycle of a technology by skill against year since start of the technology. To calculate the different measures of concentration 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 low skill occupations (with share of college educated people < 30%). Finally, the measures of concentration are calculated over $Normalized\ share_{cbsa,\tau,t,skill}$ across CBSAs by skill group, technology and time, similarly as in table 5. Standard errors are robust.

Table 7 - Concentration during the life cycle - High vs Low skill

	(1)	(2)	(3)	(4)	(5)
	Mean sq. deviation from 1	Mean abs. dev. from mean	Normalized share top 1%	SD	CV
(year since t0) * (skill == low)	-0.0905***	-0.0269***	-0.388***	-0.0552***	-0.0527***
	(0.0310)	(0.00520)	(0.0677)	(0.00920)	(0.0134)
(year since t0)	-0.0932**	-0.0102	-0.134*	-0.0214**	-0.0568***
	(0.0421)	(0.00833)	(0.0682)	(0.0107)	(0.0210)
Observations	405	405	405	405	402
R-squared	0.539	0.817	0.756	0.786	0.879
Tech FE	YES	YES	YES	YES	YES
Skill FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES

Notes: This table tests for differential dispersion trends for high and low skill technology job postings by regressing various measures of concentration (as in table 5) against the interaction term (year since t0) * (skill == low). To calculate the different measures of concentration 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 low skill occupations (with share of college educated people < 30%). Finally, the measures of concentration are calculated over $Normalized\ share_{cbsa,\tau,t,skill}$ across CBSAs by skill group, technology and time. The regression only uses data for low and high skill occupations, excluding medium skill occupations. Standard errors are robust.

Table 8 - Differential hiring for Technology hubs

Panel A: W/ intensive margin	(1)	(2)	(3)	(4)
	Normalized Share			
IHS(tech jobs at t0)	0.314***	0.346***	0.295***	0.295***
	(0.0156)	(0.0163)	(0.0152)	(0.0152)
ihs(tech jobs at t0) * (year since technology start)		-0.00297***		
		(0.000846)		
IHS(IDW(tech jobs at t0))			0.317***	0.359***
			(0.0431)	(0.0492)
ihs(IDW(tech jobs at t0)) * (year since technology start)				-0.00353
				(0.00233)
Observations	140,537	140,537	140,537	140,537
R-squared	0.220	0.221	0.224	0.224

Panel B: W/ tech hub dummy	(1)	(2)	(3)	(4)
tech hub	0.565***	0.883***		
	(0.0340)	(0.0614)		
tech hub*(year since t0))		-0.0261***		
		(0.00419)		
ihs(IDW(tech hub))			0.216***	0.201***
			(0.0357)	(0.0452)
ihs(IDW(tech hub))*(year since t0))				0.00115
				(0.00224)
Observations	140,537	140,537	134,092	134,092
R-squared	0.171	0.173	0.112	0.112
Tech FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Year since FE	YES	YES	YES	YES
CBSA FE	YES	YES	YES	YES

Notes: The table shows results from regression of $Normalized\ share_{cbsa,\tau,t,skill}$ against the total number of jobs for the cbsa at the time of start of the technology (in Panel A) and technology hub status dummy of the CBSA (in Panel B). ihs(tech jobs at t0) denotes the inverse hyperbolic sine of the absolute number of technology jobs at time t for CBSA c. Top CBSAs which account for 80% of total technology job postings at t0 are given a technology hub status for a technology. ihs(IDW(tech jobs at t0))) denotes the inverse hyperbolic sine of the total inverse distance weighted (distance between CBSAs c and c') technology jobs for CBSA c. Similarly, ihs(IDW(tech hub)) is the inverse distance weighted technology hub measure for a CBSA c. These results exclude observations at the time of and before the start year of a technology, and exclude CBSAs with less than 1000 job postings. Panel B, Column (3) and (4) exclude observations for technology hubs. Standard errors for all specifications are robust.

Table 9 - Differential Hiring for Technology Hubs by Skill

	(1)	(2)	(3)	(4)
Panel A: Low Skill	Normalized Share			
tech hub	0.449***	0.954***		
	(0.0388)	(0.0765)		
tech hub*(year since t0))		-0.0405***		
		(0.00540)		
IHS(IDW(tech hub))			0.382***	0.595***
			(0.0614)	(0.0766)
IHS(IDW(tech hub))*(year since t0))				-0.0164***
				(0.00380)
Observations	169,099	169,099	169,099	169,099
R-squared	0.085	0.089	0.082	0.083
Panel B: Medium Skill	(1)	(2)	(3)	(4)
tech hub	0.558***	0.830***		
	(0.0370)	(0.0698)		
tech hub*(year since t0))		-0.0226***		
		(0.00480)		
IHS(IDW(tech hub))			0.498***	0.576***
			(0.0523)	(0.0617)
IHS(IDW(tech hub))*(year since t0))				-0.00608**
				(0.00286)
Observations	169,099	169,099	169,099	169,099
R-squared	0.140	0.142	0.134	0.134
Panel C: High Skill	(1)	(2)	(3)	(4)
tech hub	0.660***	0.839***		
	(0.0438)	(0.0677)		
tech hub*(year since t0))		-0.0151***		
		(0.00451)		
IHS(IDW(tech hub))			0.562***	0.602***
			(0.0629)	(0.0704)
IHS(IDW(tech hub))*(year since t0))				-0.00313
				(0.00334)
Observations	169,099	169,099	169,099	169,099
R-squared	0.187	0.188	0.178	0.178
Tech FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Year Since FE	YES	YES	YES	YES
CBSA FE	YES	YES	YES	YES

Notes: This table reports the results from regressions of $Normalized\ share_{cbsa,\tau,t,skill}$ on technology hub status dummy of the CBSA and the inverse distance weighted technology hub measure for the CBSA c, separately for low skill (panel A), medium skill (panel B) and high skill (panel C). To calculate the different measures of concentration 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 low skill occupations (with share of college educated people < 30%). Finally, the measures of concentration are calculated over $Normalized\ share_{cbsa,\tau,t,skill}$across CBSAs by skill group, technology and time. Standard errors are robust. These results exclude observations at the time of and before the start year of a technology, and exclude CBSAs with less than 1000 job postings. Standard errors for all specifications are robust.

Table 10 - Technology hiring at the beginning year by skilled composition

	(1)	(2)	(3)	(4)
log(1 + # research unis)	0.284***			
	(0.0279)			
log(1+ R and D uni spend)		0.0314***		
		(0.00298)		
Share College Educated			2.852***	
			(0.260)	
Share post graduate				6.496***
				(0.649)
Observations	18,531	18,531	18,340	18,340
R-squared	0.049	0.050	0.055	0.055
Tech FE	YES	YES	YES	YES

Notes: The table presents results from a regression of $Normalized\ share_{cbsa,\tau,t}$ at time t0 on various measures of skill level of the CBSA. For some of our technologies, we do not have BG coverage throughout their lifetime. For these technologies, we take the starting year to be 2007, the starting year for our BG sample. Standard errors for all specifications are clustered by CBSA. Standard errors for all specifications are robust

Table 11 - Regression of Normalized Share on uni variables by technology

technology	start year	tstat log(1+research uni)	tstat - log(1 + research uni spend)
lcds	2000	5.06	4.69
rfid	2003	6.84	5.84
electric cars	2005	2.72	2.88
smartcard	2000	6.96	6.12
opticalfiber	2000	-0.89	-0.23
3dprinting	2005	4.41	3.96
smartphone	2006	6.53	6.29
bluetooth	2000	5.57	4.68
driverless	2014	4.02	4.21
oled	2002	3.18	3.09
monoclonal antibody	2000	4.58	4.81
fracking	2003	1.71	1.75
machine vision	2000	2.54	2.37
virtual reality	2015	5.93	6.15
solar	2003	3.34	3.5
gps	2000	1.04	1.63
ai	2014	10.88	9.83
diskdrive	2000	5.5	5.2
touchscreen	2002	1.83	1.79
cloud	2008	9.13	8.52

Notes: The table presents t-stats of regressions of $Normalized\ share_{cbsa,\tau,t}$ on university presence (in column 3) and research budget of universities (in column 4) in the CBSA, technology by technology (in column 1). T-stats are calculated using robust standard errors.

