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 to establish new facts about the role of mega firms in generating "novel patents"—innovations that introduce new combinations of technology components for the first time. While the share of mega firms in novel patents had been declining until about 2000, it has strongly rebounded since then. This coincided with a shift in the technological contents of novel patents, characterized by the transition from new combinations within Information and Communications Technology (ICT) components to new combinations integrating ICT and non-ICT components. Mega firms also generate a disproportionately large number of "hits" —new combinations that lead to the largest numbers of follow-on patents (patents that reuse the same combinations of technology components as the first novel patent). Furthermore, their novel patents tend to diffuse broadly—we find that mega firms' most successful novel patents have more follow-on patents assigned not to themselves but to other firms compared to successful novel patents generated by non-mega firms. Overall, our findings suggest that mega firms play an increasingly important role in generating new technological trajectories in recent years, especially in combining ICT with non-ICT components, creating room for other entities to conduct follow-up innovations.

May 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, 2021; 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).

A key issue in this debate is the role of mega firms in economy-wide innovation and knowledge diffusion. Similar to the increase in market concentration, Akcigit and Ates (forthcoming) show that the share of patents held by the top one percent among patenting firms has been on the rise over the past several decades. They suggest 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 deceleration 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. Examining the role played by mega firms in economy-wide innovation process is important not just from an academic perspective but also because it has major policy implications.

In this paper, we aim to provide some new evidence that could shed light on the issues above. First, we define mega firms based *not* on their patent stocks but on economic scale. More specifically, mega firms in this paper are the top 50 firms by sales in any given year among all public firms in the Compustat data.[2] Second, we utilize the concept of "novel patents"—the subject of burgeoning research in the technology literature in recent years, motivated in part by the notion that many patents may be filed for purely strategic reasons and never used in applications (for empirical evidence see, e.g., Bessen and Hunt, 2004; Noel and Schankerman, 2013; Torrisi et al., 2016). In this paper, we define novel patents as those that introduce new combinations of technological components that had never been utilized together before (Fleming et al., 2007; Verhoeven et al., 2016).[3] Such patents represent economic experimentation and, if successful, may create pathways for new technological trajectories generating new products or adding new qualities to existing products. Thus, this concept appears to be most related to Schumpeter's (1911) definition of innovations as "new combinations" (see e.g., Epicoco et al., 2022; Pezzoni et al., 2022). We also explore the sensitivity of our findings to alternative measures of patent novelty.

Figure 1 illustrates two examples of novel patents. Panel (a) displays the patent titled "Trusted agents for open electronic commerce" applied in 1994 by Citibank. This patent combines CPC groups G06Q30 and H04L63 for the first time, introducing a system that enables anonymous transaction of electronic merchandise.[4] This innovation has greatly facilitated technological advancement in electronic commerce, solving the joint problem of protecting the privacy of buyers and sellers while ensuring the delivery of merchandise and money. Panel (b) shows a patent titled "Systems for activating and/or authenticating electronic devices for operation with apparel"

[2] All the findings below remain qualitatively the same if we redefine mega firms as the top 50 firms in terms of sales in any given year after excluding non-patenting firms or top one percent firms in sales in each two-digit NAICS industry in a given year.
[3] We use the patent Classification (CPC) system designed by the US Patent Office (USPTO) to measure technology components. While we use CPC groups as the level of disaggregation in the main analysis, our findings are robust to using different levels of aggregation as well as the IPC classification. See below for more details.
[4] G06Q30 is "Commerce" but it belongs to the CPC subsection G06Q which is Information and Communication Technology, while H04L63 is "Network architectures or network communication protocols for network security."

applied in 2006 by Nike, who combines CPC group G08C17 with CPC groups A43B3 and A41D1 for the first time.[5] This technology implants a wireless transmitting device into T-shirts and shoes to enable athletes to monitor vital signs and performance. This patent was accompanied by a joint commercialization with Apple through NIKE+iPod Sports Kit in 2006, years before the first release of Apple Watch in 2015.

[Figure 1 around here]

We document several new empirical facts. First, the share of mega firms in novel patent applications by the U.S. patent applicants had been declining for over two decades but there has been a turnaround since the early-mid 2000s. By the mid-2010s, the share of mega firms was the highest since 1980 when our sample starts. The share of mega firms in novel patent applications actually follows the dynamics very similar to their share in total patent applications. Furthermore, we show that mega firms were more likely to apply for novel patents even after controlling for various firm characteristics including size, industry, and the number of patents. This finding also holds within firms—firms produce more novel patents than before as they become mega firms. This suggests that closing on market leadership is associated with more, not less new combinations. Novel patents are also generally associated with better firm performance. We also document the overall increase in the number and share of novel patents in total patent applications in the U.S. since the mid-2000s, which reversed almost two decades of the declining trend.

Second, since many new combinations represent unsuccessful experiments, we adopt the approach suggested by the previous literature (e.g., Pezzoni et al., 2022) and track the number of "follow-on patents"—the patents that use the same new technology combination as the one first

[5] G08C17 is "Arrangements for transmitting signals characterized by the use of a wireless electrical link," A43B3 is "Footwear characterized by the shape or the use" and A41D1 is "Garments."

introduced by a novel patent—to measure the degree of success of a new combination. It turns out that mega firms generate a disproportionately larger number of "hits"—new combinations that lead to the largest numbers of follow-on patents—especially so in recent years.

Third, the increase in the share of novel patent applications by mega firms coincided with a big shift in the technological contents of novel patents. Most novel patents in the 1990s were generated by new combinations involving Information and Communications Technology (ICT) components. In more recent years, however, most novel patents are underpinned by combining ICT with non-ICT components for the first time. Moreover, such new combinations are generated not just by firms whose primary industry is ICT-related but also by firms operating in non-ICT-related industries (as exemplified by the NIKE patent above).

Finally, we show that compared to other firms, mega firms have smaller shares of follow-on patents assigned to themselves. This finding suggests that mega firms contribute to knowledge diffusion beyond their boundaries through engaging in technological experiments and generating impactful new combinations, a channel that has so far been understudied in the literature.[6]

Our findings have important policy implications. If it is true that dominant mega firms are stifling innovation and slowing down knowledge diffusion, there may be a scope for regulatory intervention. If, however, mega firms are the key actors conducting experiments and generating new technological trajectories, 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

[6] Patent reassignment as well as acquisitions may be another way for mega firms to defend their technological leadership and hinder knowledge diffusion (e.g,, Akcigit and Ates, forthcoming). While this is a reasonable hypothesis, subsequent changes in patent ownership are outside the scope of our analysis as we focus on the initial applicants for novel patents.

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 measurement. More details can be found in the Appendix. In Section 3 we present some basic evidence about the changing role of mega firms in novel patents and link this to some measures of firm performance. We also examine the role of mega firms (as well as VC-backed startups) in generating most impactful novel patents. In Section 4, we document a shift in the technological contents of novel patents from new combinations based on ICT components to new combinations involving ICT and non-ICT components. Such a shift could be behind the reversal of the decades-long trend toward declining share of novel patents. We also examine the diffusion of new technological trajectories beyond the innovating firms' own boundaries and the role of mega firms in such diffusion. Section 5 concludes.

2. Data and Measurement

The primary data sources are the USPTO PatentsView and S&P's Compustat. In some parts of our analysis, we compare mega firms with ventured-backed startups the information for which is obtained from VentureXpert data. The USPTO PatentsView tracks all patents ultimately granted by the USPTO from 1976 onward. 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. We collect utility patents granted to U.S. assignees between 1976 and 2020 to track economy-wide innovation activities and in particular, the creation and trajectory of new technological combinations. We describe detailed matching procedures in the Appendix.

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

Following previous studies (Fleming et al., 2007; Strumsky and Lobo, 2015), we define a new technological combination as the first time a pairwise combination of technological

[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.
[9] Section Y represents a new addition to patent classifications introduced together with CPC, for general tagging of new technological developments which are already classified or indexed in other sections. We exclude technological components tagged under this section when constructing new combinations.

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. Because 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 three years, 1976-1979, as a buffer period to capture the history of technological components and we track novel combinations starting from 1980. We use the data starting in 1991 for much of the analysis and thus our results are unlikely to be contaminated by false positive 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 versus different assignee(s) from the 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.

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. To examine how sensitive our findings below are to this methodology of identifying ICT-related technology classes, we also utilize the "J tag" taxonomy based on the mapping between International Patent Classification (ICP) and the OECD definition of ICT-related products (Inaba and Squicciarini, 2017).

3. Mega firms and Novel Patents

3.1 Increasing share of mega firms in novel patent applications in recent years

As mentioned, we define mega firms as the top 50 in sales among all public firms in the Compustat data in each year during our sample (1980-2016), but our results are robust to using alternative definitions of mega firms such as top one percent firms in sales in each two-digit NAICS industry in each year. We then use the bridge between the U.S. patents and Compustat firms described in the previous section and in the Appendix to measure patenting activity by such mega firms.[10]

[10] We identify mega firms in the Compustat data prior to merging it with the USPTO data. 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

Figure 2 shows the dynamics of the share of mega firms in the number of all and novel patent applications by all U.S. patent assignees.[11] The first thing to note is that while the share of mega firms in novel patent applications is somewhat lower than their share in total patent applications, the dynamics are very similar; whenever mega firms' share in total patents increases, it increases also in terms of novel patents.

