MEGA FIRMS AND RECENT TRENDS IN THE U.S. INNOVATION: EMPIRICAL EVIDENCE FROM THE U.S. PATENT DATA

Yuheng Ding, Serguey Braguinsky, Joonkyu Choi, Karam Jo, Seula Kim*

ABSTRACT

We use the U.S. patent data merged with firm-level datasets and establish new facts about recent trends in "novel patents" – inventions that introduce new technology combinations. We find that such innovations were declining from 1980 to 2008, and a strong rebound has followed since then. Our evidence suggests that mega firms play an important role in driving these trends as compared to their smaller competitors and VC-backed startups. In particular, the share of novel patents by mega firms has been increasing since the early 2000s, and they are increasingly creating new combinations integrating Information and Communications Technology (ICT) components with non-ICT components. Furthermore, we find that those new technology combinations created by mega firms are becoming more influential and increasingly built on by other firms than those created by their smaller competitors or VC-backed startups by having more follow-on patents. Overall, our findings suggest that mega firms are increasingly engaging in experiments that create room for other entities to conduct follow-up innovations.

March 15, 2023

* Ding, University of Maryland, email: yuheng@umd.edu; Braguinsky (corresponding author), University of Maryland and NBER, email: sbrag@umd.edu; Choi, Board of Governors of the Federal Reserve System, email: joonkyu.choi@frb.gov; Jo, Korea Development Institute, e-mail: karamjo@gmail.com; Kim, University of Maryland, email: seulakim@umd.edu

1. Introduction

The concentration of economic activities in the largest businesses, so-called mega firms, in product and local labor markets has been increasing over the past few decades (Autor et al., 2020b; Hsieh and Rossi-Hansberg, 2022; Yeh et al., 2022). Recent literature explores two broad sets of interpretations for the rise of mega firms. Some studies have emphasized that this trend is accompanied by the rise in market power (De Loecker et al., 2020), possibly driven by the increase in entry barriers, regulation, and lobbying activities that stifle competition (Covarruias et al., 2020; Gutierrez and Philippon, 2019). Other studies have cast doubt on the increasing market power interpretation (Foster et al., 2022) and instead emphasize increased competition or winner-takes-all dynamics caused by globalization and technological advances that enable large firms to exploit economies of scale (Autor et al., 2020b; Hsieh and Rossi-Hansberg, 2021; Kwon et al., 2023).

One key issue that remains understudied in this debate is the role of mega firms in the economy-wide innovation and knowledge diffusion. Similar to the increase in market concentration, Akcigit and Ates (forthcoming) document that the share of patents held by top one percent among patenting firms has been increasing dramatically over the past three decades from 35 percent to nearly 50 percent. They argue that mega firms may be increasingly building stocks of patents that make it difficult for other firms to compete in the technology domain, leading to slower diffusion of knowledge and slowdown in business dynamism. Alternatively, mega firms may be increasingly investing in innovation and experimentation that could potentially create room for subsequent innovation by other firms.

In this paper, we establish new facts about the role of mega firms in innovation and patenting in the United States in recent decades. Patents may not always reflect true innovations as a substantial number are filed for purely strategic reasons and many are actually never used (for

empirical evidence see, e.g., Bessen and Hunt, 2004; Noel and Schankerman, 2013; and Torrisi et al., 2016). To better measure the extent of innovation and experimentation, we pay a particular attention to patents—which we refer to as novel patents—that introduce new combinations of technological components that had never been utilized together before.[2] This approach is motivated by the technology literature that provides evidence that novel patents create pathways for new technological trajectories and are thus of primary importance in the innovation process as Schumpeterian "new combinations" (see, e.g., Verhoeven et al., 2016; Pezzoni et al., 2022; Epicoco et al., 2022). We investigate how much mega firms contribute to novel patents and what type of new potential technological trajectories they generate.

Our examination of novel patents and the roles played by mega firms in generating them yields several novel empirical findings. First, there has been a secular decline in the share of novel patents in total patent applications from 1980 to 2008, followed by a strong rebound since then. After about 2008, there has been a sharp rise in both the number of novel patent applications and all patent applications, with the increase in the former proportionately much larger than the increase in the latter. The share of novel patents had reached the level last observed in the mid-1990s by 2016. Second, the share of novel patent applications accounted for by mega firms has been increasing already since the early 2000s, but especially so in more recent years, with 2/3 of the increase in the last one and a half decades coming after 2009. Third, there is a shift in the direction of technological innovations by mega firms and their industry composition; mega firms are increasingly generating novel combinations that combine Information and Communications Technology (ICT) components with non-ICT components for the first time and such novel

[2] We use the patent classification system designed by the US Patent Office (USPTO) to measure technology components. See https://www.uspto.gov/web/patents/classification/cpc/html/cpc.html for details. While we use the main group level of patent classes disaggregation as our main specification in this paper, the findings go through at different levels of aggregation.

combinations are generated not just by firms whose primary industry is ICT-related but also by firms operating in non-ICT-related industries.

Figure 1 shows one specific example—U.S. patent 8,188,868 B2, applied for in 2006 and granted in 2012. This patent meets the definition of a novel patent as it combines CPC group G08C17 with CPC groups A43B3 and A41D1 for the first time. As seen from their names, G08C17 is a component in the CPC section G (physics), while A43B3 and A41D1 both belong to the CPC section A (human necessities). Note also that the patent is assigned to NIKE, Inc. whose main industry is footwear and garments. In this case, NIKE basically implanted a wireless transmitting device into a T-shirt to enable the athlete to monitor vital signs and performance.

[Figure 1 around here]

Fourth, adopting measurement approach from the previous literature (e.g., Pezzoni et al., 2022), we use the number of follow-on patents—the patents that use the same new combination as the one first introduced by a novel patent—to directly measure the diffusion of new technological trajectories. To understand the role of mega firms in the diffusion process, we investigate whether follow-on patents are mostly created by the firm that initially introduced the novel technological combination (the "focal firm," hereafter) or whether follow-on patents are created by other firms, indicating the diffusion of knowledge beyond the focal firm's boundaries.

When examining the share of follow-on patents assigned to the focal firm based on the first five years after the new combination was first produced, we do not find evidence that a larger share of follow-on patents is being assigned to focal firms over time. And mega firms, if anything, have a smaller share of follow-on patents assigned to themselves compared to other firms. Second, the absolute number of follow-on patents generated by new combinations during the recent surge in novel patent applications is significantly higher than in previous years, especially when compared

to the first half of the 2000s. Indeed, this number had almost tripled from the 2001-2006 period to the 2007-2016 period for all U.S. patent assignees and increased by more than five times for mega firms. The number of follow-on patents generated by new combinations in 2007-2016 is also almost two times higher than what it was in the 1990s (three times higher for mega firms). Third, when looking at long-term (20-year) diffusion curves of new technology trajectories generated by most impactful new combinations, we find that mega firms have the largest average numbers of follow-on patents assigned to entities other than themselves, followed by VC-backed startups, with all other firms falling behind. Overall, our direct measure does not corroborate the conjecture of slower diffusion of new knowledge combinations, especially so with respect to mega firms.[3]

We also examine the role of VC-backed startups in novel patents, and we find that their role is high relative to their share in the economy, but their role has been relatively stable since the mid-1990s. That said, recent VC-backed startups have played a disproportionately large role in generating most impactful novel combinations. We discuss that in more detail below.

Finally, we examine the within-firm correlation between novel patents and firm outcomes, such as employment and revenue productivity, and present suggestive evidence that novel patents are generally positively associated with firm performance even after controlling for changes in overall patenting behavior (i.e., total patent stock of the firm). While these findings cannot be interpreted as causal at this stage, they do suggest that novel patents may indeed represent especially important innovations. We plan to pursue this line of investigation in future research.

Our findings have important policy implications. If it is true that dominant mega firms are stifling innovation and slowing down knowledge diffusion, there may a scope for regulatory

[3] Akcigit and Ates (forthcoming) also examine patent reassignment as another way in which mega firms may be using patents strategically. The issue of reassignment is beyond the scope of our analysis as we focus on the *production* of novel patents and their diffusion to subsequent patents produced following the same technological trajectory.

intervention. If, however, mega firms are the key actors generating novel technologies, then such an approach may backfire. Examinations of large regulatory interventions of the past paint a mixed picture. On the one hand, Watzinger and Schnitzer (2022) show a positive impact of the breakup of the Bell system on subsequent U.S. innovation. On the other hand, Klepper (2016) argues that anti-trust action against RCA was one of the triggers in the total demise of the U.S. color TV receivers industry. With the U.S. technological dominance, especially in ICT, facing increasing global challenges, the stakes couldn't be higher. We provide further discussion in the concluding section.

The rest of the paper is organized as follows. In the next section we describe data construction and the sample. In Section 3 we present some basic facts about the evolution of patenting and specifically novel patenting activity in the United States in the past several decades and the changing role of mega firms. In Section 4, we examine recent changes in the direction of novel combinations with ICT as a general-purpose technology and the role of mega firms. We also discuss the role of VC-backed startups in creating high-impact novel combinations. Section 5 links novel patents by mega firms to some measures of their performance. Section 6 concludes.

2. Data

The primary data sources are the USPTO PatentsView, S&P's Compustat and VentureXpert data.[4] The USPTO PatentsView tracks all patents ultimately granted by the USPTO from 1976 onward.[5] This database contains detailed information for granted patents including application and grant dates, technology class categories, patent inventors and citation information, and the names and addresses of patent assignees. In this paper, we collect utility patents granted to U.S. assignees

[4] We also have access to administrative datasets including the Longitudinal Business Database (LBD) and the Longitudinal Employer-Household Dynamics (LEHD) datasets from the U.S. Census Bureau. While these data are not used for the current version of the paper, we plan to utilize these data sets to develop the paper going forward.
[5] See details from https://patentsview.org/download/data-download-tables.

between 1976 and 2020 to track economy-wide innovation activities and in particular, the creation and trajectory of new technological combinations, which is discussed later in this section. Before that, we describe detailed procedures we took to match the USPTO patents with Compustat and VentureXpert to get information of public traded firms, in particular mega firms, and VC-backed startups respectively.

2.1 USPTO-Compustat

For the analyses involving mega firms, we use S&P's Compustat data to track publicly listed firms in the U.S. We created our own bridge between the U.S. patenting firms in the USPTO patent database and Compustat firms through a standard name-matching and internet-based matching algorithm as in Autor et al. (2020a).

First, we standardize firm names in both datasets using the algorithm provided by the NBER PDP and use the standardized names in the matching process. We define the patenting firms as patent assignees that are located in the U.S. with assignee type equal to 2 (U.S. company or corporation) in the USPTO data.

The first match procedure involves identifying firms with precisely the same standardized names in both datasets. Following the previous studies, we do not use address information in Compustat throughout the entire match process as the data only reports information for headquarters, which can be different from the exact address of the establishments that filed patent applications to the USPTO. For the unmatched USPTO firms, we use the stem name (standardized firm names without suffixes) to find matches.

For the rest of the unmatched U.S. patenting firms after the standard name matching, we apply the internet-based matching algorithm to identify the same firms in Compustat. Specifically, we put every patent assignee and Compustat firm name into the Google.com search engine, collect

the URLs of the top five search results, and identify any given pair of the patent assignee and Compustat firm as the same firm if they share at least two identical search results. If any of these patenting firms remain unmatched, we utilize web-URL information in Compustat and find the corresponding firms if the top five search results of the unmatched patenting firms exactly match the web-URL of the Compustat firms.