Appendix Tables

Appendix Table 1 - List of Keywords by Technology

technology	keywords
3dprinting	3d printing,3d printed,3d printer,3d printers,3d print, additive manufacturing, inkjet bioprinting, binder jetting, rapid prototyping
ai	artificial intelligence, machine learning, neural networks, deep learning, predictive analytics, language processing
bluetooth	bluetooth
cloud	cloud computing, cloud technology, cloud resources, cloud storage, cloud applications, cloud services, public cloud, cloud business, private cloud, cloud service, hybrid cloud, cloud platform, cloud infrastructure, cloud solutions, cloud providers, cloud offering, cloud revenue, cloud offerings, cloud solution, cloud based, service cloud, cloud customers, cloud data, enterprise cloud, cloud environment, mobile computing, cloud security, cloud environments, cloud deployments, cloud application, cloud platform
disk drive	disk drive, disc drive, hard drive, hard disk, hard disc
driverless	autonomous car, autonomous cars, selfdriving car, selfdriving cars, selfdriving truck, selfdriving trucks, autonomous trucks, autonomous vehicles, selfdriving vehicles, autonomous driving, autonomous vehicle, automated driving, driverless car, driverless cars, driverless vehicle, driverless vehicles
electric cars	electric car, electric vehicle, electric cars, electric vehicles, electric hybrid, hybrid car, hybrid bus, electric bus, hybrid vehicle, hybrid electric, hybrid vehicles, electric drive, plugin hybrid, hybrid cars, electric buses, vehicle charging, electrical vehicles, electrical vehicle
fracking	horizontal drilling, coalbed methane, unconventional gas, hydraulic fracturing, tight sandstones, shale rock, shale gas, shale oil, unconventional shale, oil shale, gas sands, shale plays, gas shale, hydraulic fracturing, fracturing activity, fracking
gps	global positioning, gps
lcds	flat panel, lcd, panel display, panel tv, flatpanel
machine vision	machine vision
monoclonal antibody	monoclonal antibody, monoclonal antibodies
oled	oled
optical fiber	optical fiber, optical fibre, fiber optic, fibre optic, fiber optics, fibre optics, optical transport, optical networks, fiber access, fiber channel, fiber cable
rfid	rfid
smartcard	smart card, smart cards
smartphone	smart phone, smart phones, samsung galaxy, android phone, android phones, smartphones tablets, smartphone, smartphones, iphone
solar	solar power, solar energy, solar business, solar cell, solar projects, solar market, solar capital, solar industry, solar cells, solar panels, solar project, solar panel, solar pv, solar module, solar modules, solar side, solar technology, solar farm, solar products, solar assets, solar systems, solar industries, solar generation, solar pwr, solar markets, solar wafer, solar plant, solar grade, solar capacity, solar facility, solar farms, solar companies, solar plants, solar installations, solar equipment, solar product, solar technologies, solar thermal, solar customers, solar segment, solar applications, solar installation, solar materials, solar facilities, solar sales
touch screen	touch screen, touch screens, touchscreen, touchscreens
virtual reality	virtual reality, augmented reality, mixed reality

Notes: The table provides list of all keywords (in Column 2) by technology (in Column 1) for each of the 20 technologies in our sample. This list is first constructed using most cited patents from 1976-2016, then extrapolated using embedding vectors algorithm trained on Earnings calls, and finally, corrected using human audit checks performed individually for each keyword.

Appendix Table 2 - Selected Earnings Calls Excerpts of Top Exposed Firms

Technology	Company Name (Tech Exposure)	Month	Excerpt
RFID	Progressive Gaming International Corp(0.81)	May-2005	on that and we today are tracking approximately million per day of bets using **RFID** technology so i think the speculation that is **RFID** a viable technology to track bets
Optical fiber	Optical Communication Products Inc(1.0)	Nov-2003	and optical transponders with indiscernible ranging from oc oc oc to indiscernible ethernet and **FIBER CHANNEL** oc or gbps and ocy gbps indiscernible ranging from several hundred meters to over
LCD	Samsung Electronics Co Ltd(1.0)	Jan -2009	with tvs the tv divisions third quarter market share was number one in the **FLAT PANEL** tvs and **LCD** tv segment in terms of both shipment and revenue in addition
Bluetooth	Dialog Semiconductor PLC(1.0)	Nov-2019	in new product development in iot we continue to increase our footprint with our **BLUETOOTH** low energy products which delivered year-on-year revenue growth in q consumer appetite for a growing
AI	Cloudera Inc(1.0)	Jun-2017	build scale and deploy **MACHINE LEARNING** solutions using the most popular programming languages and **DEEP LEARNING** frameworks our data science workbench delivers to the market the ip that we acquired
3dprinting	Voxeljet AG(1.0)	Nov-2019	our technology are flying in space and we are implementing with vjet an **ADDITIVE MANUFACTURING** solution for true mass production in our indirect parts segment
Cloud	Basware Oyj(1.0)	Jan-2019	of sales comes primarily or the costs in cogs is primarily related to the **CLOUD BUSINESS** cloud grew kind of doubledigit while then cost of sales in q came down
SmartCard	Next Biometrics Group ASA(1.0)	Nov-2018	we see is a new wave of biometric applications that are emerging in the **SMART CARD** and government identification space these markets require a largesized sensor so that you can

Notes: The table reports excerpts with 15 words each side of a technology keyword (in column 4) from earnings calls of top exposed firms (in column 2) for a handful of technologies (in column 1). The sample of earnings calls is between 2002 and 2019. Exposure for each of the firms is calculated as the percentage of earnings calls exposed to the technology and reported for each firm in brackets in column 2.

Appendix Table 3 - Start Year by Technology

tech	start year
3dprinting	2005
ai	2014
bluetooth	2000
cloud	2008
disk drive	2000
driverless	2014
electric cars	2005
fracking	2003
GPS	2000
LCDs	2000
machine vision	2000
monoclonal antibody	2000
OLED	2002
optical fiber	2000
RFID	2003
smart card	2000
smart phone	2006
solar	2003
touch screen	2002
virtual reality	2015

Notes: The table reports start year (in Column 2) for each of the 20 technologies (in Column 1) in our sample. The start year is calculated as the year in which the share of firms (in Earnings calls) first attains 10% of the maximum between 2002 and 2019. For technologies, which start out at 10% of the sample max in 2002, we denote their start years to be 2000.

Appendix Table 4 – Summary Statistics

Panel A: Occupation Level	N	Mean	SD	P10	P25	P50	P75	P90
technology jobs	174,140	41.19	852.93	0.00	0.00	0.00	2.00	21.00
wage	21,850	41,298.11	27,429.40	16,859.27	23,693.96	33,643.25	48,301.86	74,628.82
share phd	21,850	0.02	0.07	0.00	0.00	0.00	0.01	0.05
share post graduate	21,850	0.14	0.23	0.01	0.01	0.03	0.16	0.47
share college educated	21,850	0.33	0.31	0.04	0.07	0.18	0.54	0.84
avg. years of schooling	21,850	13.85	2.02	11.84	12.29	13.40	14.87	16.57
Panel B: CBSA Level								
normalized share of technology jobs	188,468	0.49	1.44	0	0	0	0.50	1.37
technology exposed jobs at t0	18,531	6.03	77.42	0	0	0	0	4
# of research unis in the CBSA	18,531	0.66	2.24	0	0	0	1	2
research uni expenditure	18,531	71,542.28	28,8430.10	0	0	0	2,860	162,658
PanelC: CBSA - Skill Level								
Normalized share (low skill)	203,195	0.49	1.95	0	0	0	0.08	1.30
Normalized share (medium skill)	203,960	0.45	1.78	0	0	0	0	1.23
Normalized share (high skill)	203,960	0.37	1.44	0	0	0	0	0.92

Notes: The table provides summary statistics for datasets at the occupation level (Panel A), CBSA level (Panel B), CBSA-skill level (Panel C) aggregated year by year and pooled for each of our 20 technologies. In Panel A, technology hiring data at the occupation level is aggregated over 6-digit SOC occupation codes provided by BG between 2007 and 2019. Wage, share of phds and other skill data is collected from the American Communities survey 2015 at the occupation level. In Panel B, technology hiring data is aggregated over CBSA using BG geo-code job postings. $Normalized\ share_{cbsa,\tau,t} = \frac{share\ jobs\ expsed_{cbsa,\tau,t}}{share\ jobs\ expsed_{\tau,t}}$ Where $share\ jobs\ expsed_{cbsa,\tau,t}$ is the percentage of job postings geo-coded to a CBSA and exposed to a technology. Number of research universities and their research expenditure is obtained from Higher Education Research and Development Survey and aggregated over CBSAs. In Panel C, we repeat the same process as in Panel B but separately for low, medium and high skill occupations. Normalized share of technology jobs is winsorized by technology and year.

