NBER WORKING PAPER SERIES

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

Serguey Braguinsky
Joonkyu Choi
Yuheng Ding
Karam Jo
Seula Kim

Working Paper 31460
http://www.nber.org/papers/w31460

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2023, Revised August 2023

Nothing to disclose. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research, the Korea Development Institute, or the Federal Reserve Board of Governors or its staff.

Mega Firms and Recent Trends in the U.S. Innovation: Empirical Evidence from the U.S. Patent Data
Serguey Braguinsky, Joonkyu Choi, Yuheng Ding, Karam Jo, and Seula Kim
NBER Working Paper No. 31460
July 2023, Revised August 2023
JEL No. L10,O30,O32,O33

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 importance of mega firms in novel patents had been declining until about 2000, it has strongly rebounded since then. The timing of this turnaround coincided with the ascendance of firms that newly became mega firms in the 2000s, and a shift in the technological contents, characterized by increasing integration of Information and Communication Technology (ICT) and non-ICT components. Mega firms also generate a disproportionately large number of "hits"—novel patents that lead to the largest numbers of follow-on patents (subsequent patents that use the same combinations of technology components as the first novel patent)—and their hits tend to generate more follow-on patents assigned to other firms when compared to hits 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.

Serguey Braguinsky
University of Maryland
Robert H. Smith School of Business and
Department of Economics
4558 Van Munching Hall
College Park, MD 20742
and NBER
sbraguinsky@rhsmith.umd.edu

Joonkyu Choi
Federal Reserve Board
20th ST & Constitution Avenue N.W.
Washington, DC 20551
joonkyu.choi@frb.gov

Yuheng Ding
University of Maryland
Robert H. Smith School of Business
3330 Van Munching Hall
College Park, MD 20742
yuheng@umd.edu

Karam Jo
Korea Development Institute
263, Namsejong-ro, Sejong-si 30149
South Korea
karamjo@gmail.com

Seula Kim
Princeton University
20 Washington Road
Princeton, NJ 08544
sk6285@princeton.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 (Covarrubias et al., 2020; Gutiérrez 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. Akcigit and Ates (2023, forthcoming) show that, like the increase in market concentration, the share of patents held by the top one percent firms in patent stocks 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 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. In the baseline specification in this paper, mega firms are the top 50 firms by sales in any given year among all public firms in the Compustat data.[2] Second, we examine mega firms innovation not just through the prism of all patents, but utilizing also 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). Following extant studies (e.g., Fleming et al., 2007; Verhoeven et al., 2016), we define novel patents as those that introduce new combinations of technological components that had never been utilized together before.[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 discuss the sensitivity of our findings to alternative measures of patent novelty below.

Figure 1 illustrates two examples of novel patents as defined in this paper. 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

---

[2] We discuss below alternative definitions of mega firms, such as the top 50 firms in terms of sales after excluding non-patenting firms, or top one percent firms in sales in two-digit NAICS industries 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."

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" 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 had been declining for almost 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. Furthermore, we show that mega firms are more likely to apply for novel patents even after controlling for various firm characteristics including size, industry, and the total 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, 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

[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."

combination as first introduced by a novel patent—to measure the degree of success of a new combination. It turns out that mega firms generate a disproportionately large number of "hits"—new combinations that lead to the largest numbers of follow-on patents—especially in recent years. We also examine the opposite side of the spectrum, the not-yet-public VC-backed startups and find that those also play a disproportionately large role in generating novel patents, especially "hit" novel patents, so that successful novel patents appear to be produced in a bi-modal pattern, both by super large mega firms and relatively small startups, not even in Compustat data. Relatedly, among mega firms, an outsized role in generating "hit" novel patents in recent years belongs to those that became mega firms more recently (some VC-backed startups themselves in the 1990s).

Third, we find some big changes in the technological content of new combinations that have been driving the recent resurgence in novel patents, compared to that of new combinations that underpinned novel patents in the previous decades. Most novel patents in the 1990s involved new combinations of Information and Communications Technology (ICT) components. Since the mid-2000s, however, most novel patents involve 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 example above).

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

[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, 2023, forthcoming). While this is a reasonable

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, those 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 of the triggers that led to 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 could not 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 the 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.

hypothesis, subsequent changes in patent ownership are outside the scope of our analysis as we focus on the initial applicants for novel patents.

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 venture-backed startups for which the information 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 trajectories of new technological combinations. We describe detailed matching procedures with Compustat and VentureXpert data 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

[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.

levels of classifications.[8] The 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 a pairwise combination of technological components that appears in a patent for the first time. 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 of a novel patent as subsequent patents that (re-)use the same combination of technological components as introduced by the novel patent. Specifically, we count the cumulative number of patented inventions that re-use the new technological combination up to

[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.