For all the remaining unmatched U.S. patenting firms in the USPTO data after the previous steps, we use the NBER PDP and find matches in Compustat. The NBER PDP did extensive manual matching to identify the same firms across the two datasets. Thus, this procedure helps us to reduce our burdens of manually searching the unmatched USPTO firms. Lastly, we do our own manual matching to identify matches between the USPTO and Compustat firms. We manually inspect the match results to screen out false matches, especially for firms with many patent applications at the end of each procedure.

The above procedure matches 68.0% of utility patent applications filed by U.S. patenting firms, and 24.5% of U.S. patenting firms to Compustat firms from 1976 to 2016. See more details in the Appendix.

2.2 USPTO-VentureXpert Crosswalk and VC-backed Startups

For the analyses involving VC-backed startups, we link the USPTO to VentureXpert data by using a matching algorithm based on company name and location information similar to Ma (2020), Bernstein, Giroud, and Townsend (2016), and Gonzalez-Uribe (2020). Overall, 25.6% of companies in the VentureXpert data have been matched to the USPTO based on exact name-location matching. [6]

[6] For analysis conducted in the current draft, we utilize matching results based on exact name-location matching to minimize false positive rate. Going forward, we plan to improve the USPTO-VentureXpert crosswalk by using fuzzy name matching algorithm used by previous studies (Ma, 2020; Bernstein, Giroud, and Townsend, 2016; Gonzalez-Uribe, 2020).

We utilize information on the founding date and exit date from the VentureXpert data to identify VC-backed startups. Specifically, we classify patent assignees as VC-backed startups when they file patent applications between the founding year and the year of exit (i.e., IPO, M&A, or bankruptcy, etc.), or the company remains active (i.e., does not have an exit event) by the end of our sample period. In other words, a VC-backed startup identified in earlier years will be removed from the set of startups once it exits via IPO, M&A, bankruptcy, and so on.

2.3 New technological combinations, novel patents, and follow-on patents

To identify technological components underlying an invention, we exploit the detailed information provided by the USPTO patent database on the technological content of inventions. Each patent document at the USPTO reports technology classes based on all disclosed information in the invention. Indeed, to conduct an efficient patent search, the USPTO requires patent examiners to objectively classify an invention into technology categories based on "invention information" and "additional information". In this paper, we use technology classes based on "invention information", which, according to the USPTO, contains "technical information in the total disclosure of a patent document (for example, description, drawings, claims) that represents an addition to the state of the art.[7]

We utilize the new patent classification system – the "Cooperative Patent Classification" (CPC) – introduced in 2013 to measure technological components of inventions. The CPC scheme is a hierarchical system with multiple levels of classifications.[8] In this paper, we use the "Main Group" as our main level of classifications for patents. It is the most comparable level of

[7] According to the USPTO Manual of Patent Examining Procedure (https://www.uspto.gov/web/offices/pac/mpep/s905.html), "Addition to the state of the art" means all novel and unobvious subject matter specifically disclosed in a patent document, which advances the state of the art, i.e., the technical subject matter disclosed that is not already in the public domain.

[8] See https://www.uspto.gov/web/offices/pac/mpep/s905.html for details on the hierarchical structure of the CPC scheme.

classification to the USPC subclass widely used in the previous literature. We use "technological components" and "main group" interchangeably in this paper. While new technological components are added over time, the USPTO reclassifies old patents according to the new CPC code, which ensures the comparability of patents over time. By 2016, there are 7,246 distinct main groups under the CPC scheme excluding those under CPC Section Y.[9]

Following previous studies (Pezzoni et al., 2022; Fleming et al., 2007; Strumsky and Lobo, 2015), we define a new technological combination as the first time a pairwise combination of technological components ever appears in a patent. Patents incorporating such new technological combinations are "novel patents." While our analysis is based on utility patents assigned to the U.S. entities, we identify a pair of technological components as a new combination only if it is the first combination that appears among all utility patents granted to both U.S. and non-U.S. entities since 1976. Since the earliest year of the USPTO PatentsView data is 1976, we do not observe the complete history of technological combinations created before then. We use the first couple of years, 1976-1979, as a buffer period to capture the history of technological components and we track novel combinations starting from 1980. Indeed, for much of the analysis, we start in 1991 to minimize the likelihood of misclassifying new combinations.

To study the diffusion and technological trajectories of new combinations, we identify the pool of follow-on inventions that (re-)use the same pairwise connection between technological components as first created by the focal new combination. Specifically, we count the cumulative number of patented inventions that re-use the new technological combination up to 20 years following the appearance of the focal new combination and up to the end of 2020. Furthermore, we also distinguish follow-on patents assigned to the same v. different assignee(s) as the

[9] We exclude technological components under Section Y in our analysis since it is used for general tagging of new technological developments and documents which are already classified or indexed elsewhere.

assignee(s) of the focal new combination, based on assignee information provided by the USPTO. Occasionally, patents are assigned to multiple assignees. In such cases, we classify the follow-on patent as the one assigned to the same original assignee if any of its assignees are the same as those inventing the original new technological combination.

2.4 ICT and non-ICT related new combinations

To better understand the nature of new technological combinations, we further identify technological components closely related to ICTs by utilizing the NAICS-to-CPC crosswalk created by Goldschlag, Lybbert, and Zolas (2020) and ICT industry classification by Goldschlag and Miranda (2019). We first identify ICT industries based on 4-digit NAICS by following Goldschlag and Miranda (2019). Table A1 in the Appendix provides a complete list of ICT industries. Next, we classify a given CPC technology as ICT-related if it is linked to one of those ICT industries based on the NAICS-to-CPC concordance by Goldschlag, Lybbert, and Zolas (2020), which provides a probabilistic matching from the 4-digit NAICS to the 4-digit CPC subclass by using an "Algorithmic Links with Probabilities" (ALP) approach. The ALP extracts key words, i.e., search terms, from each NAICS industry description and combing through all the patent text to retrieve patents that contain the exact phrases of each search term. Then, the underlying CPC subclass of the retrieved patents are linked to each NAICS industry with a probability score reflecting their matching frequencies. The results are then reweighted to reduce noise and possible bias. In essence, the probabilistic matching helps us identify the most frequently used CPC technologies in ICT industries.

3. Mega firms and trends in patenting and novel patents: some basic facts

In their examination of the slow-down in knowledge diffusion and declining business dynamism in the United States, Akcigit and Ates (forthcoming) present evidence of increasing concentration

of patent ownership and the declining entrants' patent share since the 1980s (Akcigit and Ates, Figure 9, p. 45). Figure 2, constructed using our own calculations, is similar to Figure 9 in Akcigit and Ates while also adding the time trend in novel patents.[10] While the share of top 1% patent assignees in the stock of all patents is indeed increasing and the share of new entrants in patent applications is decreasing over time, the picture with respect to novel patents is somewhat different.

[Figure 2 around here]

First, as seen from the left panel of Figure 2, the time trend in the share of the top 1% patent assignees in the stock of novel patents follows the trend in their share in the stock of all patents until about 2000 (with somewhat lower levels), but then starts declining after 2000 and falls all the way back to the level of the early 1990s by the mid-2010s. In contrast, as seen from the right panel of Figure 2, while the share of entrants (assignees who are patenting for the first time) in novel patent applications is decreasing after 1990, the slope of this decline is much less steep than the slope of the decline in their share in total patent applications and appears to have stabilized around 8% since about 2005. The levels are also much higher compared to the entrants' share in total patent applications.

So far the emerging picture appears to be consistent with the notion that the use of patents in the U.S. is indeed becoming more concentrated; furthermore, the fact that the top 1% assignees account for a smaller stock of novel patents over time while entrants produce a disproportionately large share of such patents seems to vindicate the view that there might indeed be "abuse" of patents by the largest patentors (Akcigit and Ates, forthcoming, p. 44).

[10] We sent an email on March 3 to both Ufuk Akcigit and Sina Ates requesting their codes, but so far have not received a reply. We were thus unable to replicate their Figure 9 exactly because we suspect they apply a somewhat different data cleaning procedure compared to us. The overall trends in their Figure 9 and our Figure 2 are, however, similar.

The above picture may be misleading for two reasons, however. First, top 1% patent assignees are not necessarily mega firms (indeed, they may be not even be firms at all). If we are interested in the role of mega firms in innovation, we need to define them based not on their patenting behavior, but on economic criteria. In this paper, we (re-)define mega firms as top 50 in sales among all public firms in the Compustat data in each year during our sample (1980-2016). We then use the bridge between the U.S. patents and Compustat firms described in the previous section to measure patenting activity by such mega firms.[11] Also, assignees that patent for the first time are not necessarily startups. To examine how the role of startups, and specifically VC-backed startups in innovation has been evolving over time we use the bridge between the USPTO and VentureXpert data also described in the previous section. We present this examination below while in this section we focus on mega firms.

Figure 3 shows the dynamics of the share of mega firms (defined, once again, as those among top 50 in sales among U.S. public firms in each given year) in the cumulative number of all and novel patents of the U.S. public firms. [12] Interestingly, there is no increase in the overall concentration of patents over the 37 years of our sample. Instead, we see a steep decline in the 1990s-early 2000s, followed by an equally steep recovery starting from the mid-2000s. Second, while the share of mega firms in novel patents is a bit less than their share in total patents, the

[11] We identify mega firms in the Compustat data prior to merging it with the USPTO data; hence, we are actually left with between 20-30 firms after the match, depending on the year. That is, our sample of mega firms is comprised of those among the top 50 firms by sales in a given year, which also have at least one patent in the same year. Alternatively, we could define mega firms as the top 50 firms in terms of sales after excluding non-patenting firms. It turns out that the results are very similar regardless of whether we assign firms to the mega firms' category before or after merging the Compustat with the USPTO data (results with the alternative definition of mega firms are available upon request). Also, while we use our own USPTO-Compustat bridge in this paper, we checked our basic findings using the DISCERN (Duke Innovation & Scientific Enterprises Research Network) bridge (https://zenodo.org/record/4320782#.ZAzaKS1h1gg) and the results were, once again, very similar.

[12] The denominator in Figure 3 is the stock of patent applications by the U.S. public firms in Compustat, which is different from the denominator in Figure 2 which is the stock of patent applications by all U.S. patent assignees. We also constructed the figure showing the share of mega firms in the stock of patents by all U.S. assignees, and while the levels of the share of mega firms were lower (because the denominator includes all patents, not just by Compustat firms), the dynamics were similar. Details are available upon request.

dynamics are very similar. Thus, in contrast to the divergence between all patents and novel patents owned by top 1% patent assignees we saw in Figure 2, there is no such divergence among economically dominant mega firms, and the rapid increase in their patent share in recent years appears to be accompanied by an equally sharp increase in the share of their novel patents as well. These findings are robust to various different cutoffs one may want to use to define economically dominant mega firms—for example, they hold qualitatively if we redefine mega firms to be the top 50 in terms of sales among all patenting public firms, and they also hold if we define mega firms as, for instance, those in the top five percent of sales among all public firms in the Compustat data in any given year. It seems, therefore, that *economically successful* mega firms (a) are not necessarily increasing their total share in patents owned by the U.S. public firms, and (b) whenever their share in total patents does increase (as it does in recent years), it increases also in terms of novel patents.