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

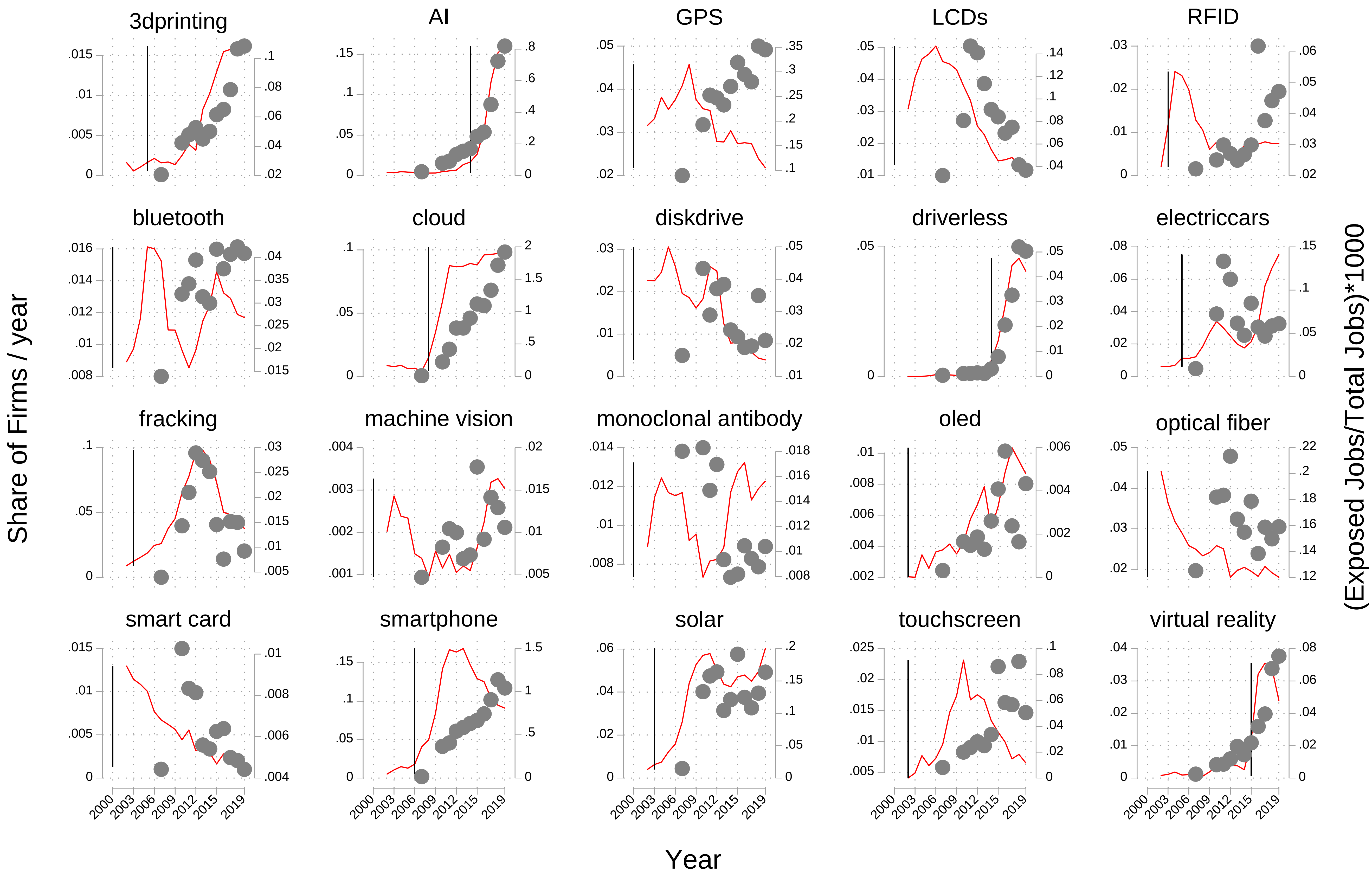
20190315

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

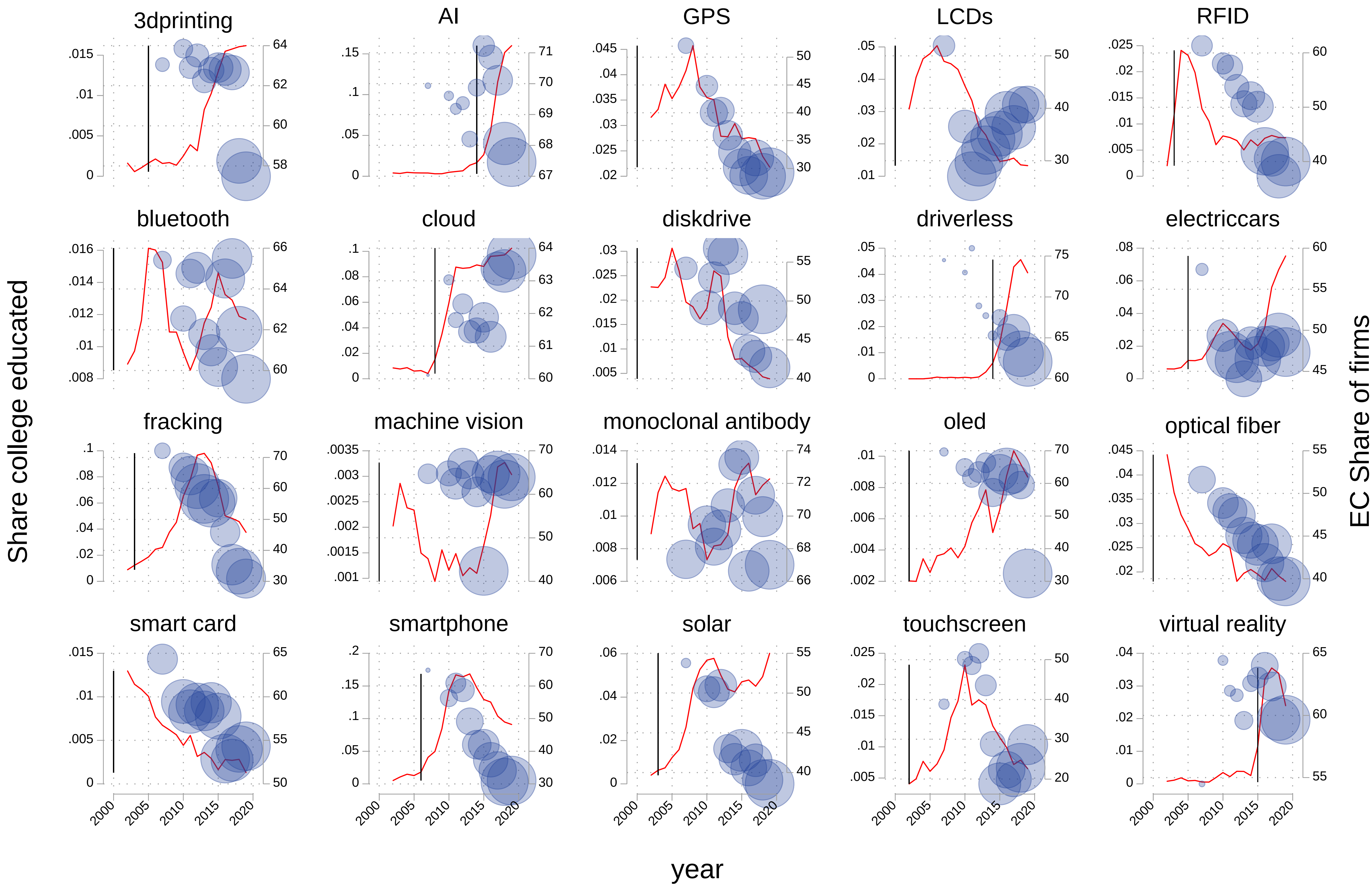
Notes: The picture is a sample job posting, which mentions solar technology related keywords, with a standardized job title, processed by Burning Glass, and the text of the job advertisement posted online on glassdoor.com.