20 years following the appearance of the focal new combination or up to the end of 2020. Furthermore, we distinguish follow-on patents assigned to the same versus different assignee(s) from the assignee(s) of the novel patent. 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 introducing 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 (2020). We first identify ICT industries based on 4-digit NAICS by following Goldschlag and Miranda (2020). 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 (Table A2). The ALP extracts key words, i.e., search terms, from each NAICS industry description and combing through all the patent texts to retrieve patents that contain the exact phrases of each search term. Then, the underlying CPC subclass of the retrieved patents is 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 (IPC) and the OECD definition of ICT-related products (Inaba and Squicciarini, 2017).

3. Mega Firms and Novel Patents

3.1 The Share of Mega Firms in Novel Patents Has Been Increasing 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, such as top 50 patenting Compustat firms or top one percent in sales in each two-digit NAICS industry. 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]

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 also increases 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.[12] 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

---

[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 (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 publicly available 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 stay very robust. Details are available upon request.

[12] Akcigit and Ates (2023, forthcoming, Figure 9) show a secular trend toward increasing concentration of patents at the top one percent of patenting firms, and we have confirmed that the same holds in our data. The difference with Figure 2 is due to changes in the composition of top one percent of patenting firms over time. See Appendix A.4.

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. In Appendix A.4, we show that this trend is even more strongly pronounced among mega firms in most actively patenting industries. In the same Appendix A.4, we discuss why we find seemingly different results for the importance of mega firms compared to those reported in Akcigit and Ates (2023, forthcoming).

We next utilize our Compustat-USPTO matched firm panel data to examine more formally 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, 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 Table 1, 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. To see how the pattern varies across different periods, we include the interaction term between the mega firm indicator and a dummy variable for the post-2007 period as follows:

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

The estimation results are shown in the last two columns of Table 1, Panel (a). The coefficient on the interaction term, 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 firm 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 pass the threshold from (to) a non-mega firm to (from) a 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, with mega firms playing a major role.

The findings with firm fixed-effects also suggest a relationship between novel patents and firm performance. To examine this 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.[13] 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 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 A4 and A5 in the Appendix).

[Table 2 around here]

Note that these estimation results just show a correlation and should not be interpreted as indicating causality, even though we control for firm fixed effects. For example, both the increase in novel patenting and the subsequent increase in sales can 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 Novel Patents Had Been Declining but Are on the Rise Once Again

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

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

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 the NLP-based measures of novel patents, and it suggests that our measure based on the co-assignment of CPC main groups captures different aspects of patent novelty. For example, it is possible that new combinations underpinning novel patents (under the definition adopted in this paper) do not qualify as dissimilar enough from previous patents using the NLP methodology because new combinations are likely to combine existing knowledge. However, novel patents described in this paper are still important from the economic point of view, as they are most related to Schumpeter's (1911) definition of innovations as "new combinations." See also Section 4.1 below.

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 (see Appendix A.5). 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 Mega Firms and VC-backed Startups Generate Most Impactful New Combinations

One interpretation of novel patents and new combinations of technological components underpinning 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.[14] 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 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

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

[14] An alternative measure widely used in the literature is the number of forward citations. This measure is less straightforward in our context than follow-on patents because a novel patent may not necessarily be cited for the new combination it contains. Nevertheless, we checked the robustness of our findings below to utilizing this alternative measure of the impact of a novel patent and found broadly similar results. See Appendix A.7 for details.

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 in Table 3 show that new combinations by mega firms were less likely than other patenting entities to generate follow-on patents during the baseline period, but the opposite was 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 them 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 the largest increase in the impact of those new combinations. Furthermore, if we break down mega firms into those that newly became mega firms in the post-2007 period and those that had appeared as mega firms even before 2007, the trend toward generating more follow-on patents and lower probability of not having any follow-on patents in post-2007 is mostly driven by the former category and to a less extent by mega firms who had been identified as such in earlier periods. See Appendix A.6 for details.

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 each application year cohort, 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 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 new combinations. It is worth noting that several firms that became mega firms in 2007 or later are in our sample of VC-backed startups in 1991-2000.[15]

[15] These include Apple and Google, two of the GAMAM firms (tech Giants including Google, Amazon, Meta (Facebook), Apple, and Microsoft). Amazon also became a mega firm in the post-2007 period.