[Figure 3 around here]

Importantly, the changes in the share of novel patents by mega firms is closely associated with changing dynamics of novel patents in the U.S. economy overall. In Figure 4 we present the dynamics of the number of novel patents and their share in total patent applications in the U.S. over time. As can be seen from this figure, the share of novel patents in total had been declining steadily from 1980 (the first year in our data) and until about 2007, while the absolute number of novel patent applications had been basically flat. In other words, the U.S. innovation system continued to generate novel patents in more or less similar numbers (between 3,000-4,000 per year) during the whole period above, although in somewhat lower numbers in the 1980s and the first half of the 2000s and somewhat higher numbers in the 1990s. Since the total number of patent applications had been increasing rapidly during those years, the share of novel patents in total

patent applications dropped all the way from 12% in 1980 (8% at the start of the 1990s) to 3% in 2007. After 2008, however, the number of novel patent applications doubled to almost 8,000 per year, and their share in total patent applications had accordingly recovered to 6%, the level last seen in the mid-1990s, by 2016.

[Figure 4 around here]

Comparing Figures 3 and 4, one cannot fail to notice that since the 1990s, the trend in the share of novel patents in the total patents filed by the U.S. patent assignees is very similar to the trend in the share of mega firms in novel patents. That is, as the share of novel patents in the total number of patents in the U.S. falls during the 1990s-first half of the 2000s, so does the share of mega firms in those novel patents, and as the share of novel patents in the total number of patents starts recovering in the late 2000s, so does the share of mega firms (if anything, the share of mega firms in novel patents had already recovered to the 1990 level by 2016, while the share of novel patents economy-wide is still below the 1990 level by about two percentage points).

Inasmuch as novel patents are a proxy for novel technologies, the above can be interpreted as suggestive evidence of a possibly turning tide in the U.S. innovation, to large extent driven by mega firms. We examine this in more detail in the next section.

4. New combinations, knowledge interdependence, mega firms, and VC-backed startups

4.1 New combinations and "hits"

Each new combination of previously not connected knowledge components adds a new "link" to the complex web of knowledge interdependence. As a result, we expect to see more connecting edges across different types of knowledge in the common stock of patented new knowledge as time goes by. Figure 5 shows that it is indeed the case. In this Figure we present the dynamics of knowledge interdependence measured by the average (valued) degree centrality of technological

components in the knowledge network,[13] using the number of patents that are co-assigned to different technology groups in a given year. More precisely, we calculate the (valued) degree centrality of each technological components normalized by the network size as $D_{i,t} = \frac{1}{N_t - 1}\sum_{j \neq i} P_{i,j,t}$, where N_t is the number of distinct CPC main groups in year t and $P_{i,j,t}$ is the number of patents that are assigned to groups i and j in year t (which may be zero). Intuitively, this measures the patent-weighted share of distinct technology groups that have been integrated into the focal technology at any given point in time.

[Figure 5 around here]

As mentioned (see Figure 4 above), the number of novel patents applied in each year was more or less constant over the 1980s and 1990s, reflected in generally gradual increase in the valued degree centrality during this period in Figure 5. Also, consistent with the big increase in the number of novel patents since the late 2000s in Figure 4, Figure 5 shows a sharp increase in knowledge interdependence in all CPC technology sections during that period.

Since Figure 5 shows the dynamics of the degree of knowledge interdependence, for this degree to increase as dramatically as it does after the mid-2000s, novel patents, establishing new combinations for the first time, must generate many follow-on patents that repeatedly utilize the same new combination. It turns out that almost half of all new combinations (48.4%, to be exact) do not generate any follow-on patents within the first five years after they were introduced by the first novel patent. Some novel patents, however, generate a very large number of follow-on patents and thus have a big impact on shaping entirely new technological trajectories. We follow the terminology in Pezzoni (2022) and call such impactful new combinations "hits." For our main

[13] The degree centrality of a node (i.e., a technological component) in the network is the same as its degree, which is simply a count of connections to other technological components.

analysis we define a "hit" as a new combination that generated the number of follow-on patents in the top one percentile of the distribution of follow-on patents within the first five years, although the findings below are robust to using other cutoffs, such as the 95th percentile.

Table 1 presents some summary statistics on the overall number of new combinations created over 10 years during two periods, 1991-2000 (including the dot.com boom era) and in 2007-2016 as well as how many of those were created by mega firms in our sample, by VC-backed startups and by those two types of firms together (that is, patents that were co-assigned to a mega firm and a VC-backed startup). The top panel shows all new combinations (with or without follow-on patents) while the bottom panel shows hits. As can be seen already from Figure 4, U.S. patenting entities generated twice as many new combinations in 2007-2016 compared to 1991-2000. More important from the point of view of this paper is that the share of mega firms in the total number of new combinations generated in the U.S. economy increased from just over 9% to almost 13% between these two periods.

The bottom panel focuses on the top 1% new combinations in terms of the number of follow-on patents, i.e., "hits." The share of mega firms in those hits is 12% in 1991-2000 but it increases to 21.2% in 2007-2016. We also see that VC-backed startups produced twice the share of hits compared to their share in all new combinations in both periods, and their share in total hits is similar to the share of mega firms. Thus, VC-backed startups generate a disproportionately large number of hits compared to their share in all novel combinations. It is worth noting that a number of mega firms in 2007-2016 are also in our sample of VC-backed startups in 1991-2000. In particular, this is true of all GAMAM firms with the exception of Amazon.[14]

[Table 1 around here]

[14] GAMAM firms refer to the Tech Giants including Google, Amazon, Meta (Facebook), Apple, and Microsoft.

We next examine how successful mega firms and VC-backed startups were in generating follow-on patents and how their success changed over time by means of a simple regression. The estimation equation is

$$y_t = \alpha + \beta_1 I_{\{mega\ firm\}t} + \beta_2 I_{\{mega\ firm\}t} X\{2007-2016\ period\} + \beta_3 I_{\{VC\}t} + \beta_4 I_{\{VC\}t} X\{2007-2016\ period\} + \delta_t + \varepsilon_t$$

where y_t is the outcome in year *t*, $I_{\{mega\ firm\}t}$ is the dummy equal to one if the novel combination was generated by a mega firm in year t and zero otherwise, $I_{\{VC\}t}$ is the dummy equal to one if the novel combination was generated by a VC-backed startup in year t and zero otherwise, δ_t are year fixed effects, and ε_t is the error term. The omitted category is all other patenting entities (that is, neither a mega firm nor a VC-backed startup). The baseline period is 1991-2000, as in Table 1. The outcome variables, in columns (1)-(3) in Table 2, respectively, are the IHS number of follow-on patents within the first five years after the new combination, the dummy equal to one if no follow-on patents within the first five years after the new combination, and the (logged) number of follow-on patents within the first five years after the new combination, conditional on at least one such patent.

[Table 2 around here]

The estimation results are presented in Table 2. New combinations by mega firms were less likely than other patenting entities to generate follow-on patents during the baseline period, but the opposite is true in the 2007-2016 period when the number of new combinations sharply increased. In particular, the coefficients in column (3) indicate that conditional on having at least one follow-on patent, mega firms had about 7% fewer follow-on patents than other entities in 1991-2000 but 9.7% more in 2007-2016. The probability of not having any follow-on patent was also higher by about 5% for mega firms in 1991-2000 but it was 3.5% lower than for other entities

in 2007-2016 (column (2)). As for the VC-backed startups, they have more follow-on patents than other entities and a lower probability of no follow-on patents, but the increase in the number of follow-on patents and the decrease in the probability of "failure" between 2007-2016 and 1991-2000 for VC-backed startups is significantly less pronounced than for mega firms. Thus, we see that mega firms not only increased the number of new combinations from the 1990s to the late 2000s-first half of 2010s, but they also had a large increase in the impact of those new combinations. The findings in this section thus present further suggestive evidence in favor of interpreting the recent increase in novel patents by mega firms as reflecting "true" innovation rather than just strategic behavior.

4.2 Changing technology combinations among hits and mega firms

We now turn our attention to another crucial aspect of the recent surge in novel patents—a big shift that is occurring in the direction of new combinations embodied in such patents, especially those that are hits in the sense of seeding whole new technological trajectories.

We already saw in Figure 5 above that the largest increase in (valued) degree centrality happens in CPC sections G and H (physics and electricity). These turn out to be the sections mostly associated with ICT-related innovations (Table A.2 in the appendix). A deeper examination reveals, however, that the recent surge in novel patents is associated not with innovation in ICTs themselves; instead, it is almost entirely driven by combining ICT with non-IC technologies. And mega firms are playing a key role in this process.

Recall that in the patent example in Figure 1 above, NIKE combined a technology component belonging to section G with two components belonging to section A for the first time, creating a new product that incorporated wireless communication technology into a garment (T-shirt). It turns out that this is not a unique case but rather a reflection of the general trend in which

mega firms, including mega firms in non-ICT industries have been playing an increasing role in generating innovations that apply ICT-related technologies to their main lines of business.[15]

In Figure 6 we decomposed the increase in valued degree centrality from Figure 5 into within- and between CPC sections and subsections. Especially in physics and electricity sections, the lion's share of new combinations in recent years are *across*, not *within* the same CPC section, even at the most aggregated, one-digit CPC classification level. This suggests that an increasing share of hits among new combinations spans technology components from more distant technological fields than before.

[Figure 6 around here]

To examine which exact technology components are being brought together (newly combined) in recent hits, we looked at all new combinations generated since 2006 and identified the top 0.1% of those in terms of the number of follow-on patents re-using the focal pair of technological components. There are 389 such new combinations and each of them had been (re-)used by at least 61 follow-on patents until 2020. Figure 7 shows the main groups in the CPC classification which were combined by those top hits for the first time. The following link: https://www.yuhengding.com/about/vis_most_successful_new_combinations provides the interactive animation of Figure 7 which can be used to follow the discussion below.

[Figure 7 around here]

The color in Figure 7 indicates different CPC sections. Yellow color in Figure 7 (and the interactive animation in the link above) denotes technology groups in the H (electricity) section, while green color denotes technology groups in the G (physics) section. The sizes of the nodes reflect its degree – the number of other nodes (i.e., technological components) connected to the

[15] NIKE is not among the top 50 firms by sales in Compustat overall, but it is among the top 50 patenting firms by sales in the Compustat data, so we count it as a mega firm for the purpose of this example.

focal one, while the thickness of the connecting lines reflects the number of patents using the pair of technological components. For example, the large yellow node at the center of Figure 7 is CPC main group H04W4 (services specially adapted for wireless communication networks; facilities therefor) and we can see that a lot of links connect it to nodes colored differently (meaning that it is being combined with technological components in various CPC sections). One of the thickest lines connects this node to orange-colored (CPC section B, performance operation and transportation) node B60W50 which is related to vehicle drive control and driver interface systems. To put it more simply, this means that we see a lot of follow-on patents combining these components to generate improvements in Advanced Driver Assistance Systems (ADAS) and self-driving cars. Another node, also connected to the H04W4 node is G06T11 (2-dimensional image generation)[16]. In its turn, this is connected to node A61B34 (computer-aided surgery), in section A (human necessities), pointing to another hit that combined ICT with an advance in health care.

In Table 3 we present systematic evidence showing a big change in the novel technological trajectories that happened in recent years (see also Figure A2 in the appendix). Specifically, we compare the type of technologies integrated by "hits" between the 1991-2000 and the 2007-2016 periods. We define a new combination to be within ICT ("ICT & ICT" in Table 3) if the new combination linked for the first time only the CPC main groups that are matched to ICT-related industries (see Section 2.4 above for how this match was constructed), to be between ICT and non-ICT ("ICT & non-ICT" in Table 3) if the new combination linked for the first time the CPC main groups that are matched to both ICT-related industries and non-ICT-related industries, and to be within non-ICT ("non-ICT & non-ICT" in Table 3) if the new combination linked for the first time the CPC main groups neither of which is matched to an ICT-related industry.