Figure 3 – Technology Exposure in Earnings Calls and Jobs Postings – Time Series



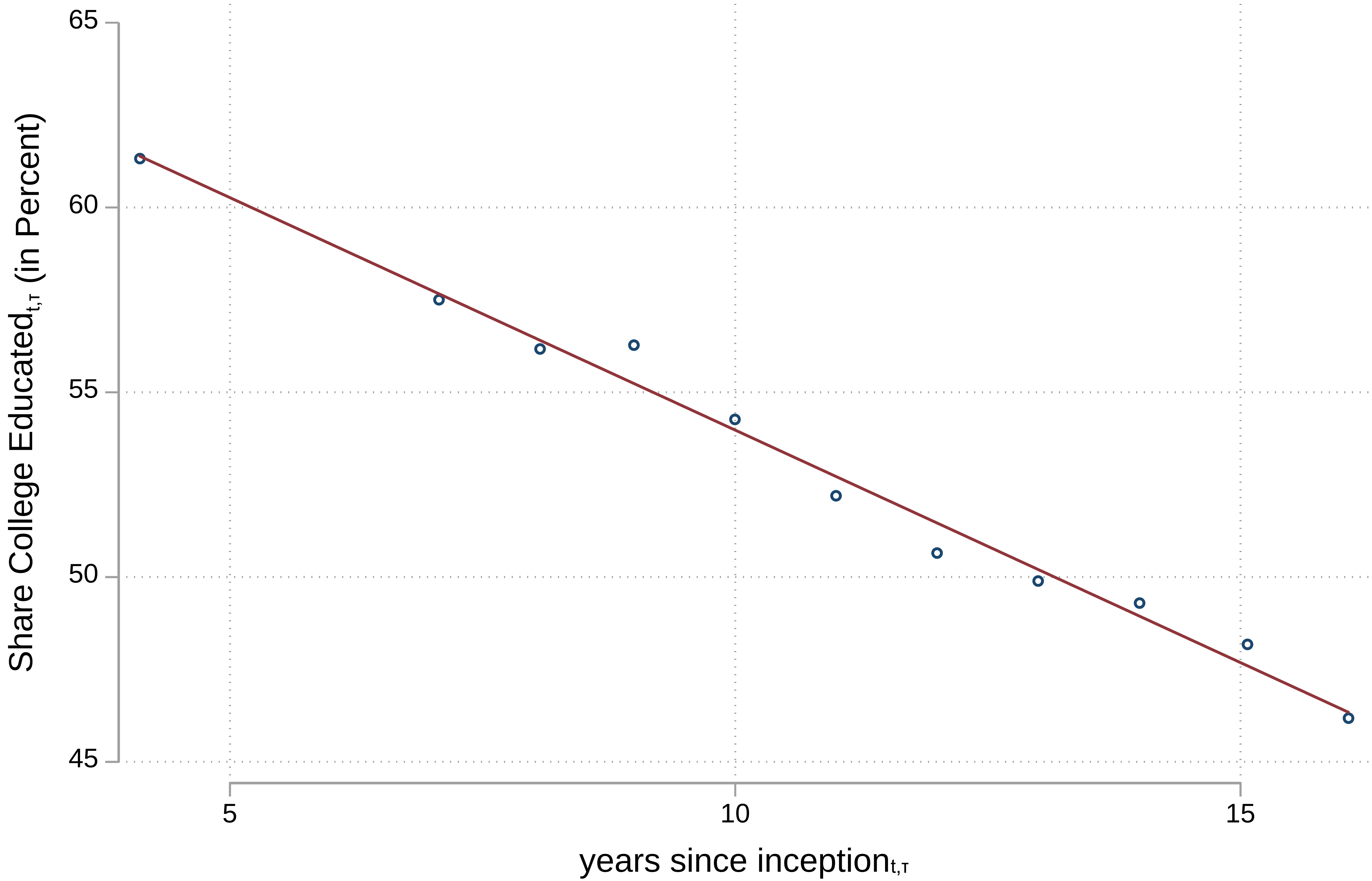
Notes: The pictures plot (year by year) the share of firms (red line) which mention technology related keywords in earnings calls, and the share of job postings (gray circles) in Burning Glass which mention technology related keywords. The vertical grey line highlights the year of inception of the technology, which is defined as the year at which firms timeseries (red line) attains at least 10\% of sample max. The overall correlation between these two time series is 68.87 \%.

Figure 4 - Share of college educated by year



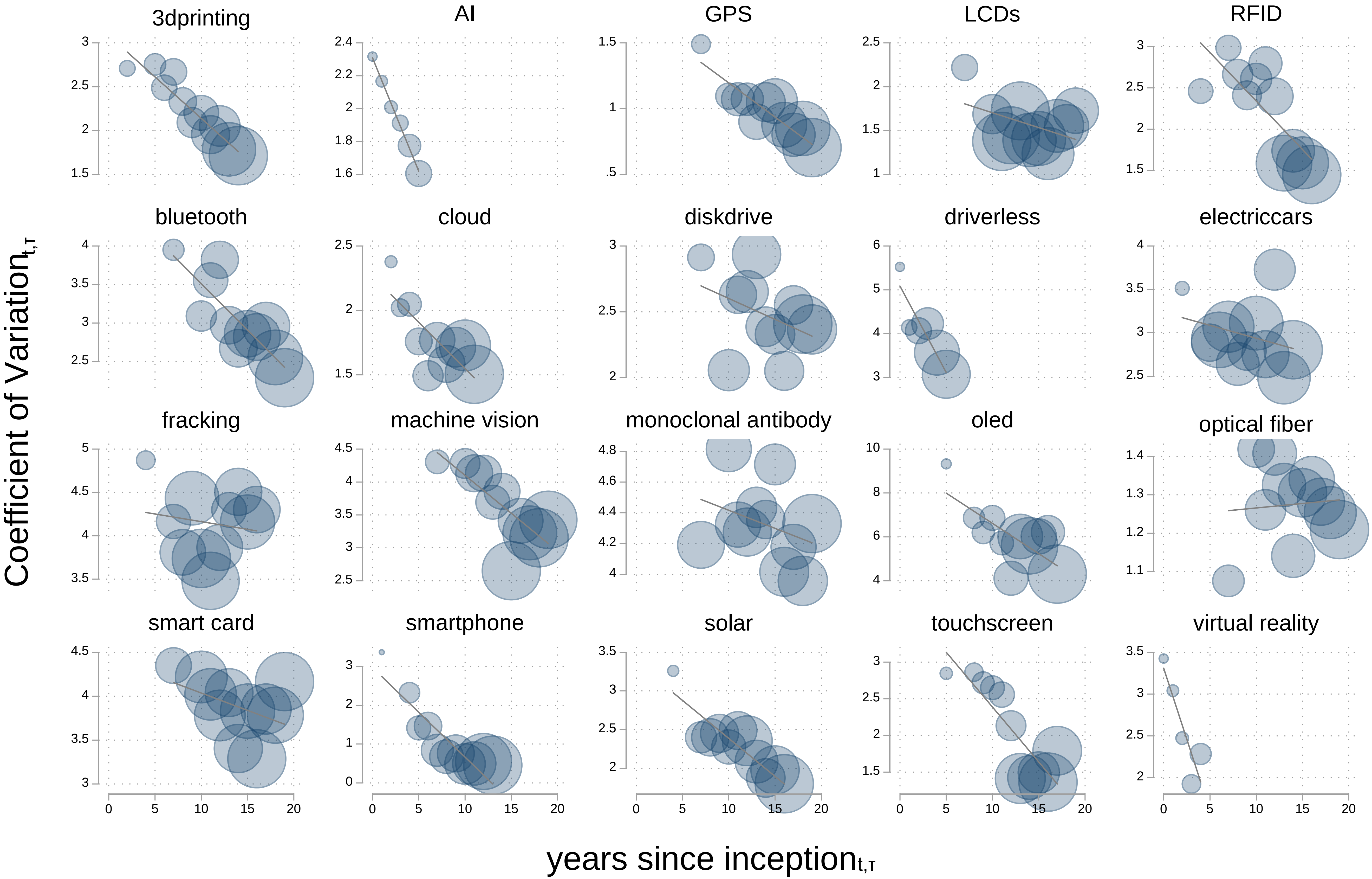
Notes: The pictures plot (year by year) the share of firms (red line) which mention technology related keywords in earnings calls, and the approximate share of technology job postings which require college educated people(blue circles, where the size of the circle represents the total number of technology job postings), calculated using $Skill_t^{\varsigma} = \frac{\sum_o N^{\tau}_{o;t}\chi_{o,2015}}{\sum_o N^{\tau}_{o;t}}$, where $\chi_{o,2015}$ is the share of college educated people in an occupation in ACS 2015 and $N^{\tau}_{o;t}$ is the number of technology job postings in technology τ. The vertical grey line highlights the year of start of the technology, which is defined as the year at which firm time series (red line) attains at least 10% of sample max.