[Figure 2 around here]

Examining the time trend, we see a steep decline from the 1990s-early 2000s, followed by an equally steep recovery since then. The share of mega firms in novel patent applications had declined by half, from about 16 percent in the early 1980s to about eight percent in 2000 but has completely recovered by 2016. Their share in novel patent applications by Compustat firms (not shown) had declined from 22-23 percent in the early 1980s to less than 14 percent by 2000 but has increased to 32-33 percent by 2016. Either way, mega firms were generating relatively more novel patents in the mid-2010s than at any time since the early 1980s.

We exploit our Compustat-USPTO matched firm panel data to examine the likelihood of producing novel patents across publicly traded firms and over time, controlling for firm characteristics. Specifically, we estimate the following regression:

$$Novel\ Patents_{ijt} = \alpha + \beta_1 I_{\{mega\ firm\}ijt} + X_{ijt} + \delta_{jt} + \varepsilon_{ijt}, \qquad (1)$$

where $Novel\ Patents_{ijt}$ is the inverse hyperbolic sine (IHS) transformation of the number of novel patents applied by firm i in industry j in year t, $I_{\{mega\ firm\}ijt}$ is an indicator for a mega firm,

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.

[11] We also constructed the same figure showing the share of mega firms in the stock of patent applications by the U.S. public firms in Compustat, and while the levels of the share of mega firms were higher (because the denominator includes only patents by Compustat firms), the dynamics were very similar. Details are available upon request.

X_{ijt} is the vector of time-varying firm characteristics, including (logged) firm employment size, (logged) sales, and the (logged) total number of patents, and δ_{jt} are industry-year fixed effects.

[Table 1 around here]

The estimation results are presented in panel (a). The coefficient of interest is β_1, which is positive and statistically significant in the first two columns, indicating that mega firms are likely to produce more novel patents even after controlling for time-varying characteristics including the total number of patents. We also include the interaction term between the mega firm indicator and a dummy variable for the post-2007 period as follows:

$$Novel\ Patents_{ijt} = \alpha + \beta_1 I_{\{mega\ firm\}ijt} + \beta_2 I_{\{mega\ firm\}ijt} X\{2007 - 2016\ period\}_t + X_{ijt} + \delta_{jt} + \varepsilon_{ijt}. \quad (1')$$

The estimation results are shown in the last two columns of Table 1, panel (a). In particular, the coefficient β_2 in column (4) is positive and statistically significant, indicating that the increase in the share of mega firms in novel patent applications in recent years observed in Figure 2 holds after controlling for size and the number of total patents, as well as industry by year fixed effects.

In panel (b) of Table 1 we present the results of similar estimations, including also firm fixed effects, δ_i in the two regressions above. Since the identification of the mega firm dummy in this case is based on firms that change their "status" from non-mega firms to mega firms (and vice versa) during the sample, the findings can be interpreted as suggesting that firms are more (less) likely to produce novel patents as they switch from (to) non-mega firm to (from) mega-firm. The pattern also gets more pronounced in the period after 2007. Once again, the results are robust to including time-variant firm characteristics in columns (2) and (4).

These results suggest that market-leading mega firms are more likely to engage in novel innovation activities, especially in recent years, rather than becoming less innovative. Inasmuch as novel patents are a proxy for novel technologies, the stronger likelihood of producing novel patents after 2007 for mega firms can be interpreted as suggestive evidence of a possibly turning tide in the U.S. innovation, to a large extent driven by mega firms.

The above findings suggest that novel patents and firm performance may be related. To examine this association further 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 three ways; as (logged) sales, (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, including the (logged) total number of patent applications (to control for the overall patenting propensity of the firm) as the baseline.[12] 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. Our goal here is simply to see if there is any empirical association between creating new combinations and firm performance, without claiming any causal relationship in either direction.

Table 2 shows the estimation results where the dependent variable is firm size measured by the (logged) real sales. For the one-year lag (s = 1), the relationship is statistically insignificant and small once we control for firm size in terms of both lagged employment and the total patent stock. For $s > 1$, however, the relationship becomes positive and statistically significant. The results are similar using employment and TFPR as the dependent variables (Tables A3 and A4 in the Appendix).

[Table 2 around here]

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

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 sales 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 render further support to the importance of novel patents in firm growth.

3.2 Changing tide in novel patents?

The changes in the share of novel patents by mega firms are closely associated with changing dynamics of novel patents in the U.S. economy overall. In Figure 3 we present the dynamics of the number of novel patents and their share in total patent applications in the U.S. over time. The absolute number of novel patent applications had been basically flat while the share of novel patents in total had been declining steadily from 1980 and until about 2007, reflecting a rapid increase in the total number of patent applications. As a result, the share of novel patents in total patent applications had dropped all the way from 12% in 1980 (8% at the start of the 1990s) to 3% in 2007. This downward trend is broadly consistent with the decline in the average creativity of patents documented in Arts et al. (2021) and Kalyani (2022) who use Natural Language Processing (NLP) measures of patent novelty. Kalyani (2022) interprets this trend as being consistent with the slowdown in aggregate productivity growth and decrease in R&D efficiency (Bloom et al., 2020).

[Figure 3 around here]

After 2007, 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. This trend reversal is not observed in NLP-based measures of novel patents, and it suggests that our measure based on co-assignment of CPC main groups captures different aspects of patent novelty. It is also worth noting that the trend toward increasing share of novel patents produced by mega firms documented in the previous subsection continues to hold in the NLP-based measures (Figure A3 in the Appendix). Thus, while there is daylight between our findings of the relative increase in novel patents in recent years and the findings from the NLP methodology that shows unabated relative decline, the increase in the role of mega firms is observed in both cases.

3.3 Most impactful new combinations and mega firms

One interpretation of novel patents and new combinations of technological components that underpin them is that they represent innovations that are inherently experimental in their nature. As with all experiments, many would fail, while others would lead to different degrees of success. Following Pezzoni et al. (2022), we empirically measure the degree of success of a novel patent by the number of follow-on patents it generates; that is, the number of patents that re-use the same new technological combination as the initial novel patent.[13] 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 the first novel patent. Some novel patents, however, quickly generate a large number of follow-on patents and thus have a big immediate impact on shaping new technological trajectories.

We first examine how successful mega firms were in generating follow-on patents and how that changed over time by means of a simple regression. To compare mega firms with other highly innovative businesses, we use patenting non-mega firms as a baseline, with an explicit indicator for VC-backed startups which are known to produce highly influential innovation and patents (Kortum and Lerner, 2000; Howell et al., 2020). We also focus specifically on the comparison

[13] An alternative measure would be the number of forward citations. The findings presented below are robust to utilizing this alternative measure of the impact of a novel patent.

between two decades—the 1991-2000 decade which includes the dot.com boom period, and the most recent, 2007-2016 decade in this subsection. 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. The outcome variables, in columns (1)-(3) in Table 3 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, respectively.

[Table 3 around here]

The estimation results are presented in Table 3. 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 2007-2016. 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 6% more in 2007-2016. The probability of not having any follow-on patent was also higher by 4.8% for mega firms in 1991-2000 but it was 1.7% lower than for other entities in 2007-2016 (column (2)). Interestingly, VC-backed startups 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" from

2007-2016 to 1991-2000 for VC-backed startups is significantly less pronounced than for mega firms. Thus, mega firms not only increased the number of new combinations from the 1990s to the most recent decade, but they also had a large increase in the impact of those new combinations.

To probe the changing role of mega firms in generating most successful novel patents, we follow Pezzoni et al. (2022) and identify "hits" (most impactful new combinations). For our main analysis we define a "hit" as a new combination that generated the number of follow-on patents reusing the same combination in the top one percentile of the distribution of follow-on patents within the first five years, although the findings are robust to using other cutoffs, such as the 95th percentile. We then investigate how mega firms generate especially "successful" patents, i.e., those associated with at least one "hit" in their technology class classification.

Figure 4 shows changes in the shares of mega firms in all as well as top "hit" new combinations between the two decades, 1991-2000 and 2007-2016. Consistent with what we saw in Figure 2, the share of mega firms in the total number of new combinations (with or without follow-on patents) increased from 9.3 percent in 1991-2000 to 12.9 percent in 2007-2016. At the same time, their share in "hits" increased by much more—indeed, it almost doubled from 11.9 to 21.2 percent from 1991-2000 to 2007-2016. Also, VC-backed startups produced twice the share of hits compared to their share in all new combinations in both periods, while 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 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]

[Figure 4 around here]

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

4. Changes in Technology Content of Novel Patents and Their Diffusion

4.1 Shift from within-ICT to ICT-non-ICT new combinations

What is behind the changes in the time trend of novel patents in recent years? In this section we first present some evidence that shows a big shift that has occurred in the technological content of new combinations embodied in novel patents, especially those that are hits in the sense of seeding whole new technological trajectories. We then speculate that this shift might be one reason why we observe the reversal in the trend in novel patents.