[16] It can be found in the interactive image by typing "G06T11" into the "Select node..." bar.

[Table 3 around here]

The difference between the 1990s and 2007-2016 we see in Table 3 is striking. In the 1990s, 62.7% of the top hits were new combinations combining technologies within ICT (57% among top hits produced by mega firms). In contrast, in 2007-2016, that is, during the recent resurgence of novel patents and new combinations, 53.2% of top hits (and 54.6% of top hits produced by mega firms) newly combined ICT and non-ICT components, with those combining technologies within ICT accounting for just about 10% of all hits.

Robert Solow once famously quipped that "we see computers everywhere except in the productivity statistics." In a similar vein, it has been some time since IC technologies were recognized as general-purpose technologies (GPT) and their potential to generate a new industrial revolution has also been noted (e.g., Hsieh and Rossi-Hansberg, 2021). Our findings here suggest that IC technologies may finally be indeed "coming of age" as GPT in recent years, as they are now being relatively less combined between themselves and instead increasingly used in a broad array of "complementary innovations" that combine them with non-IC technological components.

4.3 What is happening to knowledge diffusion?

It has been argued that one reason for declining business dynamism in the U.S. in recent decades may be a slow-down in new knowledge diffusion from leading to laggard firms (e.g., Akcigit and Ates, forthcoming). The extant analysis has relied on indirect measure or inference, however.

New combinations represent new knowledge, and follow-on patents represent the diffusion of this knowledge. Hence, one way to examine directly if the diffusion of new knowledge from leading to laggard firms is indeed slowing down is to look at the dynamics of the follow-on patents that are assigned to the same firm as the original new combination assignee or whether they are assigned to different firms. If the share of follow-on patents that are assigned to the same firm that

came up with the new combination is increasing over time, it can perhaps be interpreted as a slow-down in new knowledge diffusion.

In Figure 8 we present the dynamics of the number of follow-on patents and the share of follow-on patents that are assigned to the focal assignee (the entity that came up with the original new combination) over the first five years after the new combination was generated. There is no particular time trend in this Figure, while the number of follow-on patents can be seen to be increasing in more recent years as the number of novel patents, as we saw in Figure 4 above, surged.

[Figure 8 around here]

In Table 4 we present the results of a regression estimation where the share of follow-on patents over the first five years that are assigned to the focal assignee is the dependent variable and the independent variables are time trend in column (1), time trend and dummies equal to one for mega firms and VC-backed startups in column (2), and all the above, plus the interaction terms between time trend and those two dummies in column (3).[17] We repeat the same exercise for the subsample of top 1% of new combinations in terms of follow-on patents they generated over the first five years in columns (4)-(6).

[Table 4 around here]

Interestingly, the estimation results in columns (2) and (4) indicate that mega firms have more follow-on patents that are assigned to entities other than themselves compared to the baseline category, while the opposite is true of VC-backed startups. Including interaction with the time trend in column (3) reveals that the share of follow-on patents assigned to the focal assignee if the new combination was generated by a mega firm or a VC-backed startup is increasing over time in all observations but it is decreasing over time for mega firms if we only look at top hits.

[17] Note that new combinations without any follow-on patents are excluded from these regressions since the dependent variable is the *share* of self-use among all follow-on patents.

Thus, at least at this level of analysis we do not find much evidence to support the declining diffusion of knowledge from new combinations over time, neither economy-wide nor specifically for mega firms. There is a limitation to this analysis, however, as the first five years after a new combination may be too short of an observation period. Indeed, Pezzoni et al. (2022) show that diffusion curves of new technological trajectories are S-shaped, with considerable variation in diffusion time. On average, technological impact only reaches the takeoff stage (10% of the impact measured over 20 years) at the five-year mark, while the midpoint of the diffusion curve is not reached until about 12-year mark (Pezzoni et al., 2022, Table 2). We do not yet have enough of an observation period to redo the estimations over a longer time period for most recent new combinations, but we did look at the 20-year diffusion curves following Pezzoni et al. (2022) methodology for new combinations that had been generated prior to 2001 and for which we do observe their diffusion patterns for 20 years. The findings are presented in Figure 9 where in Panel A we present the average cumulative number of follow-on patents by assignees other than the focal firm for all new combinations generated by mega firms, VC-backed startups, and other firms (patenting entities), while in Panel B we present the same findings for the subsample of hits (new combinations with follow-on patents in the top one percentile).

Note, once again, that the diffusion curves in Figure 9 exclude the follow-on patents by the focal assignee (the one who created the new combination), hence, we are looking at the diffusion of knowledge outside of the focal firm's boundary. From the evidence in Panel A, we see that looking at all new technological trajectories, the diffusion from VC-backed startups take off at the fastest pace and they also have most follow-on patents after 20 years. Mega firms, however, are closely behind, while new technological trajectories generated by other assignees take longer to take off and the number of follow-on patents is lower than for VC-backed startups or mega firms.

[Figure 9 around here]

Panel B shows that the difference between the three types of the firms is much less pronounced among hits (new technological trajectories in the top one percentile by the number of follow on patents in 20 years). This is not surprising, of course, as we are selecting on hits. What is interesting, however, is that mega firms now eventually outpace even VC-backed startups in the number of follow-on patents by other assignees: on average, they have close to 600 follow-on patents after 20 years as opposed to about 450 for VC-backed startups and 400 for all other assignees. Once again, we find no evidence here that economically successful mega firms generate less knowledge spillovers to others compared to other firms. It remains to be seen how this picture will look after 20 years with respect to new technological trajectories which started after the recent surge in novel patents, but estimation results in Table 4, column (6) suggest that at least based on the first five years of observations, mega firms may be contributing to even more knowledge spillovers outside of their boundaries than they had been doing before.

5. Novel patents and firm outcomes

Finally, we briefly examine the relationship between novel patents and firm performance. Our goal at this stage is simply to see if there is any empirical association between creating new combinations and relative firm performance, without claiming any causal relationship in any direction.

To examine this association we estimate the following regression model:

$$\mathrm{Y}_{ijt} = \alpha + \beta \text{Novel Patents}_{ijt-s} + X_{ijt-s}\,\gamma + \delta_i + \delta_{jt} + \varepsilon_{ijt},$$

where Y_{ijt} is the outcome variable measured in two ways; as (logged) employment and as logged total factor productivity measured in terms of revenue (TFPR). $\text{Novel Patents}_{ijt-s}$ is the total number of novel patents by firm *i* in industry j applied in year $t-s$, and X_{ijt-s} is a vector of firm-

level controls where we include (logged) real sales and (logged) total number of patent applications (to control for the overall patenting propensity of the firm) as the baseline.[18] δ_i represents firm fixed effects, and δ_{jt} represents industry-year fixed effects. We use lagged independent variables to allow for the possible delay in the "impact" of novel patents.

Table 5 shows the estimation results where the depended variable is firm size measured by employment. For the one-year lag (s = 1), the relationship is statistically insignificant and small once we control for firm size in terms of both real output and its total patent stock. For $s > 1$, however, the relationship becomes positive and statistically significant. For brevity, we only show the results for $s = 5$, but the findings are similar for different lags s, $1 < s < 5$.

[Table 5 around here]

Table 6 presents the estimation results with firm TFPR as the dependent variable. Similar to the above, we find a positive effect of novel patents on firm revenue productivity across different specifications. As before, the results stay robust across different lag values s and different combinations of the firm controls.

[Table 6 around here]

As mentioned, these estimation results cannot be interpreted as indicating any causality, even though we control for firm fixed effects. For example, both the increase in novel patenting and the subsequent increase in TFPR can both be due to the firm adopting a different growth strategy at some point in time. But the fact that novel patents and improved performance at the firm level are positively correlated even after controlling for overall patenting behavior does suggest, in our view, that novel patents may be an important channel of innovation. We plan to examine this further going forward.

[18] We explored different combinations of the firm controls, and the results stay robust.

6. Discussion

The share of economic activities accounted for by mega firms has dramatically increased over the past several decades and their innovation behavior has profound implications for economic growth, technological progress, and the appropriate policy response. In this regard, it is important to understand whether mega firms are occupying the technology frontier and causing a slowdown in knowledge diffusion using patents and intellectual property regulations. While mega firms may be increasingly protecting their technological superiority using patents in certain dimensions, we provide new evidence that they may be also leading the technological experimentation by introducing new technology combinations that enable other firms to conduct follow-on innovation.

We find that the pace of novel innovation had declined over several decades until 2008, followed by a strong rebound since then and mega firms played a large role in this trend reversal. Mega firms are increasingly combining ICT components with non-ICT components in their experimentation and the extent to which these new combinations generate follow-on innovation by other firms is higher for mega firms even compared to VC-backed startups, which are firms often considered to be at the heart of technological experimentation.

If it is true that mega firms are predominantly stifling innovation and slowing down knowledge diffusion, there could be a scope for regulatory intervention to be introduced. If, however, mega firms are the key actors generating novel technologies, then such an approach may backfire. We believe better understanding the strengths of these countervailing forces would be an important research agenda in this debate.

References

Akcigit, U., & Ates, S. (2023). What Happened to U.S. Business Dynamism? *Journal of Political Economy*. (forthcoming).

Autor, D., Dorn, D., Hanson, G. H., Pisano, G., & Shu, P. (2020a). Foreign Competition and Domestic Innovation: Evidence from US Patents. *American Economic Review: Insights*, *2*(3), 357–374.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., & Van Reenen, J. (2020b). The Fall of the Labor Share and the Rise of Superstar Firms. *The Quarterly Journal of Economics*, *135*(2), 645–709.

Bernstein, S., Giroud, X., & Townsend, R. R. (2016). The Impact of Venture Capital Monitoring. *The Journal of Finance*, *71*(4), 1591–1622.

Covarrubias, M., Gutiérrez, G., & Philippon, T. (2020). From Good to Bad Concentration? US Industries over the Past 30 Years. *NBER Macroeconomics Annual*, *34*, 1–46.

De Loecker, J., Eeckhout, J., & Unger, G. (2020). The Rise of Market Power and the Macroeconomic Implications. *The Quarterly Journal of Economics*, *135*(2), 561–644.

Epicoco, M., Jaoul-Grammare, M., & Plunket, A. (2022). Radical technologies, recombinant novelty and productivity growth: A cliometric approach. *Journal of Evolutionary Economics*, *32*(2), 673–711.

Fleming, L., Mingo, S., & Chen, D. (2007). Collaborative Brokerage, Generative Creativity, and Creative Success. *Administrative Science Quarterly*, *52*(3), 443–475.

Foster, L. S., Haltiwanger, J. C., & Tuttle, C. (2022). *Rising Markups or Changing Technology?* (No. w30491). National Bureau of Economic Research.

Goldschlag, N., Lybbert, T. J., & Zolas, N. J. (2020). Tracking the technological composition of industries with algorithmic patent concordances. *Economics of Innovation and New Technology*, *29*(6), 582–602.

Goldschlag, N., & Miranda, J. (2020). Business dynamics statistics of High Tech industries. *Journal of Economics & Management Strategy*, *29*(1), 3–30.

González-Uribe, J. (2020). Exchanges of innovation resources inside venture capital portfolios. *Journal of Financial Economics*, *135*(1), 144–168.

Gutiérrez, G., & Philippon, T. (2019). *The Failure of Free Entry* (No. w26001). National Bureau of Economic Research.

Hsieh, C.-T., & Rossi-Hansberg, E. (2021). *The Industrial Revolution in Services*. Working Paper.

Hunt, R., & Bessen, J. (2004). The software patent experiment. *Business Review*, *2004*, 22–32.