Figure 5 - Share of college educated by year since start of technology



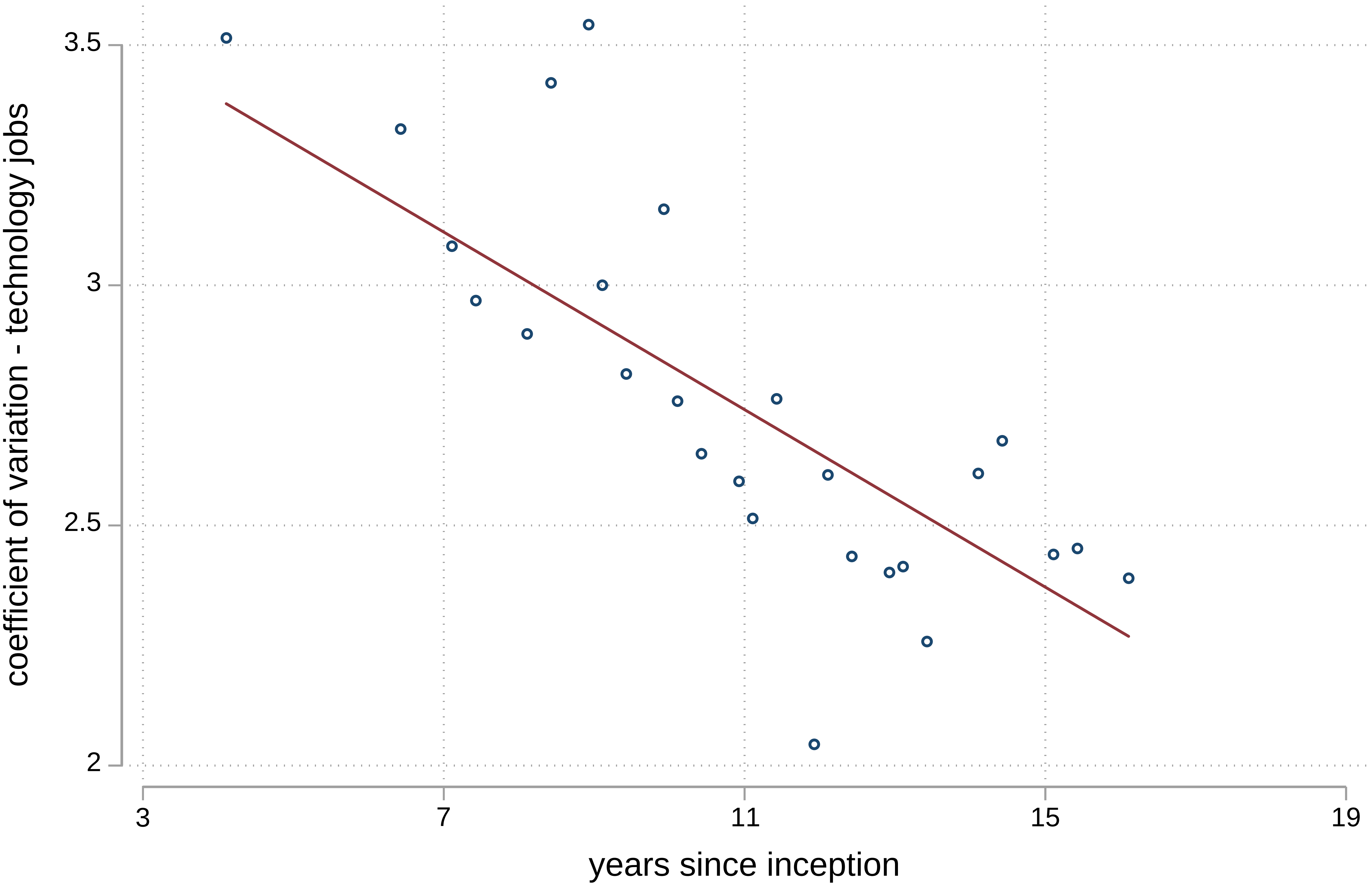
Notes: The figure plots a bin scatter (with 10 bins) of the approximate share of college educated technology job postings for each technology and calendar year against the years since start of the technology. The approximate share of college educated job postings for a technology is measured as $Skill_t^{\varsigma} = \frac{\sum_o N_{o;t}^{\tau} \chi_{o;2015}}{\sum_o N_{o;t}^{\tau}}$, 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. The figure controls for technology fixed effects.

Figure 6 – Coefficient of Variation by year since start of technology



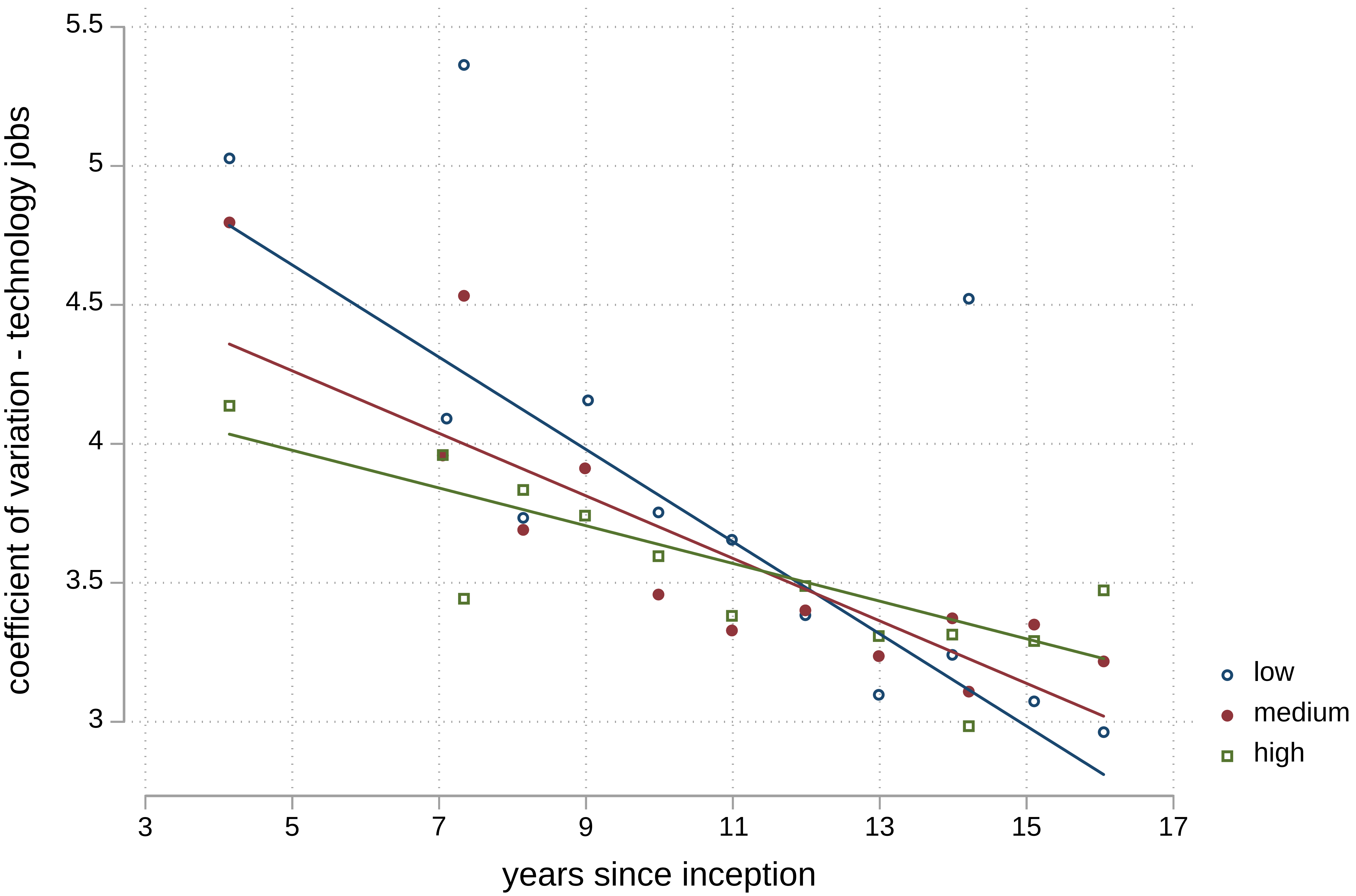
Notes: The figure plots coefficient of variation measured as coefficient of variation of normalized share of technology jobs for each of 20 technologies by year from 2007-2019 against the years since start of the technology, where $Normalized\ share_{cbsa,\tau,t} = \frac{share\ jobs\ expsed_{cbsa,\tau,t}}{share\ jobs\ expsed_{\tau,t}}$.