[Figure 4 around here]

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

4.1 New Combinations Have Been Shifting from Within-ICT to ICT-Non-ICT

What is behind the changes in the time trend of novel patents in recent years? In this section we show that a big shift 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. This 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 A7 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.[16]

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,[17] 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 component 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 could 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

[16] 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.
[17] 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.

sections during that period but especially in Sections G (physics) and H (electricity) which dominate ICT-related main groups (Table A2 in the Appendix).

[Figure 5 around here]

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.[18]

[Figure 6 around here]

The color in Figure 6 (and the interactive animation in the link) indicates different CPC sections. For example, the yellow color denotes technology groups in the H (electricity) section, while the 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

[18] The following link: https://www.yuhengding.com/about/vis_most_successful_new_combinations provides the interactive animation of Figure 6.

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 the 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 Mega Firms Contribute 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 slowdown in new knowledge diffusion from leading to laggard firms (e.g., Akcigit and Ates, 2023, forthcoming). The extant analysis, however, has relied on indirect measures or inference.

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— whether those are assigned to the same firm as the original new combination assignee or 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 slowdown 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 surged, as seen in Figure 3 above.

[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).[19] We repeat the same exercise for the subsample of top 1% of new combinations in terms of follow-on patents they generated within each application year cohort 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

[19] 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.

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.

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 5-year mark, while the midpoint of the diffusion curve is not reached until about the 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 takes 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 firms is much less pronounced among hits. This is not surprising, of course, as we are selecting on success. 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 fewer 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 also be leading technological experimentation by introducing new technology combinations and enabling 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 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.

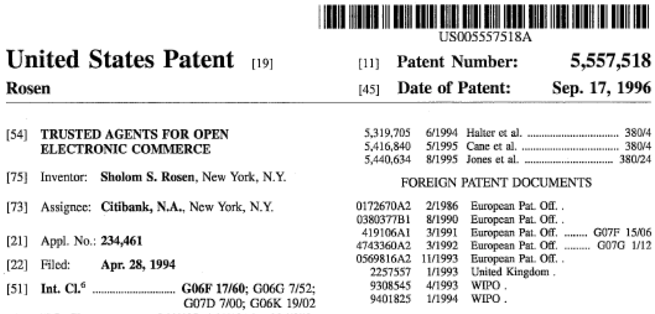
Verhoeven, D., Bakker, J., & Veugelers, R. (2016). Measuring technological novelty with patent-based indicators. *Research Policy*, *45*(3), 707–723.

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]
Rosen

[11] Patent Number: 5,557,518
[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

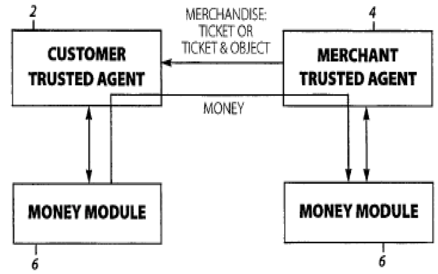
[22] Filed: Apr. 28, 1994

[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
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)

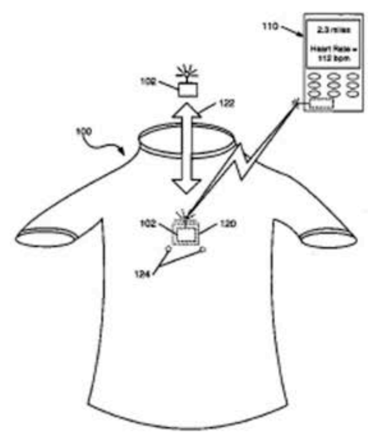
FOREIGN PATENT DOCUMENTS

EP 0 589 607 A1 9/1993
(Continued)

OTHER PUBLICATIONS

First Office Action issued in corresponding Chinese Application, Application No. 200780018533O, issued Jul. 2, 2010.
(Continued)