Klepper, S. (2015). *Experimental Capitalism: The Nanoeconomics of American High-Tech Industries*. Princeton University Press.

Kwon, S. Y., Ma, Y., & Zimmermann, K. (2023). *100 Years of Rising Corporate Concentration* (SSRN Scholarly Paper No. 4362319).

Ma, S. (2020). The Life Cycle of Corporate Venture Capital. *The Review of Financial Studies*, *33*(1), 358–394.

Noel, M., & Schankerman, M. (2013). Strategic Patenting and Software Innovation. *The Journal of Industrial Economics*, *61*(3), 481–520.

Pezzoni, M., Veugelers, R., & Visentin, F. (2022). How fast is this novel technology going to be a hit? Antecedents predicting follow-on inventions. *Research Policy*, *51*(3), 104454.

Strumsky, D., & Lobo, J. (2015). Identifying the sources of technological novelty in the process of invention. *Research Policy*, *44*(8), 1445–1461.

Torrisi, S., Gambardella, A., Giuri, P., Harhoff, D., Hoisl, K., & Mariani, M. (2016). Used, blocking and sleeping patents: Empirical evidence from a large-scale inventor survey. *Research Policy*, *45*(7), 1374–1385.

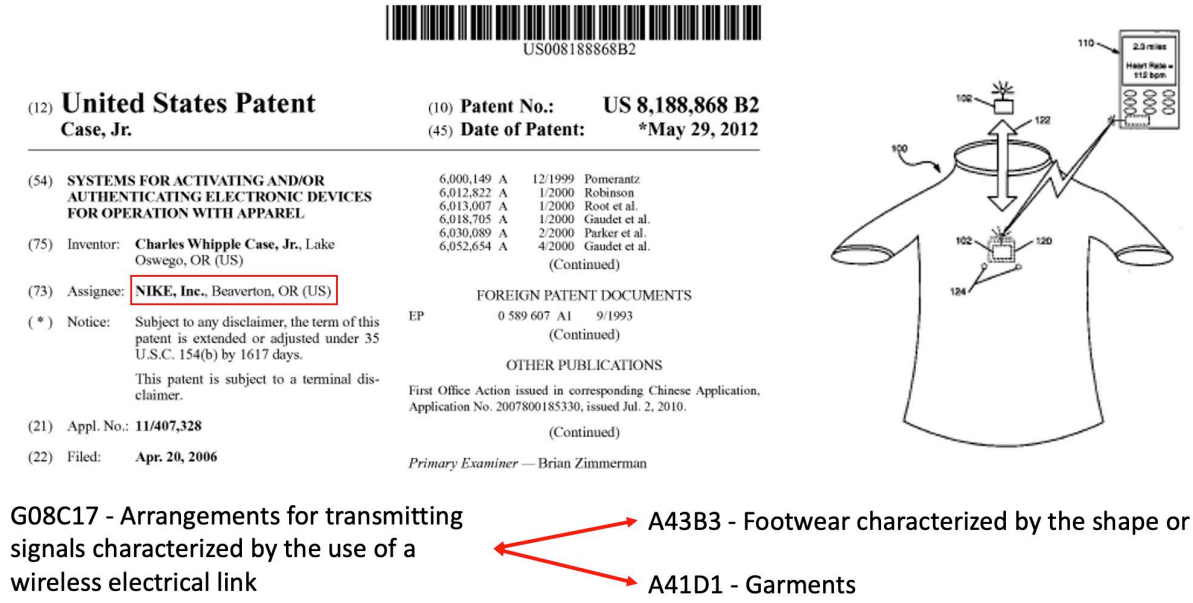
Verhoeven, D., Bakker, J., & Veugelers, R. (2016). Measuring technological novelty with patent-based indicators. *Research Policy*, *45*(3), 707–723.

Watzinger, M., & Schnitzer, M. (2022). The Breakup of the Bell System and its Impact on US Innovation. *Rationality and Competition Discussion Paper Series*, Article 341.

Yeh, C., Macaluso, C., & Hershbein, B. (2022). Monopsony in the US Labor Market. *American Economic Review*, *112*(7), 2099–2138.

Figures and Tables

Figure 1. A novel patent by NIKE combining ICT-related and non-related components.



US008188868B2

(12) **United States Patent**
Case, Jr.

(10) **Patent No.: US 8,188,868 B2**
(45) **Date of Patent: *May 29, 2012**

(54) **SYSTEMS FOR ACTIVATING AND/OR AUTHENTICATING ELECTRONIC DEVICES FOR OPERATION WITH APPAREL**

(75) Inventor: **Charles Whipple Case, Jr.**, Lake Oswego, OR (US)

(73) Assignee: **NIKE, Inc.**, Beaverton, OR (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 1617 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **11/407,328**

(22) Filed: **Apr. 20, 2006**

6,000,149 A 12/1999 Pomerantz
6,012,822 A 1/2000 Robinson
6,013,007 A 1/2000 Root et al.
6,018,705 A 1/2000 Gaudet et al.
6,030,089 A 2/2000 Parker et al.
6,052,654 A 4/2000 Gaudet et al.
(Continued)

FOREIGN PATENT DOCUMENTS

EP 0 589 607 A1 9/1993
(Continued)

OTHER PUBLICATIONS

First Office Action issued in corresponding Chinese Application, Application No. 2007800185330, issued Jul. 2, 2010.
(Continued)

Primary Examiner — Brian Zimmerman

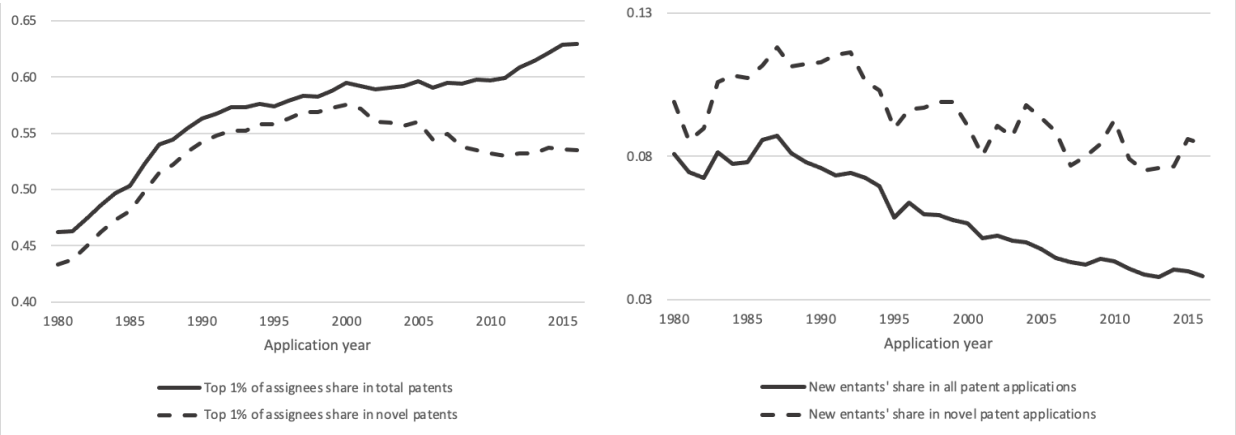
G08C17 - Arrangements for transmitting signals characterized by the use of a wireless electrical link

A43B3 - Footwear characterized by the shape or the use

A41D1 - Garments

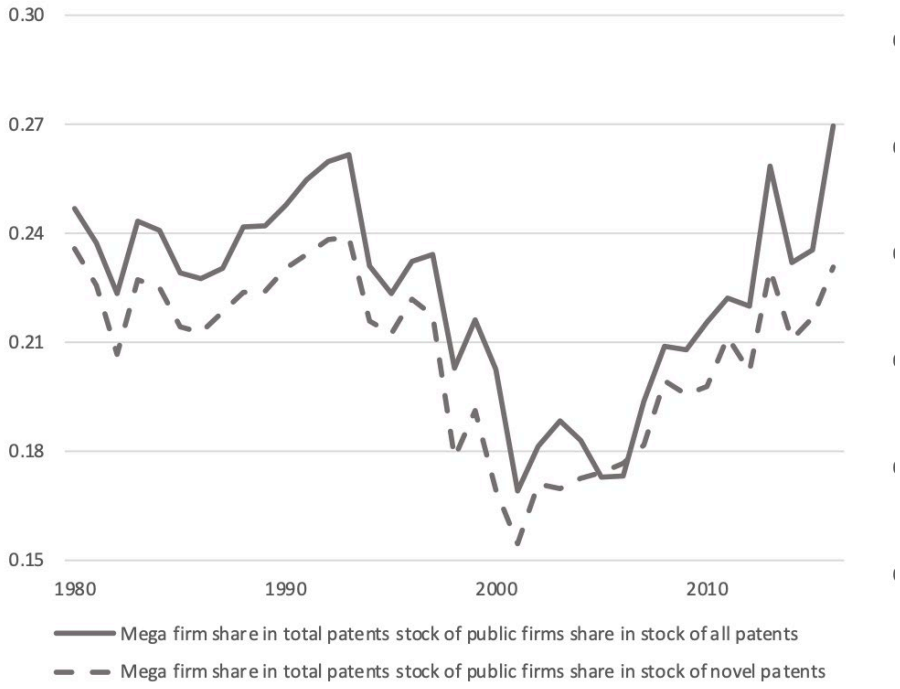
Source: USPTO data.

Figure 2. Shares of top 1% of patent assignees and new entrants in total and novel patents



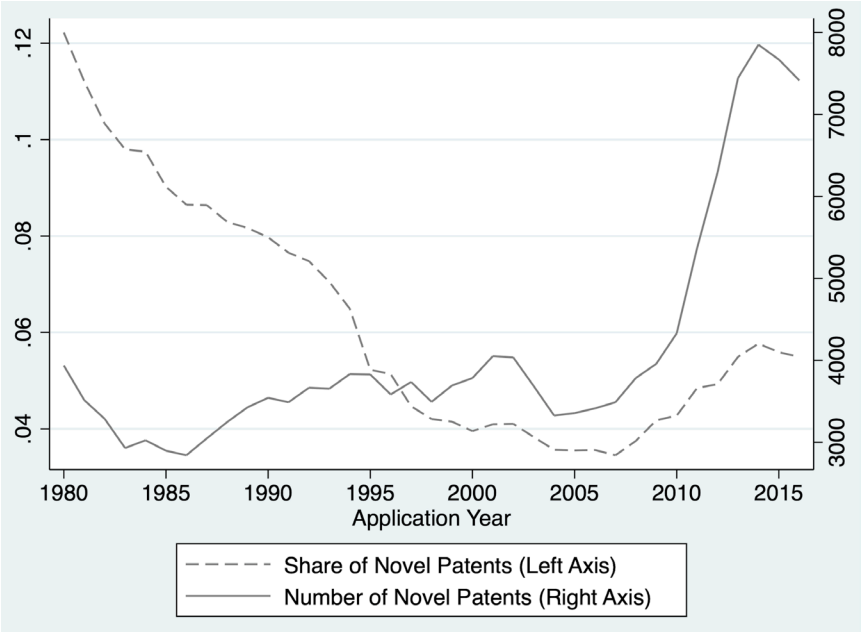
Source: Authors' own calculation using USPTO data. "Total patents" is the cumulative number of all patent applciatoins up to application year *t*. "Novel patents" are those that have at least one pair of CPC main groups co-assigned to the patent for the first time. See the main text for details. "New entrants" are patent assignees who are assigned a patent for the first time in application year *t*.