Recall that in the novel patent examples in Figure 1 above, the first patent (in Figure 1a) combined two ICT-related components, but the second patent (in Figure 1b) combined an ICT-related technology component with two non-ICT components. These examples are illustrative of the big change in the novel technological trajectories that has happened in recent years.

In Table 4 (see also Figure A2 in the appendix), 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 4) if the new combination linked for the first time only the CPC main groups that are matched to ICT-related industries (see Section 2 above for how this match was constructed), to be between ICT and non-ICT (“ICT & non-ICT” in Table 4) 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 4) if the new combination linked for the first time the CPC main groups neither of which is matched to an ICT-related industry.

[Table 4 around here]

The difference between the 1990s and 2007-2016 we see in Table 4 is striking. In the 1990s, 62% of the top hits were new combinations combining technologies within ICT (68% among top hits produced by mega firms). In contrast, in 2007-2016, that is, during the recent resurgence of novel patents and new combinations, 54.7% of top hits (and 53.1% 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.[15]

The changing technological contents of new combinations and novel patents they generate can also be seen directly in the patent data. Note that each new combination of previously not connected knowledge components adds a new connecting edge across different types of knowledge in the common stock of patented new knowledge as time goes by. In Figure 5 we present the dynamics of the average (valued) degree centrality of technological components in the knowledge network,[16] 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. Consistent with the big increase in the number of novel patents since the late 2000s, Figure 5, Panel A shows a sharp increase in the (valued) degree centrality in all CPC technology sections during that period but especially in Sections G (physics) and H (electricity) which dominate ICT-related subclasses (Table A2 in the Appendix).

[Figure 5 around here]

[15] We also did these calculations using an alternative definition of ICT-related technology classes proposed by Inaba and Squicciarini (2017) based on a very different matching algorithm. The results in Table 4 remained very similar.
[16] 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.

In Figure 5 Panel B we decompose the increase in valued degree centrality from Panel A into within- and between CPC sections and subsections. The recent shift from within-ICT to ICT-non-ICT new combinations is manifested in the lion's share of new combinations being generated *across*, not *within* the same CPC section, even at the most aggregated, one-digit CPC classification level. Once again, this is especially pronounced in the physics and electricity sections.

The role of ICT-related components being combined with non-ICT components to generate "hit" novel patents can be further examined by looking at all new combinations generated since 2006. We identified the top 0.1% among 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 6 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 6.

[Figure 6 around here]

The color in Figure 6 (and the interactive animation in the link above) indicates different CPC sections. For example, yellow color 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 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 6 is CPC main group H04W4 (services specially adapted for wireless communication networks; facilities therefor) and we can see a lot of links connecting 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. This means a lot of follow-on patents combining these components to generate improvements in Advanced Driver Assistance Systems (ADAS) and self-driving cars.

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 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. We conjecture that this could be one reason behind the recent change in the time trend in novel patents.

4.2 What is happening to the diffusion of new technologies?

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 7 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 3 above, surged.

[Figure 7 around here]

In Table 5 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 5 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 for new combinations that had been generated prior to 2001. The findings are presented in Figure 8 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 8 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 mega firms.

[Figure 8 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 the trend in the number of follow-on patents assigned to the non-focal firm is now virtually the same for mega firms and VC-backed startups. 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 5, 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. Conclusions

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 clogging 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 new combinations had declined over several decades until the mid-2000s, followed by a rebound since then and mega firms played a large role in this trend reversal.

This seems to be closely related to mega firms increasingly combining ICT components with non-ICT components in their experimentation. The extent to which these new combinations generate follow-on innovation by other firms is high for mega firms even compared to VC-backed startups, which are the entities often considered to be at the heart of technological experimentation.

The recent shift toward combining ICT and non-ICT components in novel patents may also be behind some of the differences between the overall trend in novel patents we find in this paper and the same trend using the NLP-based methodology. NLP-based measures may be better suited to capturing creative, ground-breaking inventions (for instance, they better capture patents linked to awards such as the Nobel Prize— Arts et al., 2021). Such inventions, however, are neither necessary nor sufficient for economically important innovations (e.g., Mowery and Rosenberg, 1989). In contrast, patents that generate new combinations of existing technological components may not represent ground-breaking inventions, but they are linked to experimentation with new products and/or new product qualities and as such could lead to economic success.

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).

Arts, S., Hou, J., & Gomez, J.C. (2021). Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures. *Research Policy*, 50(2), 104144.

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.

Howell, Sabrina T., Josh Lerner, Ramana Nanda, and Richard R. Townsend. (2020). How resilient is venture-backed innovation? evidence from four decades of us patenting (No. W27150). 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.

Inaba, T. & Squicciarini, M. ICT: A new taxonomy based on the international patent classification. OECD Science, Technology and Industry Working Papers 2017/01, 2017.

Kalyani, A. (2022). The Creativity Decline: Evidence from U.S. Patents. Working paper, Boston University.

Klepper, S. (2016). *Experimental Capitalism: The Nanoeconomics of American High-Tech Industries*. Princeton University Press.

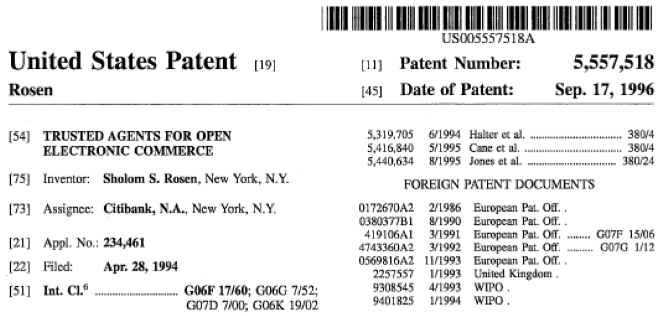
Kortum, S., & Lerner, J. (2000). Assessing the contribution of venture capital to innovation. *RAND journal of Economics*, 674-692.

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.
Mowery D., & Rosenberg N. (1989). Technology and Pursuit of Economic Growth. Cambridge University Press, New York, NY.
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.
Schumpeter, J. (1934; first German edition 1911). The Theory of Economic Development. Harvard University Press, Cambridge, MA.
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 1a. A novel patent by Citibank combining two ICT-related components



US005557518A

United States Patent [19] — [11] **Patent Number: 5,557,518**

Rosen — [45] **Date of Patent: Sep. 17, 1996**

[54] **TRUSTED AGENTS FOR OPEN ELECTRONIC COMMERCE**

[75] Inventor: **Sholom S. Rosen**, New York, N.Y.

[73] Assignee: **Citibank, N.A.**, New York, N.Y.

[21] Appl. No.: **234,461**

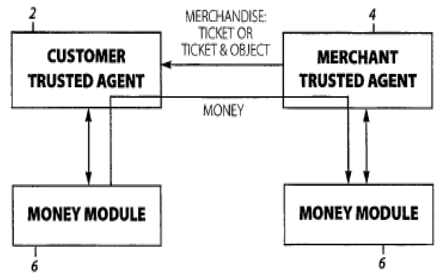
[22] Filed: **Apr. 28, 1994**

[51] **Int. Cl.**[6] **G06F 17/60**; G06G 7/52; G07D 7/00; G06K 19/02

5,319,705	6/1994	Halter et al.	380/4
5,416,840	5/1995	Cane et al.	380/4
5,440,634	8/1995	Jones et al.	380/24

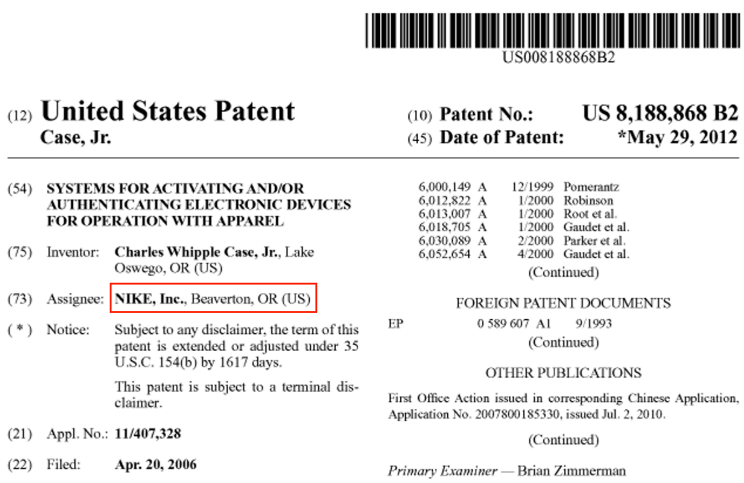
FOREIGN PATENT DOCUMENTS

0172670A2	2/1986	European Pat. Off. .	
0380377B1	8/1990	European Pat. Off. .	
419106A1	3/1991	European Pat. Off.	G07F 15/06
4743360A2	3/1992	European Pat. Off.	G07G 1/12
0569816A2	11/1993	European Pat. Off. .	
2257557	1/1993	United Kingdom .	
9308545	4/1993	WIPO .	
9401825	1/1994	WIPO .	