Figure 7 – Coefficient of variation by year since start of technology



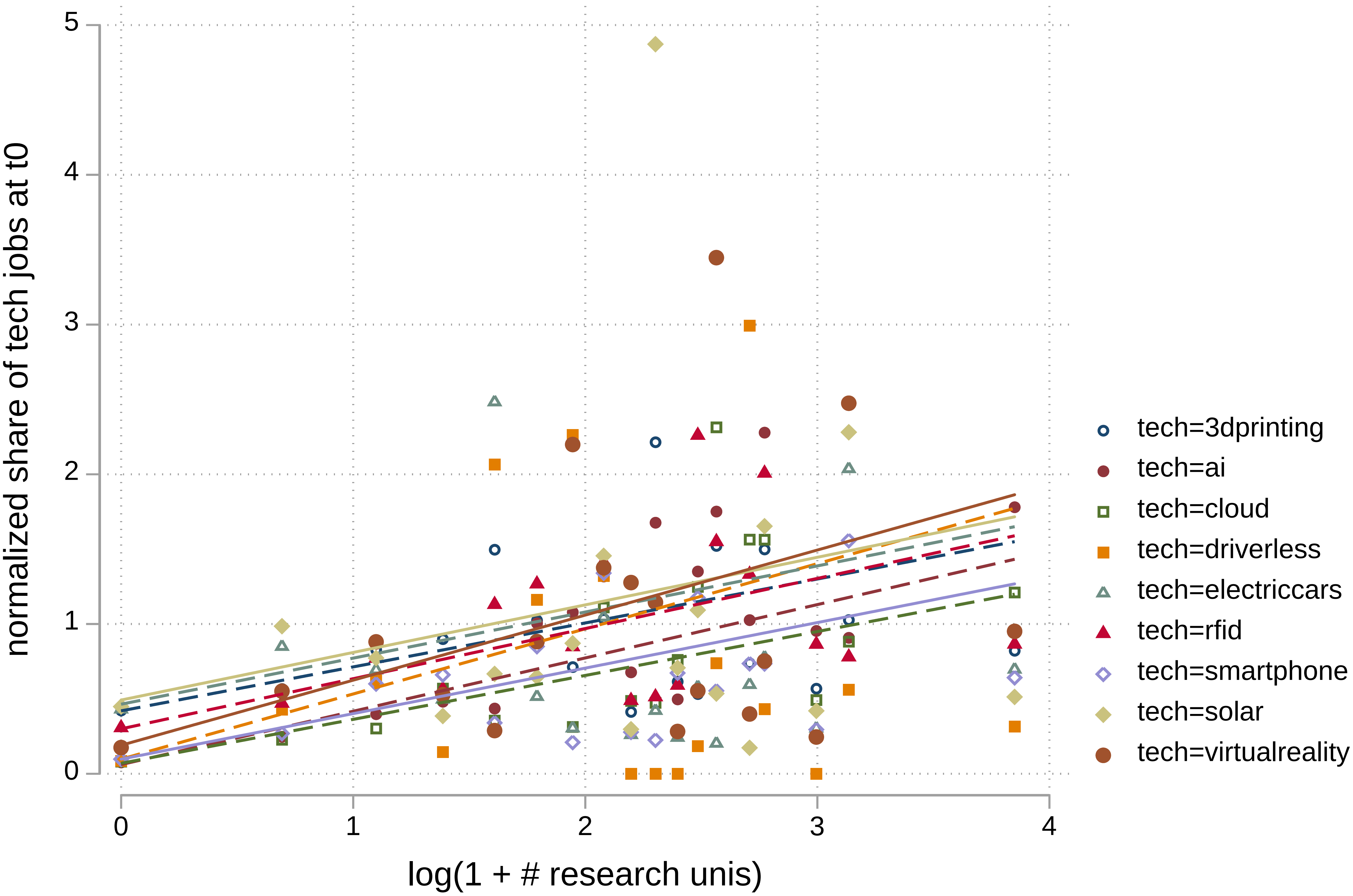
Notes: The picture shows binscatter of coefficient of variation by technology and time against year since start of the technology controlling for technology level fixed effects. Coefficient of variation is calculated over normalized share of technology job postings for each technology and for each year between 2007 and 2019.

Figure 8 – Coefficient of Variation by year since start of technology – by skill



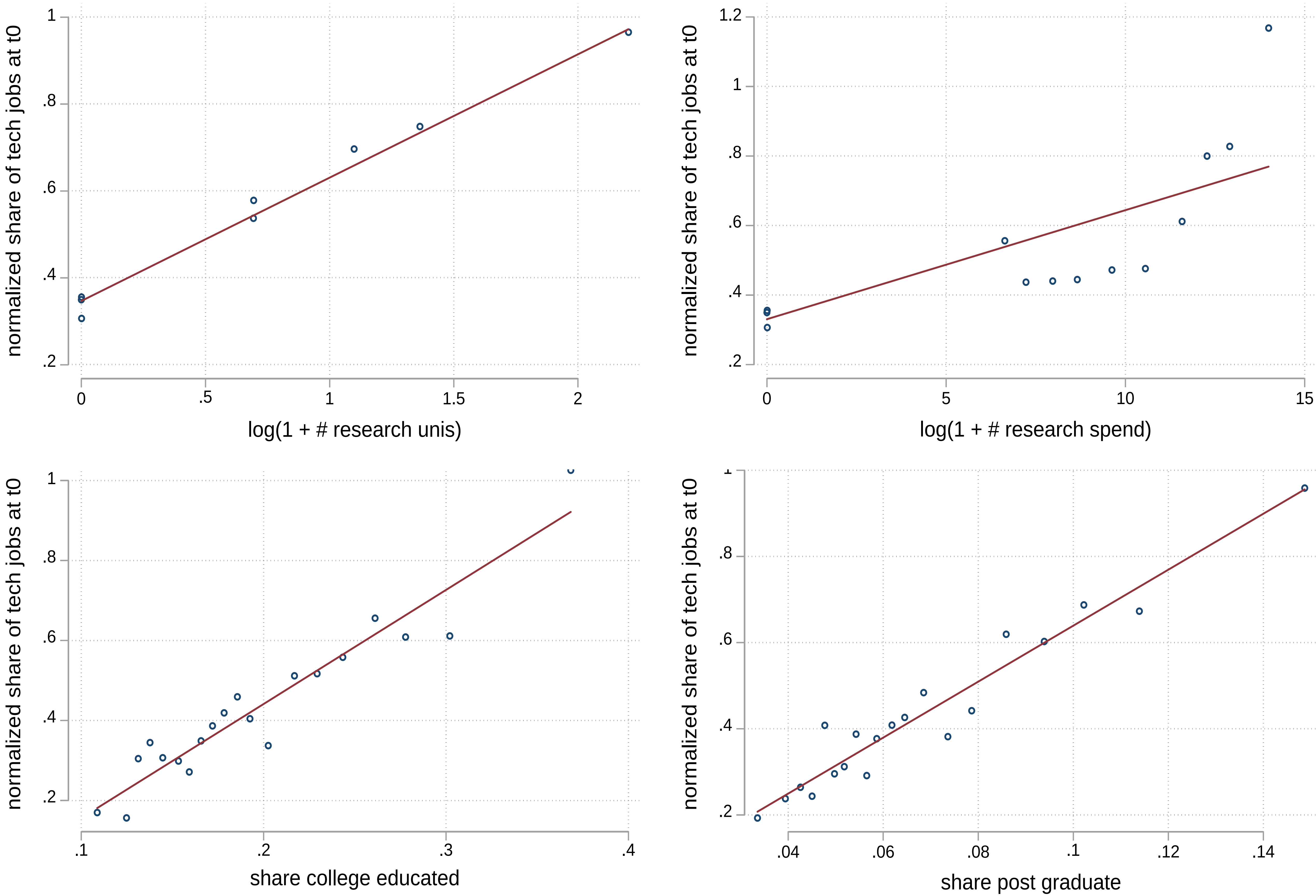
Notes: This figure plots a binscatter with 30 bins of coefficient of variation by technology and time against year since start of the technology by skill. To calculate the coefficient of variation by skill, we aggregate the job postings data over occupation, CBSA and year, and then separately for high skill occupations (with share of college educated people > 60%), medium skill occupations (with share of college educated people > 30%), and low skill occupations (with share of college educated people < 30%). Coefficient of variation is calculated over Normalized share of technology jobs by skill group, technology and time.

Figure 9 - Technology jobs vs local skill composition



Notes: The figure plots a binscatter of normalized share of jobs at time t0 on various measures of skill level of the CBSA. For some of our technologies, we do not have BG coverage throughout their lifetime. For these technologies, we take the starting year to be 2007, the starting year for our BG sample.

Figure 10 - Technology jobs vs local skill composition



Notes: The figure plots a binscatter of Normalized Share at time t0 (start date of the technology) on various measures of skill level of the CBSA. For some of our technologies, we do not have BG coverage throughout their lifetime. For these technologies, we take the starting year to be 2007, the starting year for our BG sample. The figures control for technology level averages.