Figure 3. Share of mega firms in cumulative number of patents by all public firms



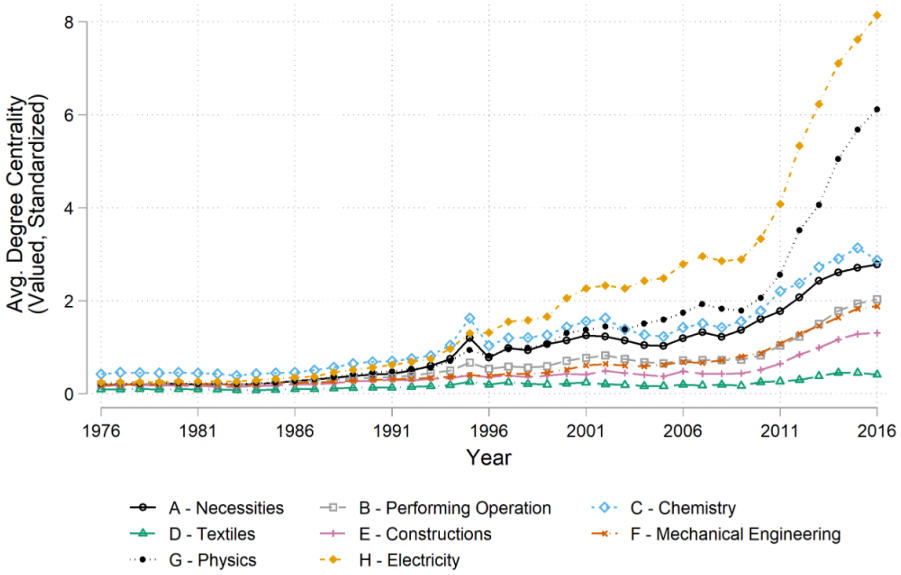
Source: Authors' own calculation using the USPTO data matched with Compustat data.

Figure 4. The number of novel patents and their share in total patent applications among the U.S. patent assignees.



Source: Authors' own calculation using USPTO data.

Figure 5. Time trend of economy-wide knowledge interdependence



Source: Authors' calculation using USPTO data. See the main text for the definition of valued degree centrality.

Figure 6. Decomoposition of valued degree centrality into within- and between sections

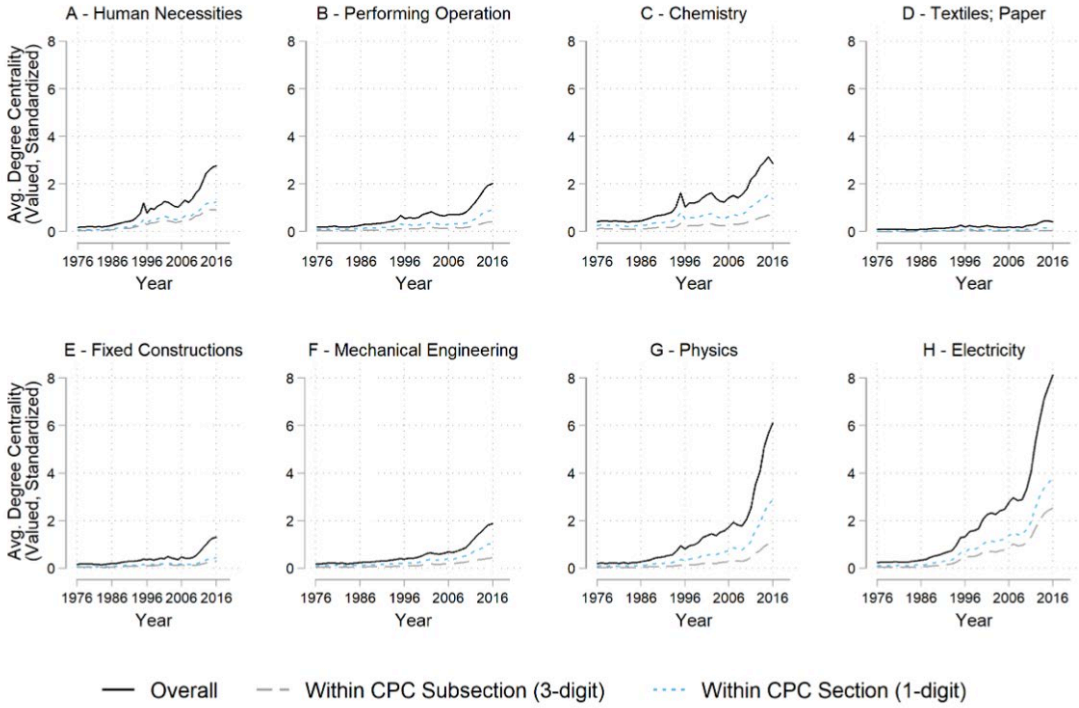
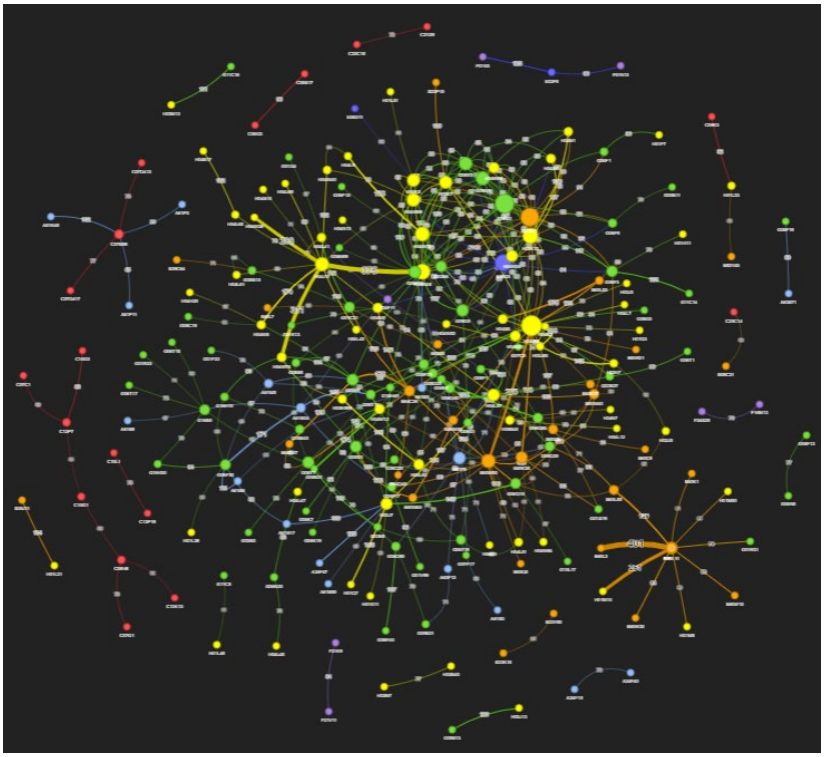
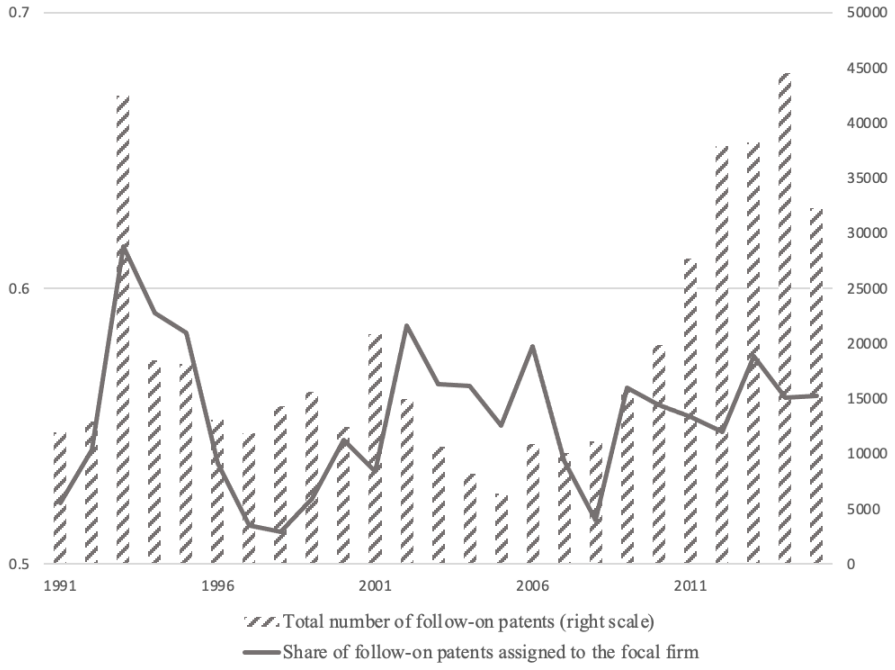


Figure 7. Most frequently used new combinations invented after 2006.



Source: created by Yuheng Ding using USPTO data.
Online interactive version: https://www.yuhengding.com/about/vis_most_successful_new_combinations

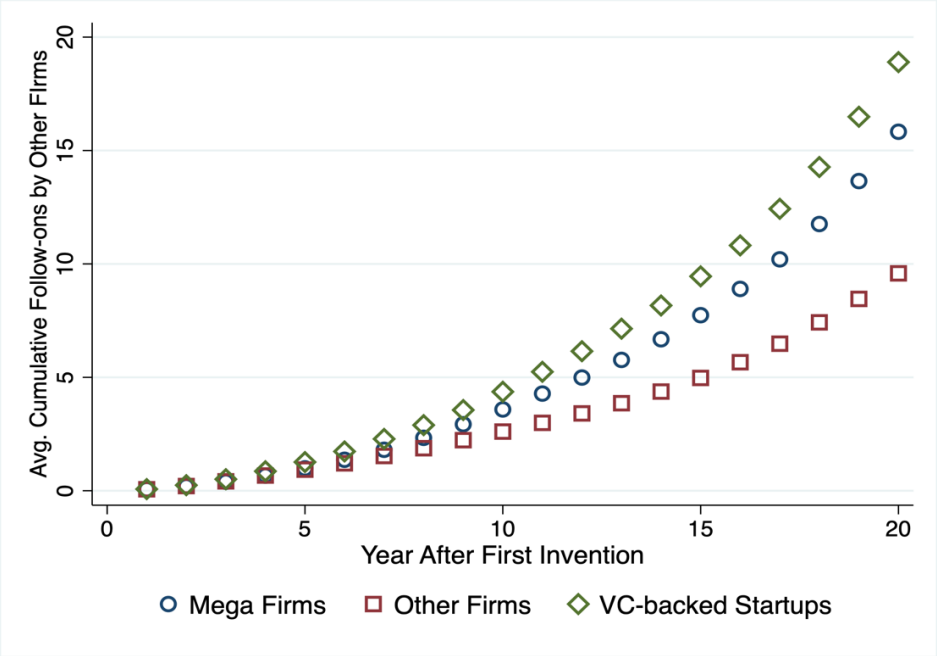
Figure 8. Number of follow-on patents and share assigned to the focal assignee



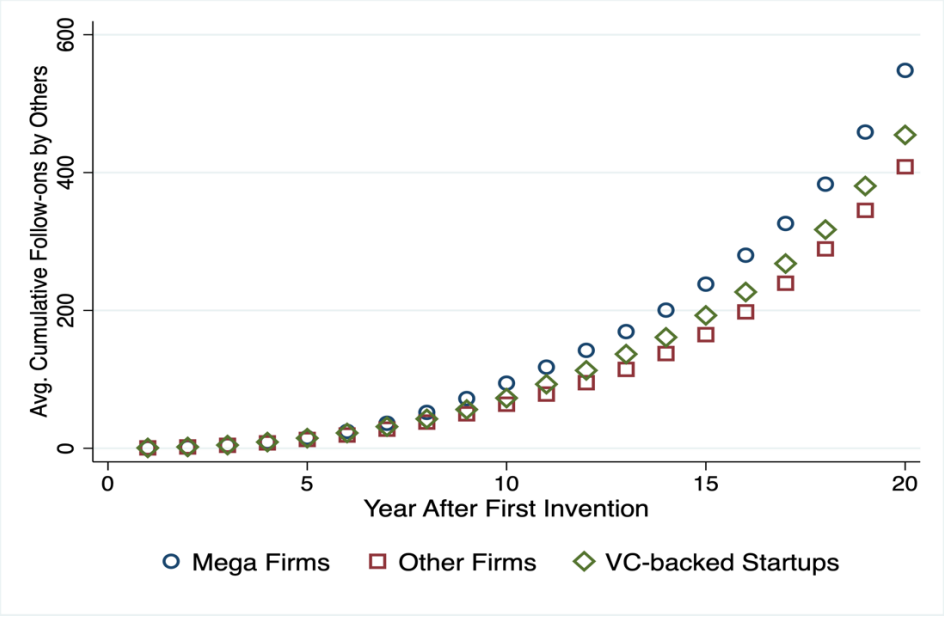
Source: Authors' own calculation using USPTO data.