H04L63 - Network architectures or network communication protocols for network security → G06Q30 – Commerce (G06Q – Information and communication technology specially adapted for administrative, commercial, financial, managerial or supervisory purposes)

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



US008188868B2

(12) **United States Patent** — (10) **Patent No.: US 8,188,868 B2**

Case, Jr. — (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

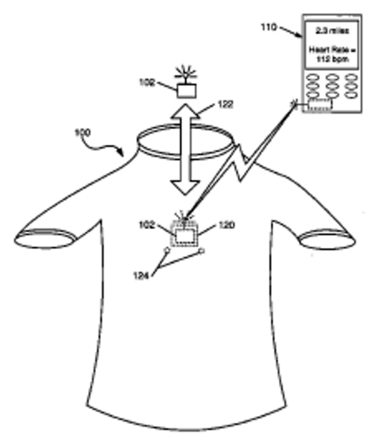
(Continued)

OTHER PUBLICATIONS

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

(Continued)

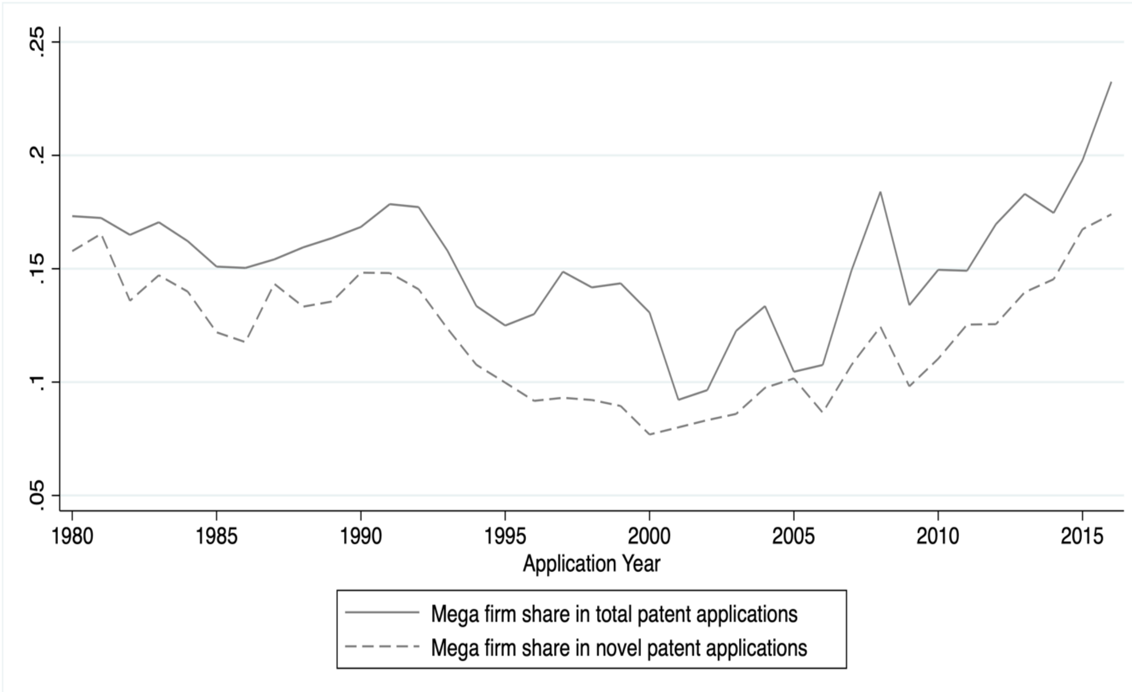
Primary Examiner — Brian Zimmerman



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

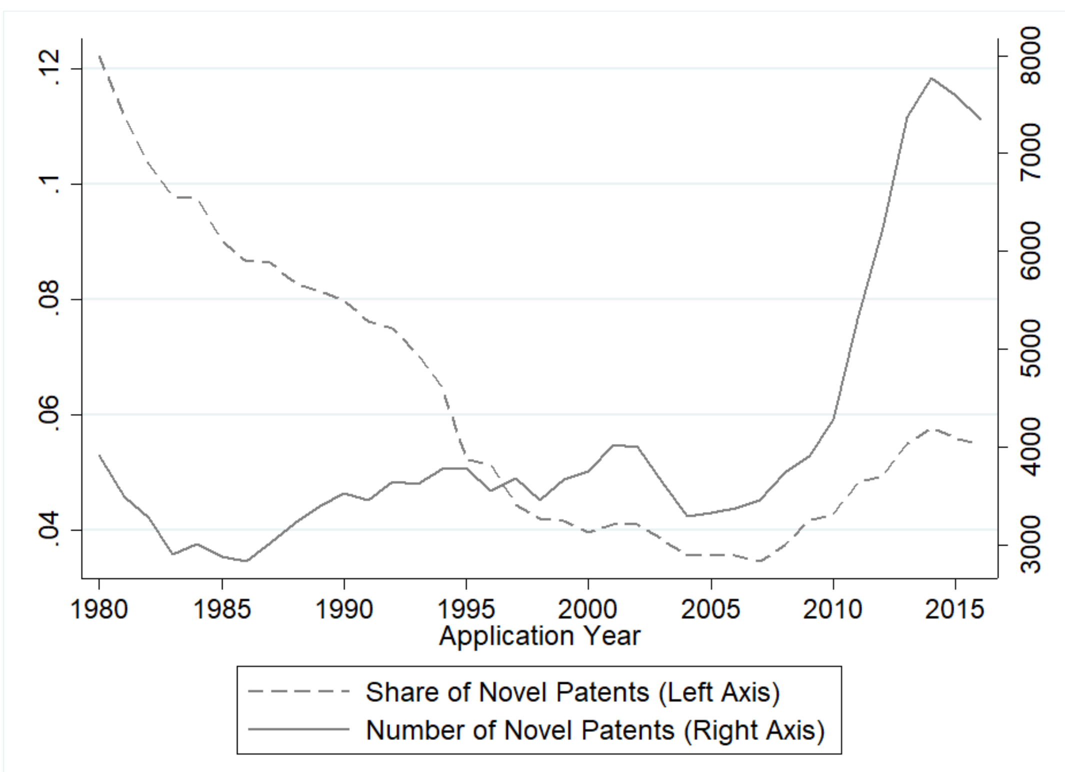
Source: USPTO data.

Figure 2. Share of mega firms in total number of patents applications



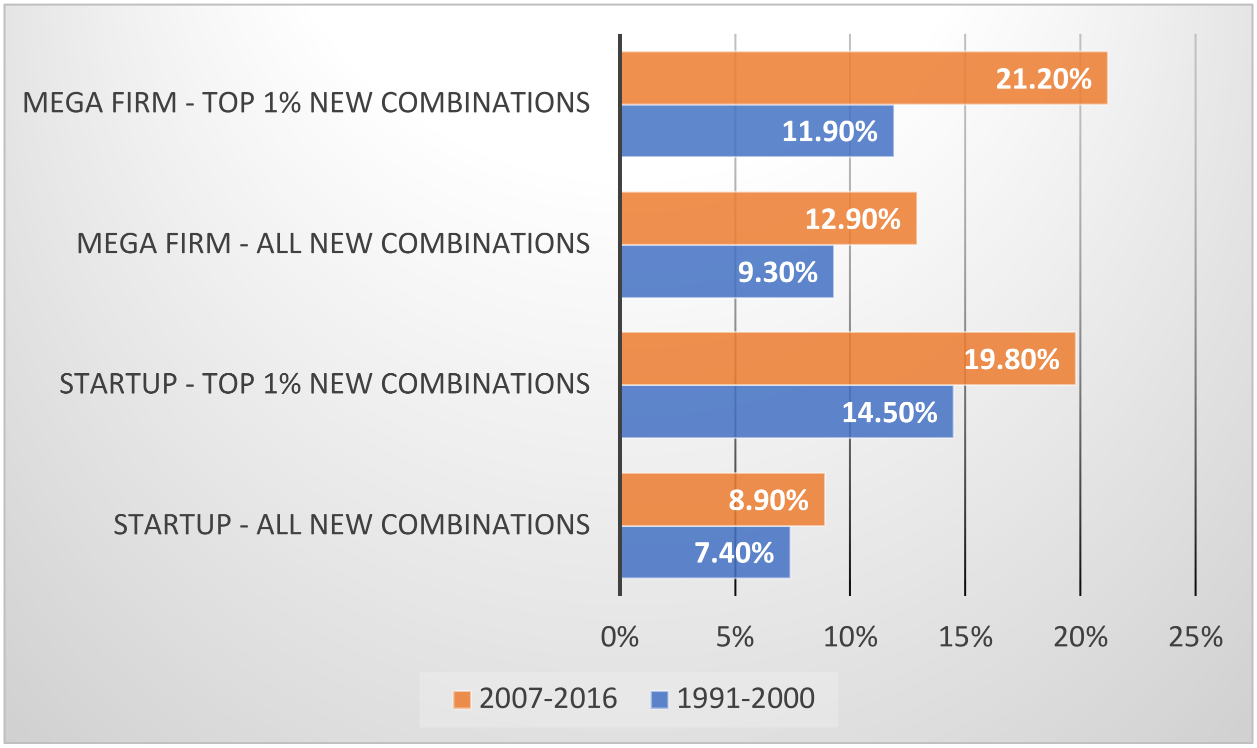
Source: Authors' own calculation using the USPTO data matched with Compustat data.