Appendix Figure 1 - Sample USPTO Patent and Technical Bigrams

United States Patent [19] — [11] **3,994,012**

Warner, Jr. — [45] **Nov. 23, 1976**

[54] **PHOTOVOLTAIC SEMI-CONDUCTOR DEVICES**

[75] Inventor: **Raymond M. Warner, Jr.,** Edina, Minn.

[73] Assignee: **The Regents of the University of Minnesota,** Minneapolis, Minn.

[22] Filed: **Feb. 17, 1976**

[21] Appl. No.: **658,307**

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 575,258, May 7, 1975.

[52] **U.S. Cl.** **357/30;** 357/4; 357/20; 357/49; 357/89

[51] **Int. Cl.**^{2} **H01L 27/14;** H01L 27/12

[58] **Field of Search** 357/30, 4, 20, 89, 49, 357/50

[56] **References Cited**

UNITED STATES PATENTS

3,509,433	4/1970	Schweder	317/234
3,579,391	5/1971	Buie	148/175
3,598,664	8/1971	Kilby	148/175
3,666,548	5/1972	Brack	117/212
3,761,326	9/1973	Weckler	148/175
3,829,889	8/1974	Allison	357/49
3,836,793	9/1974	Haitz	307/303
3,877,052	4/1975	Dixon	357/17
3,886,587	5/1975	Newlay	357/49
3,896,485	7/1975	Early	357/24
3,943,555	3/1976	Mueller	357/49

Primary Examiner—Martin H. Edlow
Attorney, Agent, or Firm—Merchant, Gould, Smith, Edell, Welter & Schmidt

[57] **ABSTRACT**

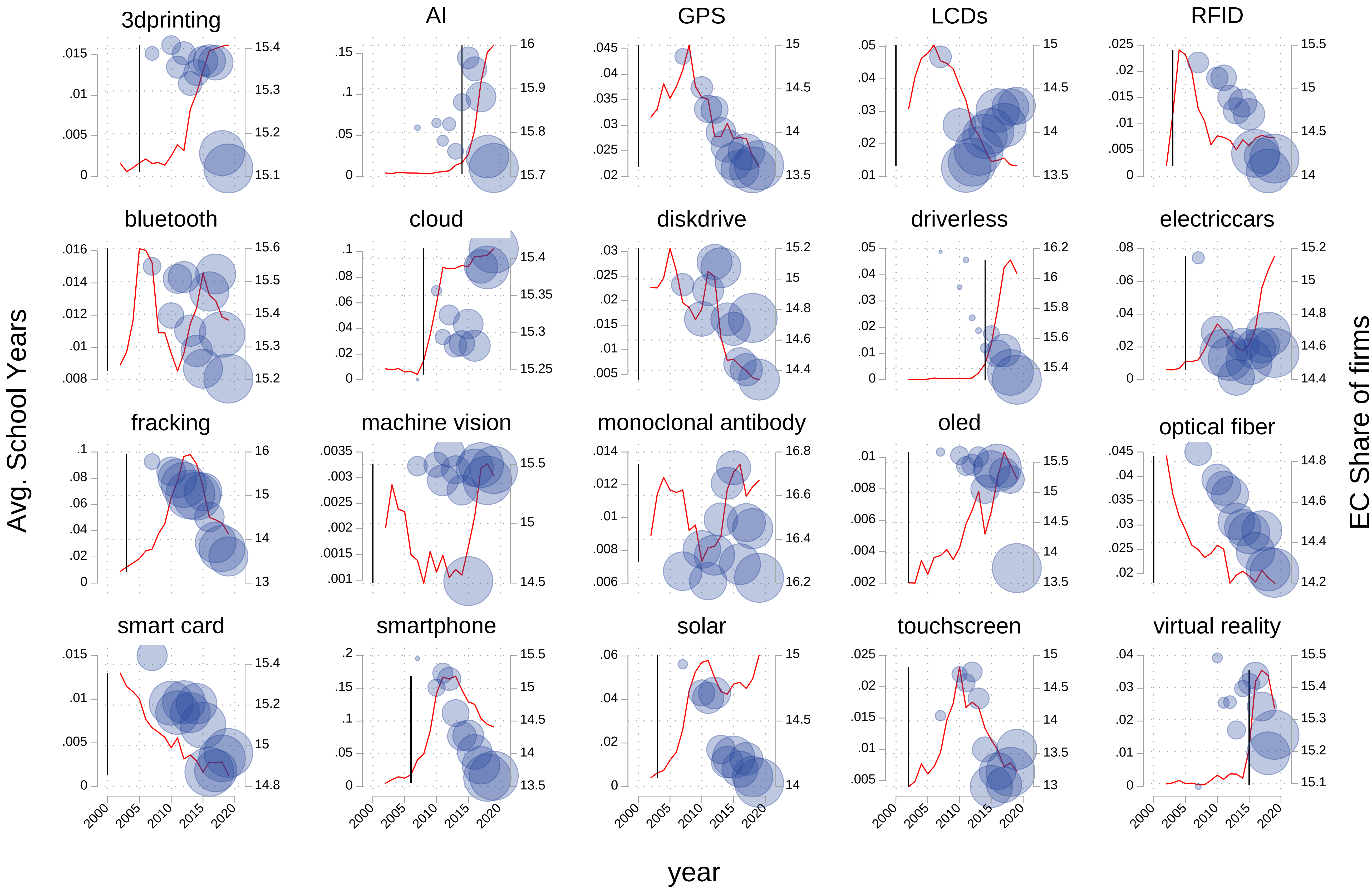
Apparatus and method for constructing by means of standard high-yield microelectronic batch fabrication processes, reliable, monolithic high-voltage photovoltaic cells and highly efficient photovoltaic arrays therewith. A thin layer of single-crystalline semiconductor material containing a plurality of sublayers defining one or more active junctions in planes parallel to an upper irradiated surface thereof, overlies a supportive insulating substrate body. Widely spaced pairs of elongate heavily doped zones of opposite conductivity types produced by two short diffusion steps extend into the thin layer, defining photovoltaic cells therebetween and providing low-impedance conductive paths for photovoltaic carriers generated in the thin layer to the upper irradiated surface. By overlapping opposite-conductivity pairs of the heavily doped elongate zones, simultaneous dielectric isolation and series connection of adjacent cells is achieved. The elongate zones of individual cells can be interdigitated to decrease parasitic series resistance of the cells. A dielectric barrier interposed between adjacent cells can also be used to isolate the cells from one another. Means for creating an inversion layer at the irradiated surface, reflective means for causing multiple passes of received solar energy through the photovoltaic devices and use of an underlying junction formed by heavily and lightly doped regions of the same conductivity type can readily be employed to further increase cell efficiency.

59 Claims, 44 Drawing Figures

Panel A: All bigrams		Panel B: Technical bigrams	
Bigram	Frequency	Bigram	Frequency
of the	223	conductivity type	147
of said	175	crystalline semiconductor	111
thin layer	147	semiconductor material	105
conductivity type	147	elongate zones	77
layer of	141	irradiated surface	69
single crystalline	139	photovoltaic cell	69
crystalline semiconductor	111	upper irradiated	59
semiconductor material	105	photovoltaic cells	54
said first	97	solar cell	43
the thin	94	type layer	40