Figure 9. Twenty-year diffusion curves to other assignees by firm types

Panel A. All new technological trajectories



Panel B. Top 1% hits



Source: Authors' calculation using USPTO matched with Compustat and VentureXpert data. Follow-on patents for all new combinations produced before 2000, for which we have 20 years of follow-on observations.

Table 1. All and top 1% successful new combinations by firm type: 1991-2000 and 2007-2016

			Mega firms	VC-backed startups	VC-backed startups & mega firms	Total
All new combinations	1991-2000	Number	7,215	5,695	292	77,888
		Percentage	9.3%	7.3%	0.4%	100.0%
	2007-2016	Number	19,531	13,064	266	150,943
		Percentage	12.9%	8.7%	0.2%	100.0%
Top 1% new combinations	1991-2000	Number	95	113	32	791
		Percentage	12.0%	14.3%	4.0%	100.0%
	2007-2016	Number	345	323	11	1,631
		Percentage	21.2%	19.8%	0.7%	100.0%

Source: Authors' calculation using USPTO matched with Compustat and VentureXpert data.

Table 2. Follow-on patents on new combinations by mega firms and VC-backed startups

DV:	(1) IHS # follow-on patents	(2) "Failed"	(3) Logged # follow-on patents
VARIABLES			
Dummy equal to one if mega firm	-0.117***	0.048***	-0.072***
	(0.012)	(0.006)	(0.014)
Mega firm X 2007-2016 period	0.202***	-0.073***	0.169***
	(0.014)	(0.007)	(0.017)
Dummy equal to one if VC-backed startup	0.226***	-0.072***	0.190***
	(0.013)	(0.007)	(0.014)
VC-backed startup X 2007-2016 period	0.094***	-0.048***	0.051***
	(0.015)	(0.008)	(0.017)
Constant	0.814***	0.462***	0.717***
	(0.002)	(0.001)	(0.003)
Year FE	Included	Included	Included
Observations	228,831	228,831	125,342
R-squared	0.031	0.024	0.027

Note: Estimation method: OLS, absorbing year fixed effects. Robust standard errors in parentheses. *** p < 0.01. Follow-on patents with first five years after application year. "Failed" means zero follow-on patents. IHS is inverse-hyperbolic sine transformation: $y = ln\left(x + \sqrt{x^2 + 1}\right)$.

Table 3. Combination types of 1% successful new combinations: 1991-2000 and 2007-2016

		Period			
		1991-2000		2007-2016	
Mega firms	ICT & ICT	35	57.4%	34	9.6%
	ICT & non-ICT	17	27.9%	189	53.2%
	non-ICT & non-ICT	9	14.8%	132	37.2%
	Total	61	100.0%	355	100.0%
All assignees	ICT & ICT	298	62.7%	194	11.9%
	ICT & non-ICT	113	23.8%	891	54.6%
	non-ICT & non-ICT	64	13.5%	546	33.5%
	Total	475	100.0%	1631	100.0%

Source: Authors' calculation using USPTO and Compustat data.

Table 4.

Share of follow-on patents over the first five years assigned to the focal assignee (1991-2016)

	(1)	(2)	(3)	(4)	(5)	(6)
DV:	Share of follow-on patents assigned to the focal assignee: All new combinations			Share of follow-on patents assigned to the focal assignee: Top 1% new combinations		
VARIABLES						
Time	0.000	0.000	-0.000**	0.017***	0.016***	0.012***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Dummy equal to one if mega firm		-0.024***	-0.061***		-0.072***	0.008
		(0.004)	(0.009)		(0.016)	(0.027)
Mega firm X time			0.002***			-0.004**
			(0.000)			(0.002)
Dummy equal to one if VC-backed startup		0.047***	-0.019**		0.196***	-0.177***
		(0.004)	(0.009)		(0.017)	(0.035)
VC-backed startup X time			0.004***			0.023***
			(0.001)			(0.002)
Constant	0.553***	0.550***	0.560***	0.022*	0.004	0.070***
	(0.003)	(0.003)	(0.003)	(0.012)	(0.013)	(0.014)
Observations	142,733	142,733	142,733	2,660	2,660	2,660
R-squared	0.000	0.001	0.002	0.136	0.195	0.237

Note: Estimation method: OLS. Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table 5. Novel patents and firm employment

	(1)	(2)	(3)	(4)	(5)	(6)
	Logged employment					
Total novel patents$_{t-1}$	0.117***	0.011***	0.002			
	(0.017)	(0.003)	(0.001)			
Total novel patents$_{t-5}$				0.139***	0.020***	0.010***
				(0.021)	(0.003)	(0.002)
Logged sales$_{t-1}$			0.399***			
			(0.015)			
Logged sales$_{t-5}$						0.218***
						(0.018)
Logged total patents$_{t-1}$			0.098***			
			(0.007)			
Logged total patents$_{t-5}$						0.069***
						(0.009)
Constant	0.401***	0.513***	-0.141***	0.544***	0.693***	0.474***
	(0.045)	(0.005)	(0.023)	(0.051)	(0.006)	(0.032)
Fixed effects	none	Firm & industry-year		none	Firm & industry-year	
Observations	38,488	34,067	32,765	30,862	26,805	22,745
adj. within R2	0.099	0.005	0.398	0.122	0.010	0.143

Note: Estimation method: OLS. Robust standard errors clustered at the firm level in parentheses. *** p < 0.01.

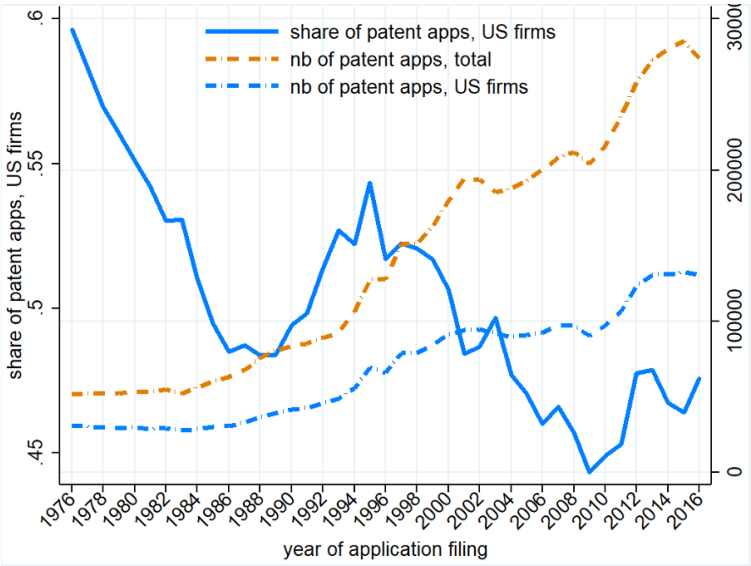
Table 6. Novel patents and TFPR

	(1)	(2)	(3)	(4)	(5)	(6)
	Logged TFPR					
Total novel patents$_{t-1}$	0.032***	0.005**	0.003*			
	(0.007)	(0.002)	(0.002)			
Total novel patents$_{t-5}$				0.032***	0.004***	0.003**
				(0.007)	(0.001)	(0.001)
Logged sales$_{t-1}$			0.152***			
			(0.015)			
Logged sales$_{t-5}$						0.007
						(0.016)
Logged total patents$_{t-1}$			-0.024***			
			(0.007)			
Logged total patents$_{t-5}$						-0.007
						(0.008)
Constant	-1.451***	-1.461***	-1.740***	-1.433***	-1.444***	-1.423***
	(0.023)	(0.003)	(0.032)	(0.025)	(0.003)	(0.035)
Fixed effects	none	Firm & industry-year		none	Firm & industry-year	
Observations	29,243	23,325	23,316	24,105	18,871	16,867
adj. within R2	0.0209	0.009	0.030	0.022	0.001	0.000

Note: Estimation method: OLS. Robust standard errors clustered at the firm level in parentheses. *** p < 0.01.

Appendix

Figure A1. UPSTO-Compustat Match Rates



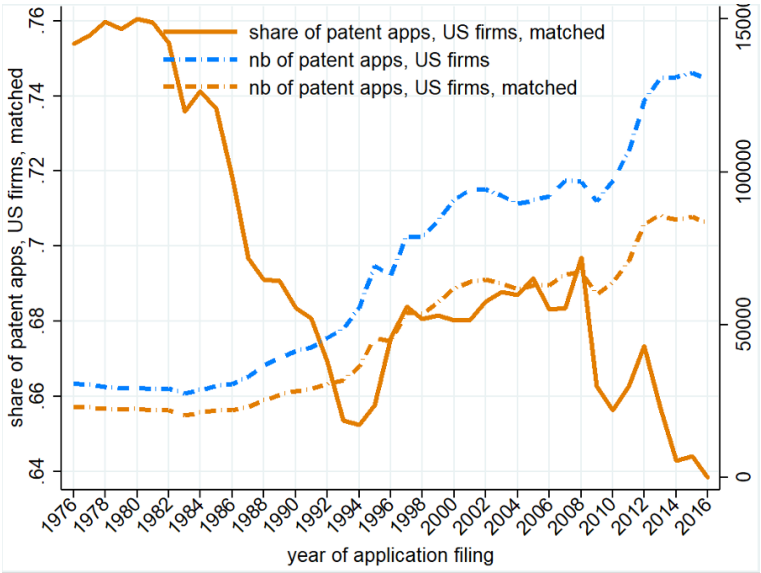
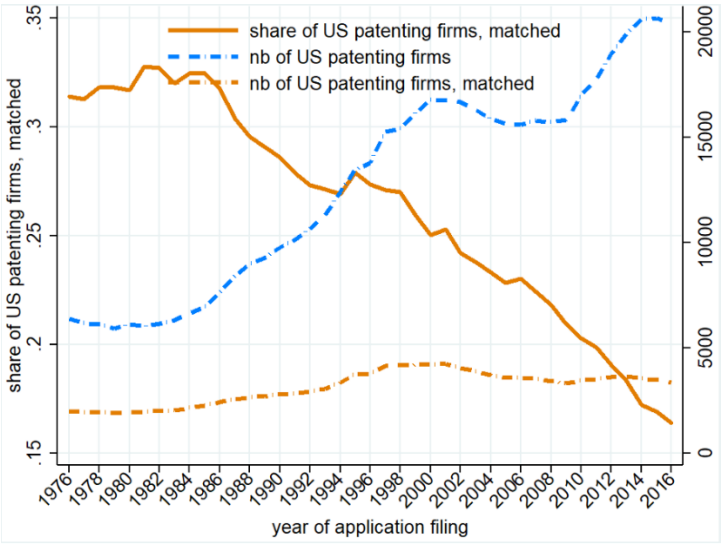
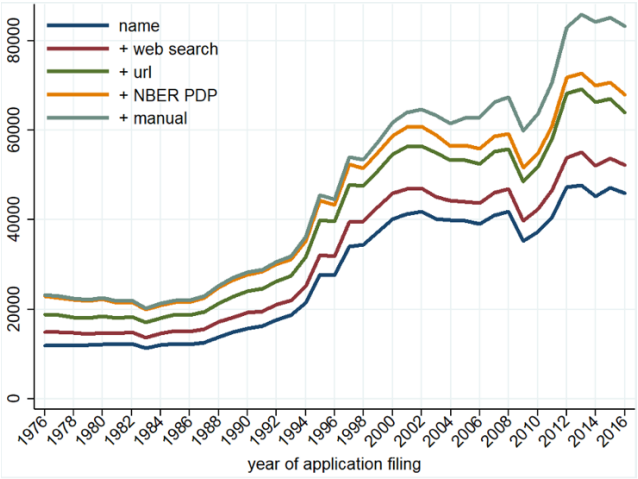