Figure 3. The number of novel patents and their share in total patent applications



Source: Authors' own calculation using USPTO data.

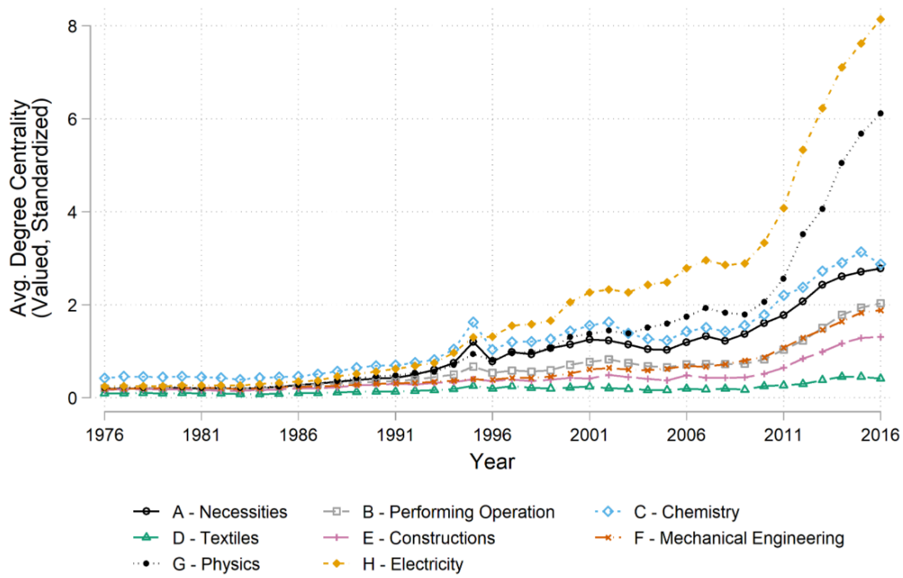
Figure 4. All and top one percent new combinations by firm type.



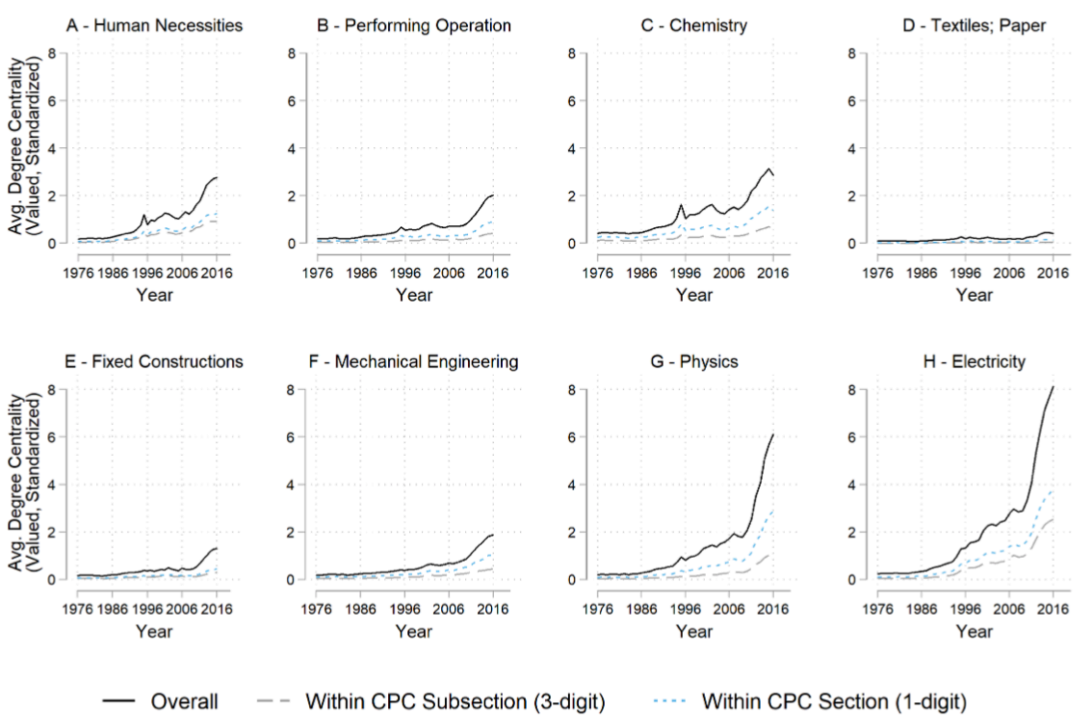
Source: Authors' own calculation using the USPTO data matched with Compustat data. Top one percent new combinations: based on the number of follow-on patents using the same technological combinations.

Figure 5. Time trend in valued degree centrality by CPC sections

Panel A. Overall

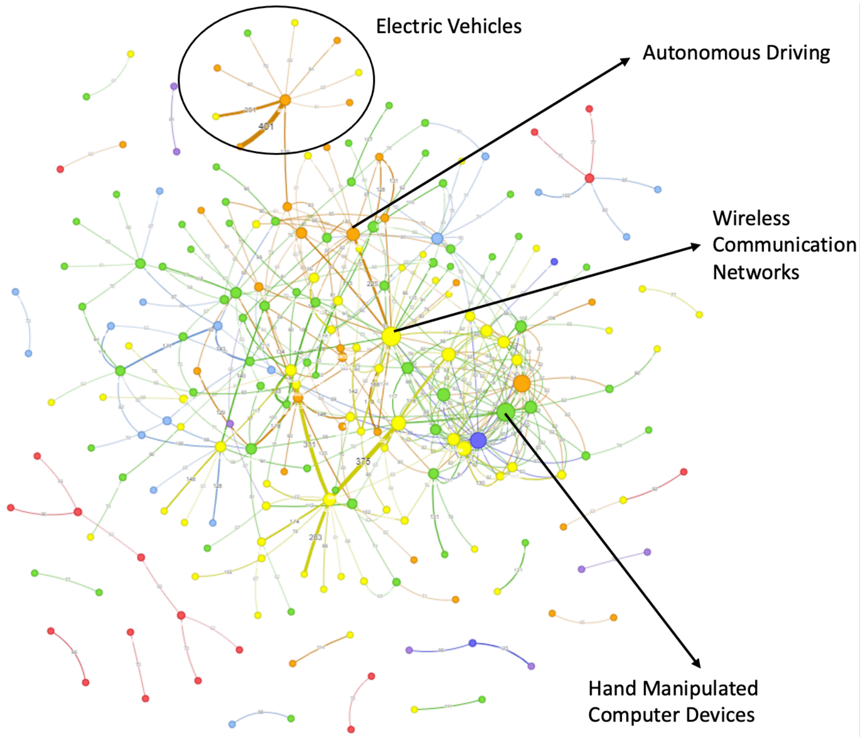


Panel B. Decomposition into within and between CPC sections



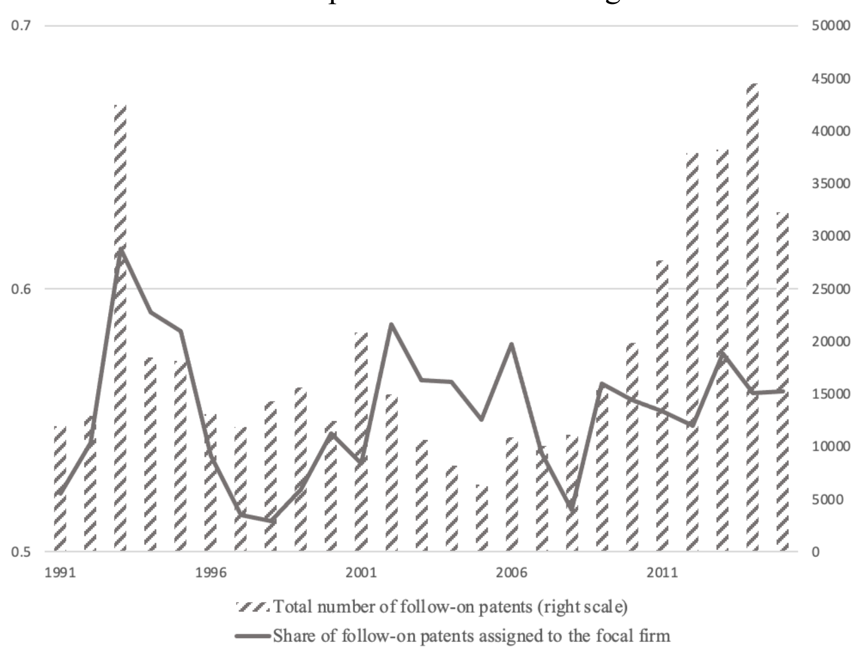
Source: Authors' calculation using USPTO data. See the main text for the definition of valued degree centrality.

Figure 6. Technological content of most frequently used new combinations invented after 2006.