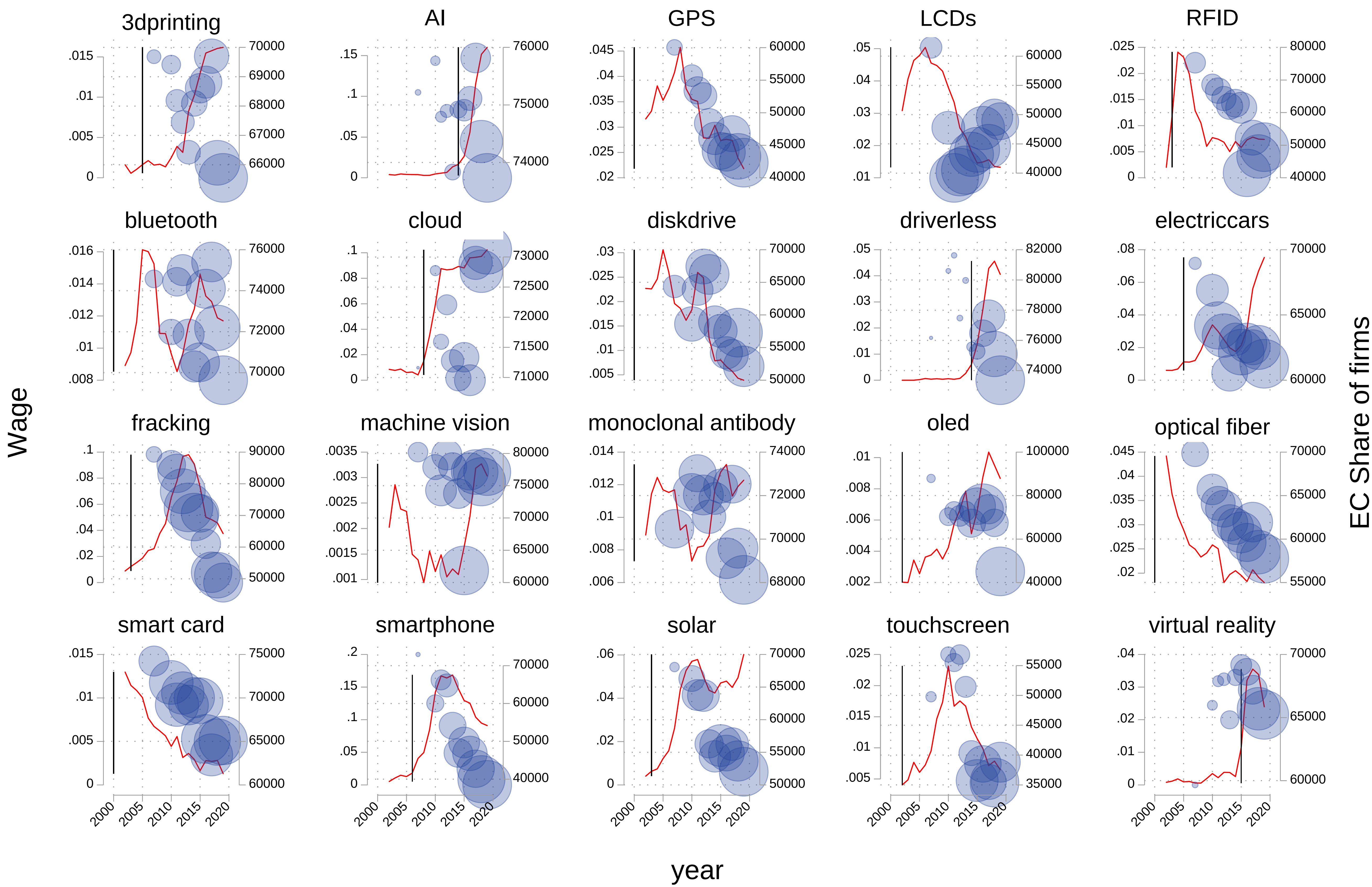
Notes: This figure shows all (Panel A Column 1) and technical bigrams (Panel B Column 1) along with their frequencies (in Column 2) obtained technical bigrams (detailed in section 2) for a sample USPTO patent titled “Photovoltaic semi-conductor devices”.

Appendix Figure 2 – Average Years of Schooling by Year



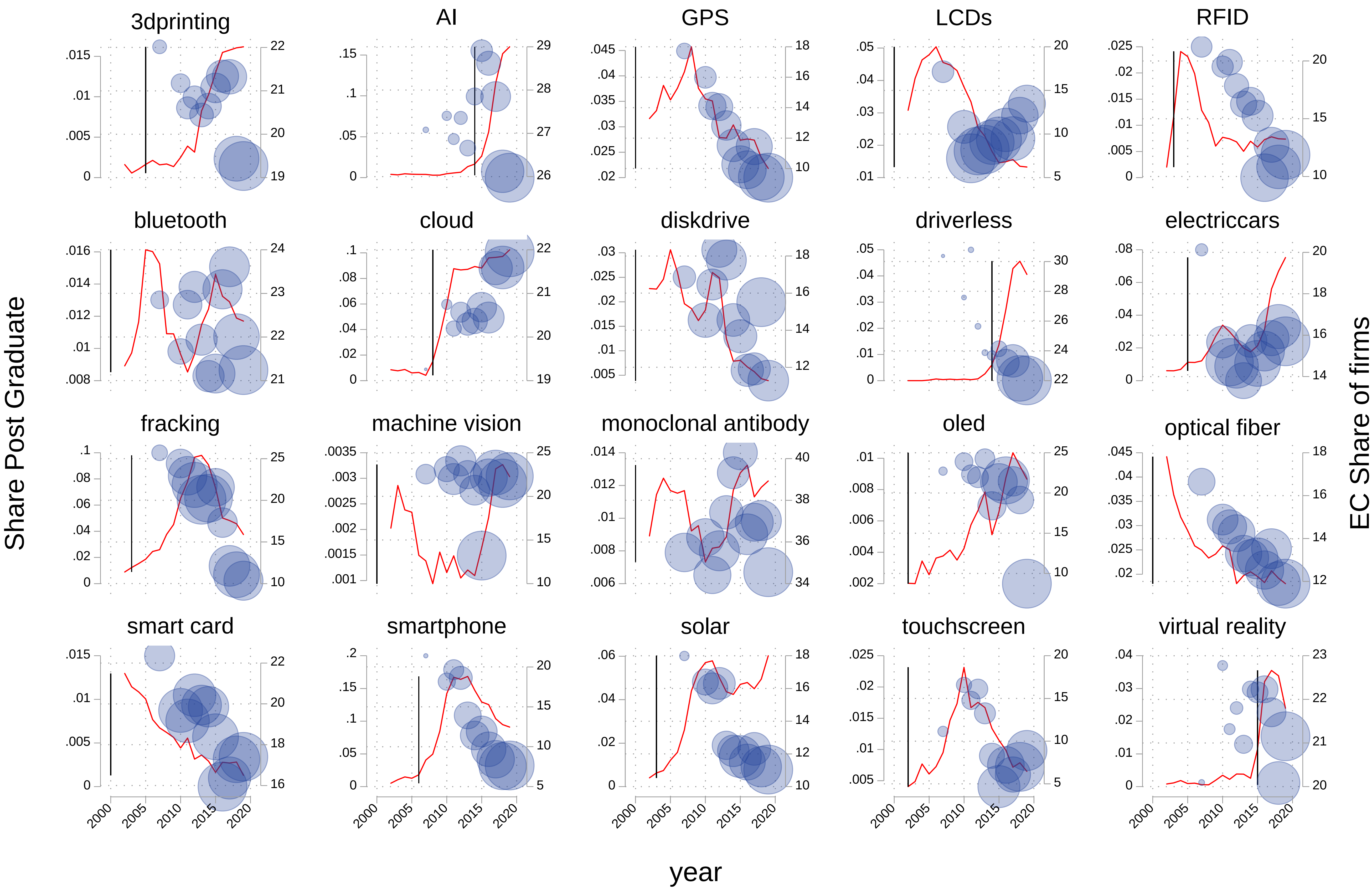
Notes: The pictures plot (year by year) the share of firms (red line) which mention technology related keywords in earnings calls, and the approximate average years of schooling for technology jobs (blue circles, where the size of the circle represents the total number of technology job postings), calculated using $Skill_t^{\varsigma} = \frac{\sum_o N_{o;t}^{\tau} \chi}{\sum_o N_{o;t}^{\tau_{o,2015}}}$, where $\chi_{o,2015}$ is the average years of schooling for persons employed in an occupation in ACS 2015 and $N_{o;t}^{\tau}$ is the number of technology job postings in technology τ. The vertical grey line highlights the year of start of the technology, which is defined as the year at which firm time series (red line) attains at least 10% of sample max.

Appendix Figure 3 – Average Wage by Year



Notes: The pictures plot (year by year) the share of firms (red line) which mention technology related keywords in earnings calls, and the approximate average wage offered for technology jobs (blue circles, where the size of the circle represents the total number of technology job postings), 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 average wage offered for persons employed in an occupation in ACS 2015 and $N_{o;t}^{\tau}$ is the number of technology job postings in technology τ. The vertical grey line highlights the year of start of the technology, which is defined as the year at which firm time series (red line) attains at least 10% of sample max.

Appendix Figure 4 – Share Post Graduate by Year



Notes: The pictures plot (year by year) the share of firms (red line) which mention technology related keywords in earnings calls, and the approximate share of post graduate employees in technology jobs (blue circles, where the size of the circle represents the total number of technology job postings), 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 post graduates employed in an occupation in ACS 2015 and $N_{o;t}^\tau$ is the number of technology job postings in technology τ. The vertical grey line highlights the year of start of the technology, which is defined as the year at which firm time series (red line) attains at least 10% of sample max.