Figure A1. UPSTO-Compustat Match Rates (continued)





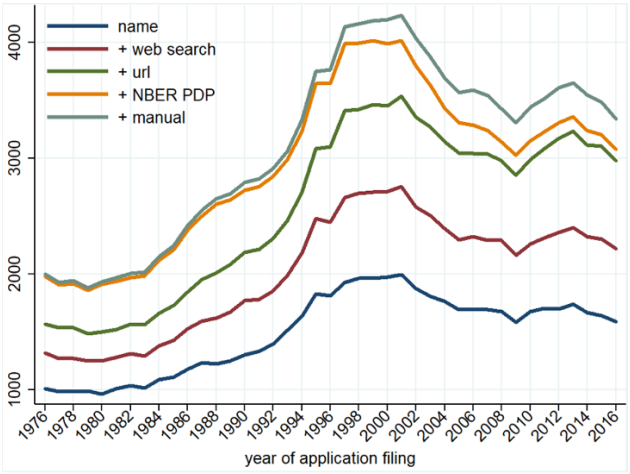


Figure A2.

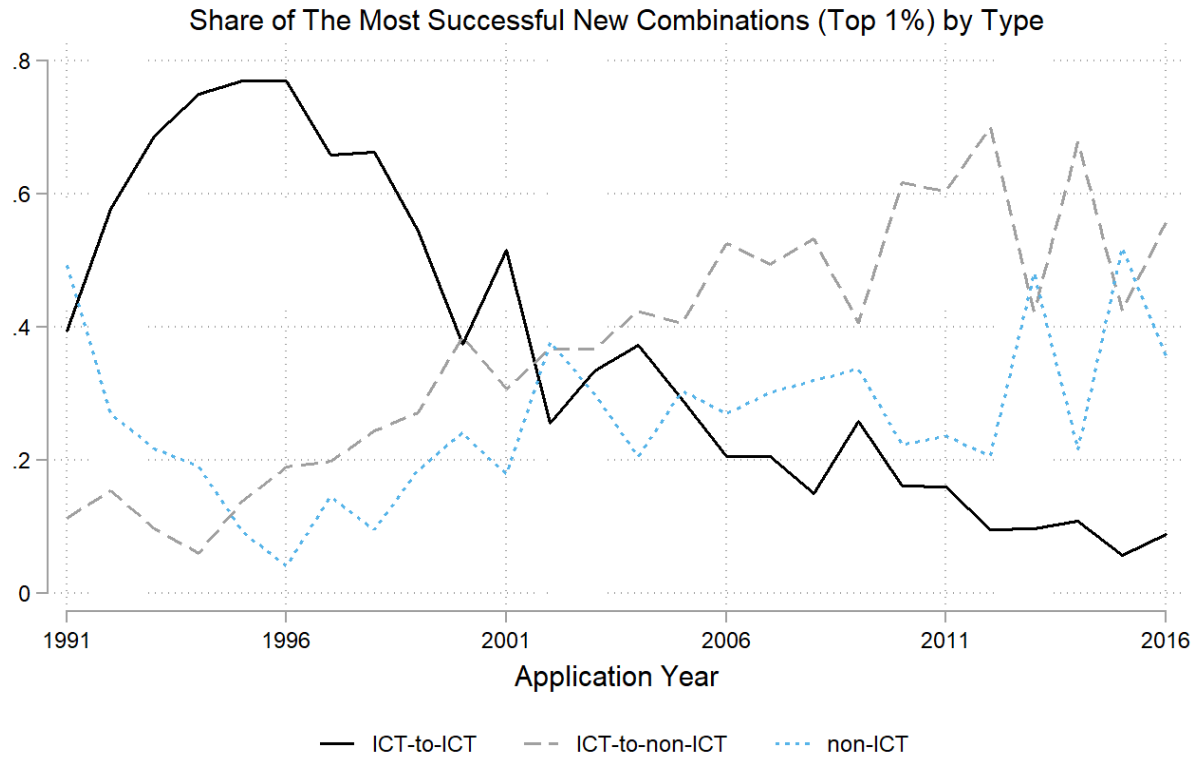


Table A1. List of ICT Industries

NAICS	Industry Description
3341	Computer and Peripheral Equipment Manufacturing
3342	Communications Equipment Manufacturing
3344	Semiconductor and Other Electronic Component Manufacturing
3345	Navigational, Measuring, Electromedical, and Control Instruments Manufacturing
5112	Software Publishers
5171	Wired Telecommunications Carriers
5179	Other Telecommunications
5182	Data Processing, Hosting, and Related Services
5191	Other Information Services
5415	Computer Systems Design and Related Services

Source: Table lists 4-digit 2007 NAICS industries that are identified as ICT industries based on Goldschlag and Miranda (2019)

Table A2. List of ICT Technologies

CPC Subclass	Description
A61B	DIAGNOSIS; SURGERY; IDENTIFICATION
B41B	MACHINES OR ACCESSORIES FOR MAKING, SETTING, OR DISTRIBUTING TYPE; TYPE; PHOTOGRAPHIC OR PHOTOELECTRIC COMPOSING DEVICES
B81B	MICROSTRUCTURAL DEVICES OR SYSTEMS, e.g. MICROMECHANICAL DEVICES
B81C	PROCESSES OR APPARATUS SPECIALLY ADAPTED FOR THE MANUFACTURE OR TREATMENT OF MICROSTRUCTURAL DEVICES OR SYSTEMS
F02D	CONTROLLING COMBUSTION ENGINES
G01C	MEASURING DISTANCES, LEVELS OR BEARINGS; SURVEYING; NAVIGATION; GYROSCOPIC INSTRUMENTS; PHOTOGRAMMETRY OR VIDEOGRAMMETRY
G01F	MEASURING VOLUME, VOLUME FLOW, MASS FLOW OR LIQUID LEVEL; METERING BY VOLUME
G01R	MEASURING ELECTRIC VARIABLES; MEASURING MAGNETIC VARIABLES
G01S	RADIO DIRECTION-FINDING; RADIO NAVIGATION; DETERMINING DISTANCE OR VELOCITY BY USE OF RADIO WAVES; etc.
G01V	GEOPHYSICS; GRAVITATIONAL MEASUREMENTS; DETECTING MASSES OR OBJECTS
G04B	MECHANICALLY-DRIVEN CLOCKS OR WATCHES; MECHANICAL PARTS OF CLOCKS OR WATCHES IN GENERAL; TIME PIECES USING THE POSITION OF THE SUN, MOON OR STARS
G04D	APPARATUS OR TOOLS SPECIALLY DESIGNED FOR MAKING OR MAINTAINING CLOCKS OR WATCHES
G05B	CONTROL OR REGULATING SYSTEMS IN GENERAL; FUNCTIONAL ELEMENTS OF SUCH SYSTEMS; MONITORING OR TESTING ARRANGEMENTS FOR SUCH SYSTEMS OR ELEMENTS
G06F	ELECTRIC DIGITAL DATA PROCESSING
G06K	RECOGNITION OF DATA; PRESENTATION OF DATA; RECORD CARRIERS; HANDLING RECORD CARRIERS
G06Q	DATA PROCESSING SYSTEMS OR METHODS, SPECIALLY ADAPTED FOR ADMINISTRATIVE, COMMERCIAL, FINANCIAL, MANAGERIAL, SUPERVISORY OR FORECASTING PURPOSES
G06T	IMAGE DATA PROCESSING OR GENERATION, IN GENERAL
G09G	ARRANGEMENTS OR CIRCUITS FOR CONTROL OF INDICATING DEVICES USING STATIC MEANS TO PRESENT VARIABLE INFORMATION
G11B	INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER
G11C	STATIC STORES
G21H	OBTAINING ENERGY FROM RADIOACTIVE SOURCES; APPLICATIONS OF RADIATION FROM RADIOACTIVE SOURCES, NOT OTHERWISE PROVIDED FOR; UTILISING COSMIC RADIATION
H01L	SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR

H01Q	ANTENNAS, i.e. RADIO AERIALS
H01R	ELECTRICALLY-CONDUCTIVE CONNECTIONS; STRUCTURAL ASSOCIATIONS OF A PLURALITY OF MUTUALLY-INSULATED ELECTRICAL CONNECTING ELEMENTS; etc.
H01S	DEVICES USING THE PROCESS OF LIGHT AMPLIFICATION BY STIMULATED EMISSION OF RADIATION [LASER] TO AMPLIFY OR GENERATE LIGHT; etc.
H02P	CONTROL OR REGULATION OF ELECTRIC MOTORS, ELECTRIC GENERATORS OR DYNAMO-ELECTRIC CONVERTERS; CONTROLLING TRANSFORMERS, REACTORS OR CHOKE COILS
H03B	GENERATION OF OSCILLATIONS, DIRECTLY OR BY FREQUENCY-CHANGING, BY CIRCUITS EMPLOYING ACTIVE ELEMENTS WHICH OPERATE IN A NON-SWITCHING MANNER; etc.
H03C	MODULATION
H03D	DEMODULATION OR TRANSFERENCE OF MODULATION FROM ONE CARRIER TO ANOTHER
H03K	PULSE TECHNIQUE
H03L	AUTOMATIC CONTROL, STARTING, SYNCHRONISATION, OR STABILISATION OF GENERATORS OF ELECTRONIC OSCILLATIONS OR PULSES
H03M	CODING; DECODING; CODE CONVERSION IN GENERAL
H04B	TRANSMISSION
H04J	MULTIPLEX COMMUNICATION
H04L	TRANSMISSION OF DIGITAL INFORMATION, e.g. TELEGRAPHIC COMMUNICATION
H04M	TELEPHONIC COMMUNICATION
H04N	PICTORIAL COMMUNICATION, e.g. TELEVISION
H04Q	SELECTING
H04T	INDEXING SCHEME RELATING TO STANDARDS FOR ELECTRIC COMMUNICATION TECHNIQUE
H04W	WIRELESS COMMUNICATION NETWORKS
H05K	PRINTED CIRCUITS; CASINGS OR CONSTRUCTIONAL DETAILS OF ELECTRIC APPARATUS; MANUFACTURE OF ASSEMBLAGES OF ELECTRICAL COMPONENTS