Combinations that are used more than 61 times by follow-on patents. (i.e., top 99.9% among all new combinations). Size of the node indicates the number of other technologies linked to the focal one. Source: created by Yuheng Ding using USPTO data.
Online interactive version: https://www.yuhengding.com/about/vis_most_successful_new_combinations

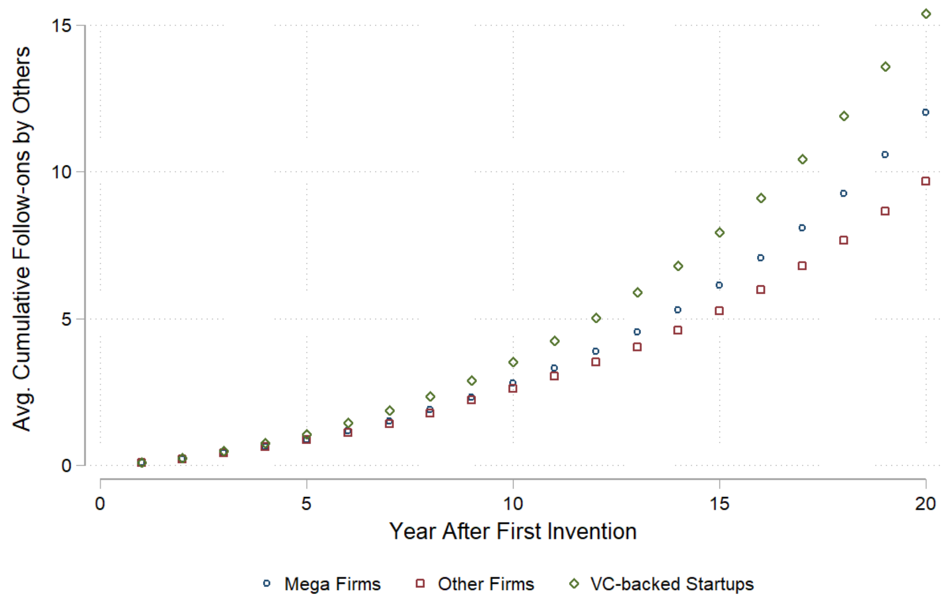
Figure 7. Number of follow-on patents and share assigned to the focal assignee



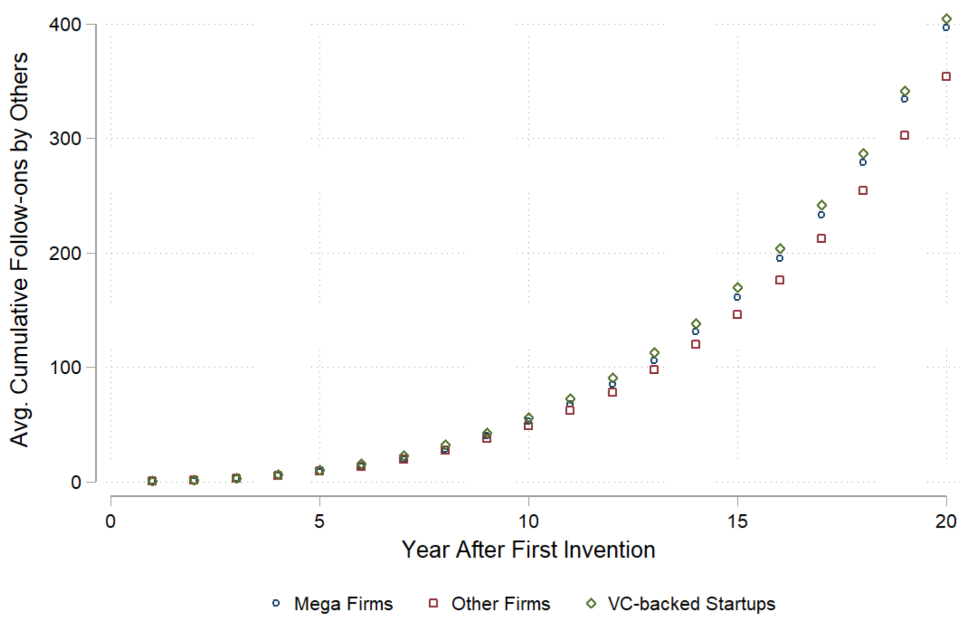
Source: Authors' own calculation using USPTO data.

Figure 8. Twenty-year diffusion curves to other assignees by firm types

(a) All new technological trajectories



(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. Novel Patents by Mega Firms in 1980-2016

(a) Without firm fixed effects

	(1)	(2)	(3)	(4)
DV:	IHS # novel patents	IHS # novel patents	IHS # novel patents	IHS # novel patents
VARIABLES				
Dummy equal to one if mega firm	1.793***	0.650***	1.716***	0.619***
	(0.025)	(0.020)	(0.029)	(0.023)
Mega firm X 2007-2016 period			0.277***	0.111***
			(0.056)	(0.042)
Logged employment		0.038***		0.038***
		(0.004)		(0.004)
Logged real sales		-0.030***		-0.030***
		(0.003)		(0.003)
Logged # total patents		0.440***		0.440***
		(0.003)		(0.003)
Constant	0.365***	-0.381***	0.365***	-0.381***
	(0.003)	(0.006)	(0.003)	(0.006)
FE	Year-Industry	Year-Industry	Year-Industry	Year-Industry
Observations	53,819	47,524	53,819	47,524
adj. within R2	0.0939	0.5456	0.0943	0.5457

(b) With firm fixed effects

	(1)	(2)	(3)	(4)
DV:	IHS # novel patents	IHS # novel patents	IHS # novel patents	IHS # novel patents
VARIABLES				
Dummy equal to one if mega firm	0.305***	0.211***	0.184***	0.110***
	(0.031)	(0.030)	(0.034)	(0.032)
Mega firm X 2007-2016 period			0.419***	0.350***
			(0.044)	(0.042)
Logged employment		0.011*		0.013**
		(0.006)		(0.006)
Logged real sales		-0.001		-0.001
		(0.005)		(0.005)
Logged # total patents		0.370***		0.369***
		(0.004)		(0.004)
Constant	0.411***	-0.267***	0.322***	-0.266***
	0.305***	0.211***	(0.002)	(0.009)
FE	Firm, Year-Industry	Firm, Year-Industry	Firm, Year-Industry	Firm, Year-Industry
Observations	51,852	45,650	51,852	45,650
adj. within R2	0.0023	0.1938	0.0044	0.1954

Note: Estimation method: OLS, absorbing year fixed effects. Robust standard errors in parentheses. *** p < 0.01. Successful patents mean novel patents associated with hits (top 1% new combinations). IHS is inverse-hyperbolic sine transformation: $y = ln\left(x + \sqrt{x^2 + 1}\right)$.

Table 2. Novel patents and Sales

	(1)	(2)	(3)	(4)	(5)	(6)	(7))	(8)	(9)
					Logged Sales				
Total novel patents$_{t-1}$	0.130***	0.012***	0.002						
	(0.019)	(0.004)	(0.001)						
Total novel patents$_{t-2_}$				0.138***	0.015***	0.004**			
				(0.021)	(0.004)	(0.002)			
Total novel patents$_{t-3}$							0.142***	0.016***	0.005**
							(0.022)	(0.004)	(0.002)
Logged employment$_{t-1}$			0.851***						
			(0.020)						
Logged employment$_{t-2}$						0.704***			
						(0.022)			
Logged employment$_{t-3}$									0.587***
									(0.024)
Logged total patents$_{t-1}$			0.017**						
			(0.007)						
Logged total patents$_{t-2}$						0.023***			
						(0.009)			
Logged total patents$_{t-3}$									0.024**
									(0.010)
Constant	1.271***	1.398***	1.024***	1.312***	1.450***	1.179***	1.364***	1.514***	1.330***
	(0.050)	(0.006)	(0.016)	(0.052)	(0.006)	(0.019)	(0.053)	(0.006)	(0.021)
Fixed effects	none	Firm & industry-		none	Firm & industry-		none	Firm & industry-	
Observations	39,686	35,274	31,444	38,254	33,808	28,363	36,435	32,046	25,383
adj. within R2	0.08771	0.00278	0.41338	0.09436	0.00390	0.31132	0.10026	0.00444	0.21849

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

Table 3. Follow-on patents on new combinations by mega firms and VC-backed startups

VARIABLES	DV: (1) IHS # follow-on patents	(2) "Failed"	(3) Logged # follow-on patents
Dummy equal to one if mega firm	-0.115***	0.048***	-0.068***
	(0.011)	(0.006)	(0.014)
Mega firm X 2007-2016 period	0.166***	-0.065***	0.129***
	(0.013)	(0.007)	(0.017)
Dummy equal to one if VC-backed startup	0.225***	-0.072***	0.190***
	(0.012)	(0.006)	(0.014)
VC-backed startup X 2007-2016 period	0.046***	-0.025***	0.049***
	(0.015)	(0.008)	(0.017)
Constant	0.773**	0.493***	0.7731***
	(0.002)	(0.001)	(0.003)
Year FE	Included	Included	Included
Observations	228,831	228,831	125,342
R-squared	0.086	0.114	0.025

Note: Estimation method: OLS, absorbing year fixed effects. Robust standard errors in parentheses. *** $p < 0.01$. Follow-on patents within 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 4. Combination types of 1% successful new combinations: 1991-2000 and 2007-2016

		Period			
		1991-2000		2007-2016	
Mega firms	ICT & ICT	86	68.3%	34	9.6%
	ICT & non-ICT	23	18.3%	189	53.1%
	non-ICT & non-ICT	17	13.5%	133	37.4%
	Total	126	100.0%	356	100.0%
All assignees	ICT & ICT	488	62.0%	194	11.9%
	ICT & non-ICT	147	18.7%	891	54.7%
	non-ICT & non-ICT	152	19.3%	545	33.4%
	Total	787	100.0%	1630	100.0%

Source: Authors' calculation using USPTO and Compustat data.

Table 5.

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***	0.007
		(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**		0.197***	-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***	0.021*	0.002	0.068***
	(0.003)	(0.003)	(0.003)	(0.012)	(0.013)	(0.014)
Observations	142,733	142,733	142,733	2,652	2, 652	2, 652
R-squared	0.000	0.001	0.002	0.137	0.196	0.239

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

Appendix

A.1 USPTO-Compustat Matching

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.

A.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. [18]

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

[18] 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).

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.

A.3. Additional Figures and Tables

Figure A1. UPSTO-Compustat Match Rates

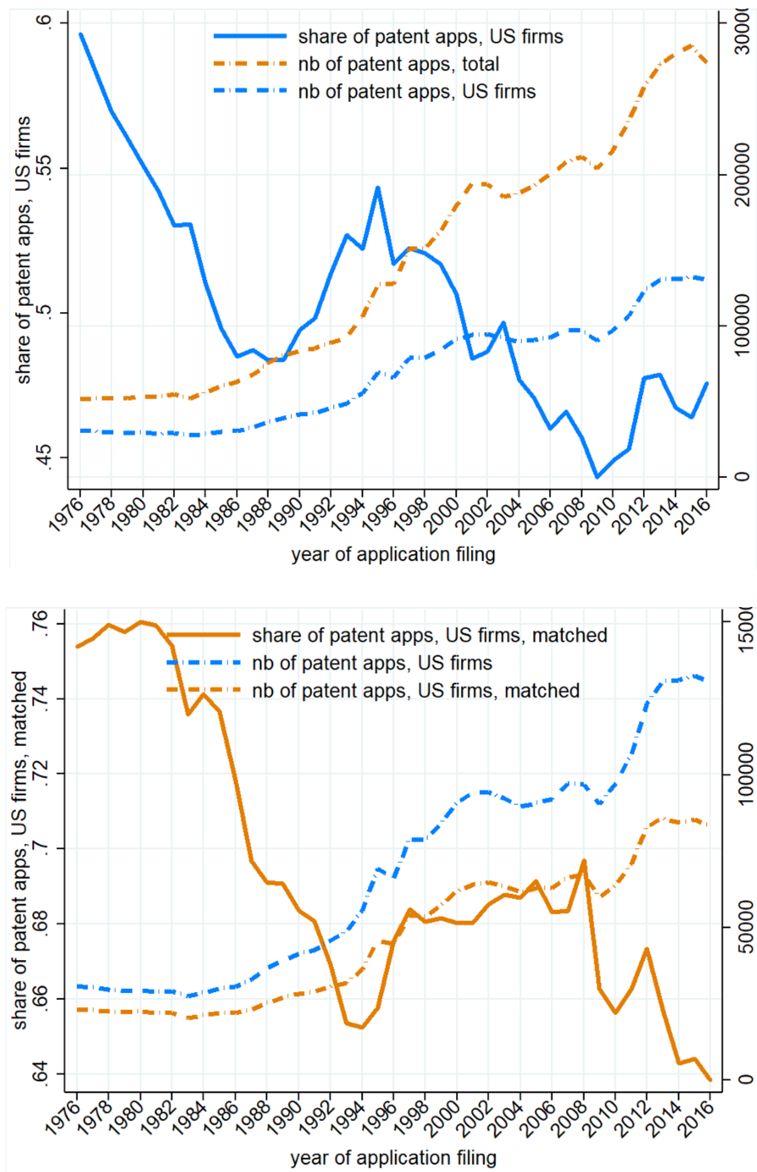
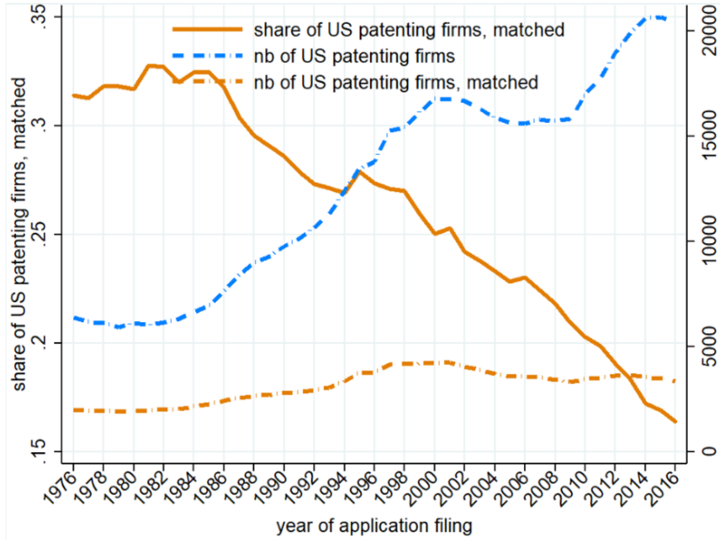
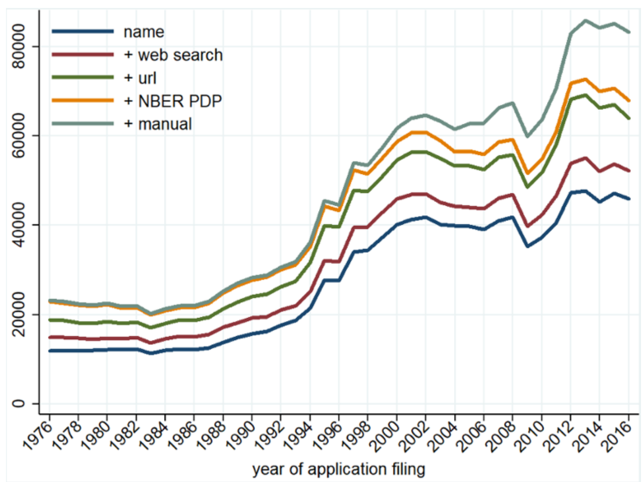


Figure A1. UPSTO-Compustat Match Rates (continued)





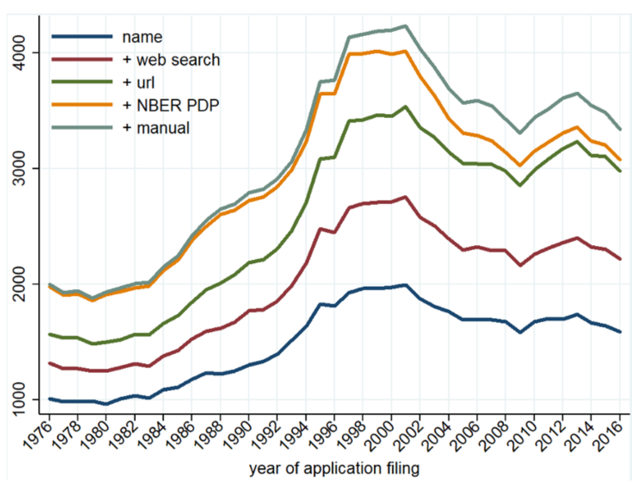


Figure A2.

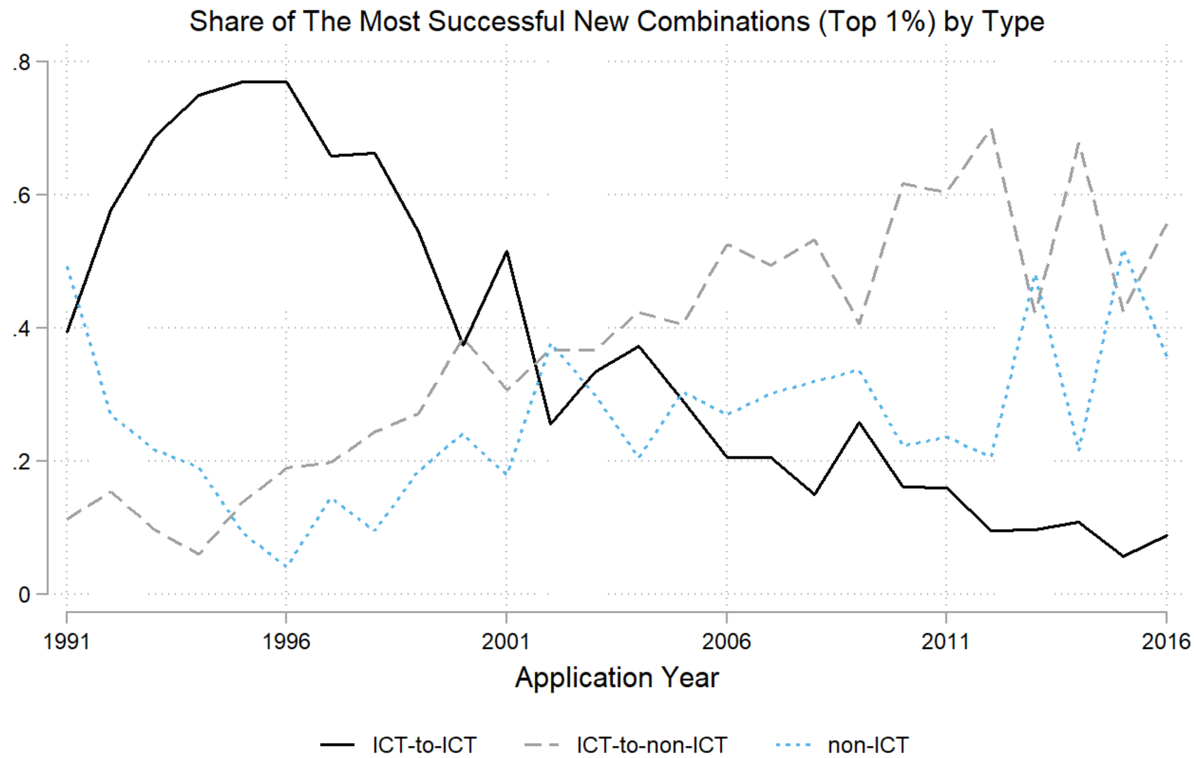


Table A1. List of ICT Industries

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

Table A3. Novel patents and firm employment

	(1)	(2)	(3)	(4)	(5)	(6)	(7))	(8)	(9)
	Logged employment								
Total novel patents$_{t-1}$	0.117***	0.011***	0.002						
	(0.017)	(0.003)	(0.002)						
Total novel patents$_{t-2_}$				0.125***	0.013***	0.003*			
				(0.019)	(0.003)	(0.002)			
Total novel patents$_{t-3}$							0.130***	0.016***	0.005**
							(0.020)	(0.003)	(0.002)
Logged sales$_{t-1}$			0.399***						
			(0.015)						
Logged sales$_{t-2}$						0.329***			
						(0.015)			
Logged sales$_{t-3}$									0.280***
									(0.015)
Logged total patents$_{t-1}$			0.098***						
			(0.007)						
Logged total patents$_{t-2}$						0.094***			
						(0.008)			
Logged total patents$_{t-3}$									0.082***
									(0.008)
Constant	0.401***	0.513***	0.099***	0.430***	0.553***	0.019	0.470***	0.606***	0.180***
	(0.045)	(0.005)	(0.007)	(0.047)	(0.005)	(0.024)	(0.048)	(0.005)	(0.026)
Fixed effects	none	Firm & industry-		none	Firm & industry-		none	Firm & industry-	
Observations	38,528	34,107	32,218	37,224	32,821	29,453	35,487	31,092	26,300
adj. within R2	0.09924	0.00476	0.39411	0.10646	0.00657	0.29866	0.11380	0.00812	0.22627

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

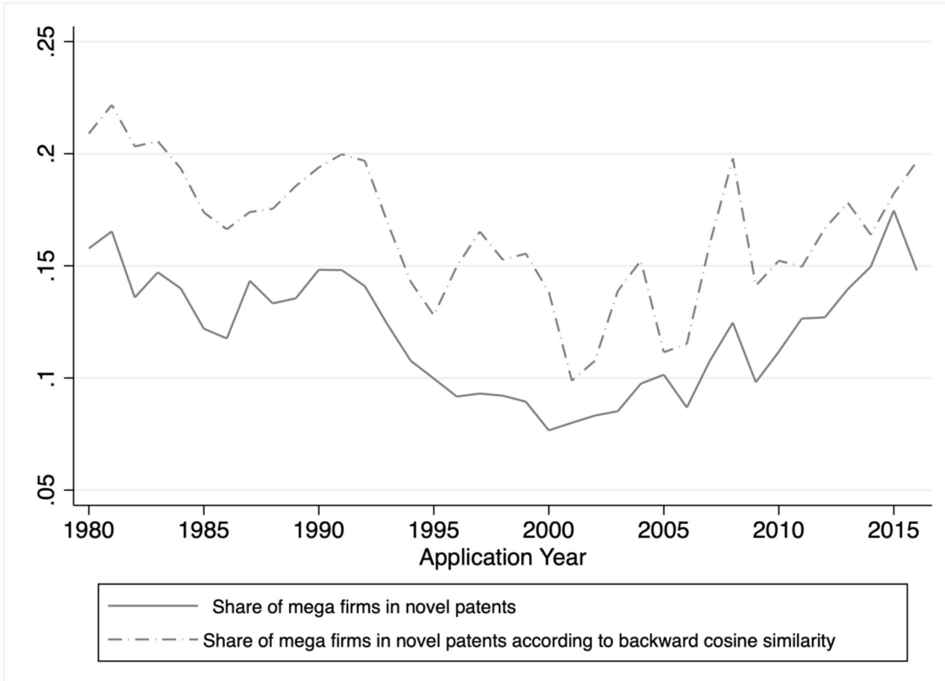
Table A4. Novel patents and TFPR

	(1)	(2)	(3)	(4)	(5)	(6)	(7))	(8)	(9)
					Logged TFPR				
Total novel patents$_{t-1}$	0.032***	0.004**	0.003*						
	(0.007)	(0.002)	(0.002)						
Total novel patents$_{t-2_}$				0.033***	0.004**	0.003*			
				(0.007)	(0.002)	(0.002)			
Total novel patents$_{t-3}$							0.031***	0.004**	0.004**
							(0.007)	(0.002)	(0.002)
Logged sales$_{t-1}$			0.152***						
			(0.015)						
Logged sales$_{t-2}$						0.071***			
						(0.015)			
Logged sales$_{t-3}$									0.049***
									(0.015)
Logged total patents$_{t-1}$			-0.024***						
			(0.007)						
Logged total patents$_{t-2}$						-0.006			
						(0.007)			
Logged total patents$_{t-3}$									-0.008
									(0.008)
Constant	-1.451***	-1.460***	-1.736***	-1.453***	-1.465***	-1.583***	-1.453***	-1.464***	-1.523***
	(0.023)	(0.003)	(0.033)	(0.023)	(0.003)	(0.033)	(0.024)	(0.003)	(0.033)
Fixed effects	none	Firm & industry-year		none	Firm & industry-year		none	Firm & industry-year	
Observations	29,301	23,343	22,788	28,450	22,589	20,893	27,333	21,572	18,838
adj. within R2	0.02030	0.00083	0.03007	0.02079	0.00062	0.00765	0.02042	0.00046	0.00408

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

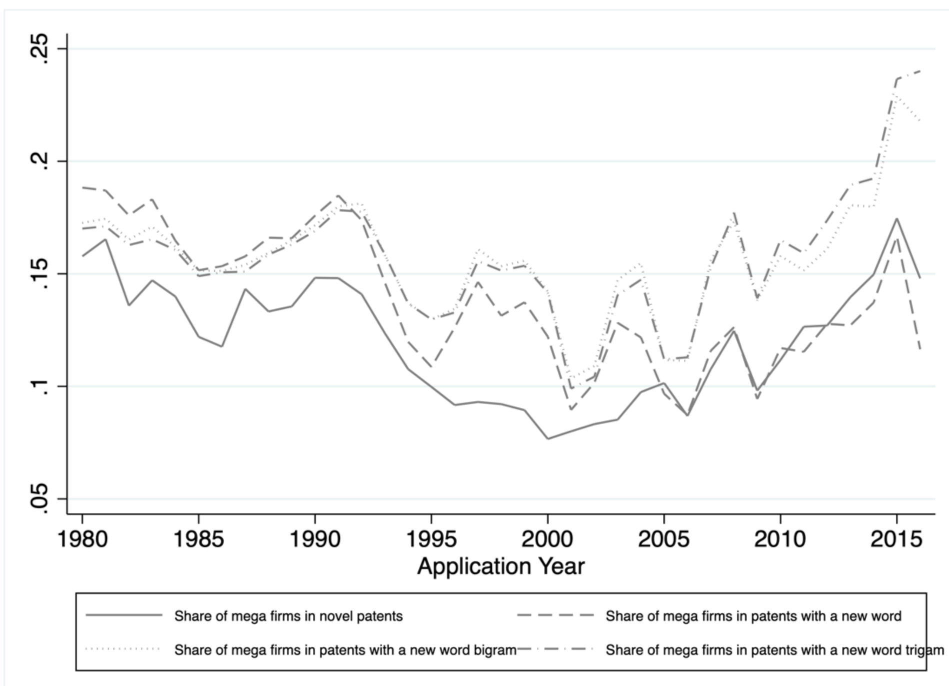
Figure A3. Share of mega firms in novel patents identified using NLP methodology

Panel A.



The solid line depicts the share of mega firms in novel patent applications defined as in the main text. The dashed line depicts the share of mega firms in novel applications below the median backward cosine similarity constructed by Arts et al. (2021). Source: Authors' own calculation using the USPTO data matched with Compustat data.

Panel B.



The solid line depicts the share of mega firms in novel patent applications defined as in the main text. The dash, dot, and dash-dot lines depict the share of mega firms in novel applications based whether the patent contains at least one novel word, novel bigram, or novel trigram, respectively (see Arts et al., 2021). Source: Authors' own calculation using the USPTO data matched with Compustat data.