Primary Examiner — Brian Zimmerman



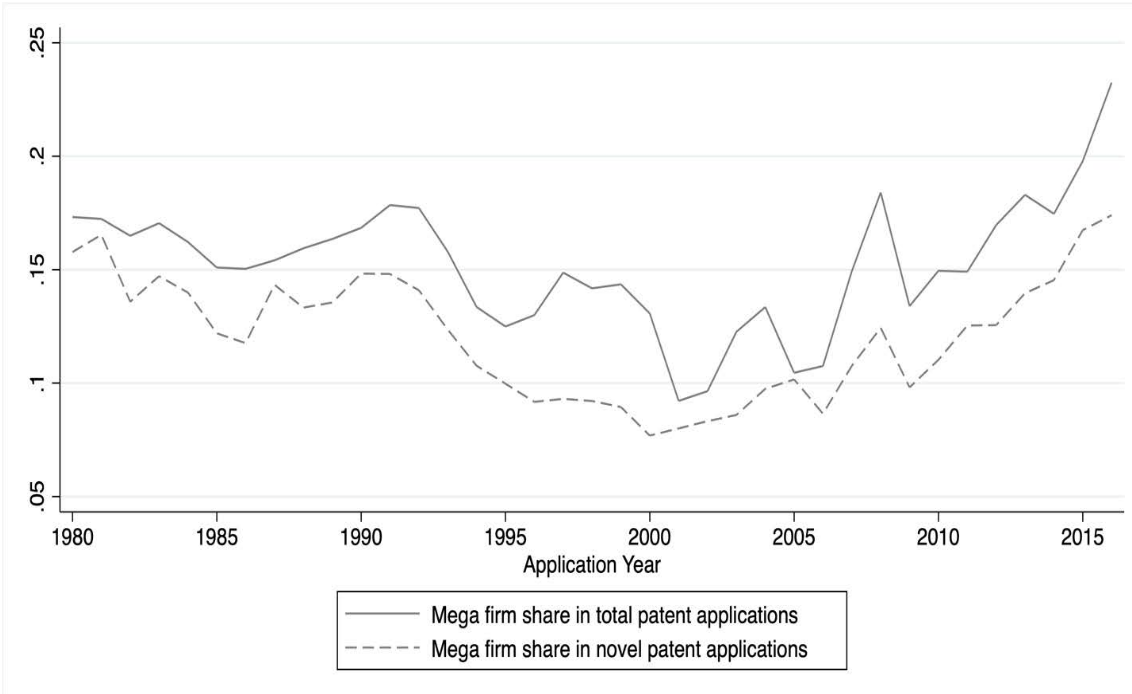
G08C17 - Arrangements for transmitting signals characterized by the use of a wireless electrical link

A43B3 - Footwear characterized by the shape or the use

A41D1 - Garments

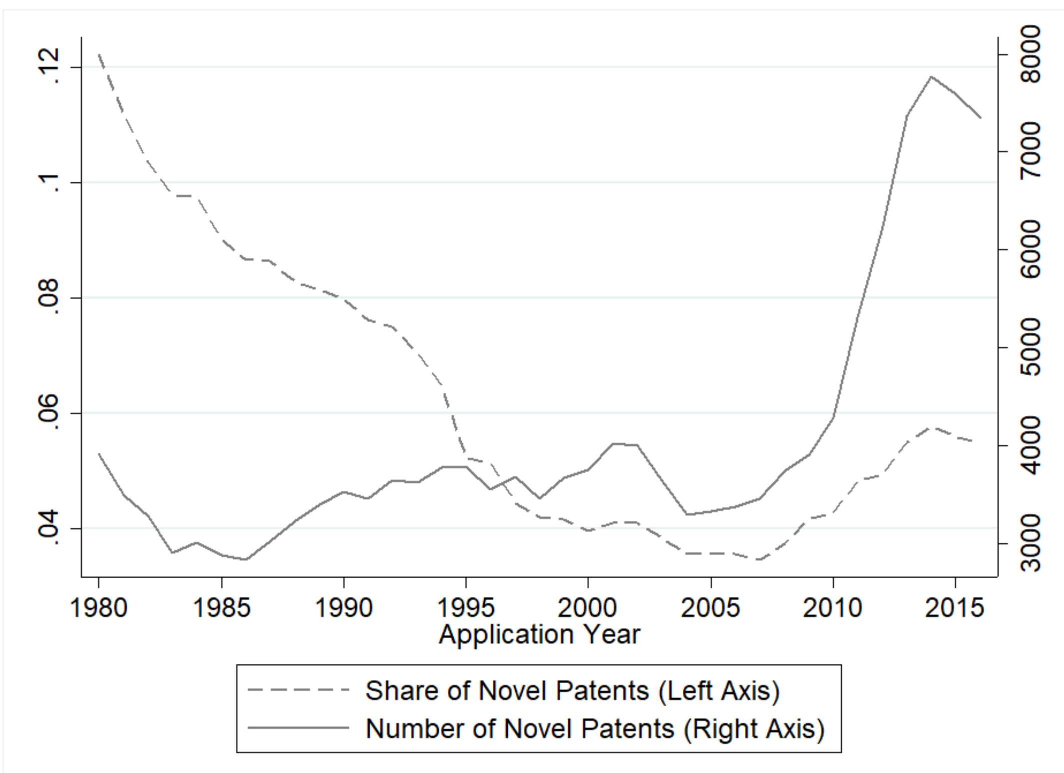
Source: USPTO data.

Figure 2. Share of Mega firms in Total Number of Patents Applications



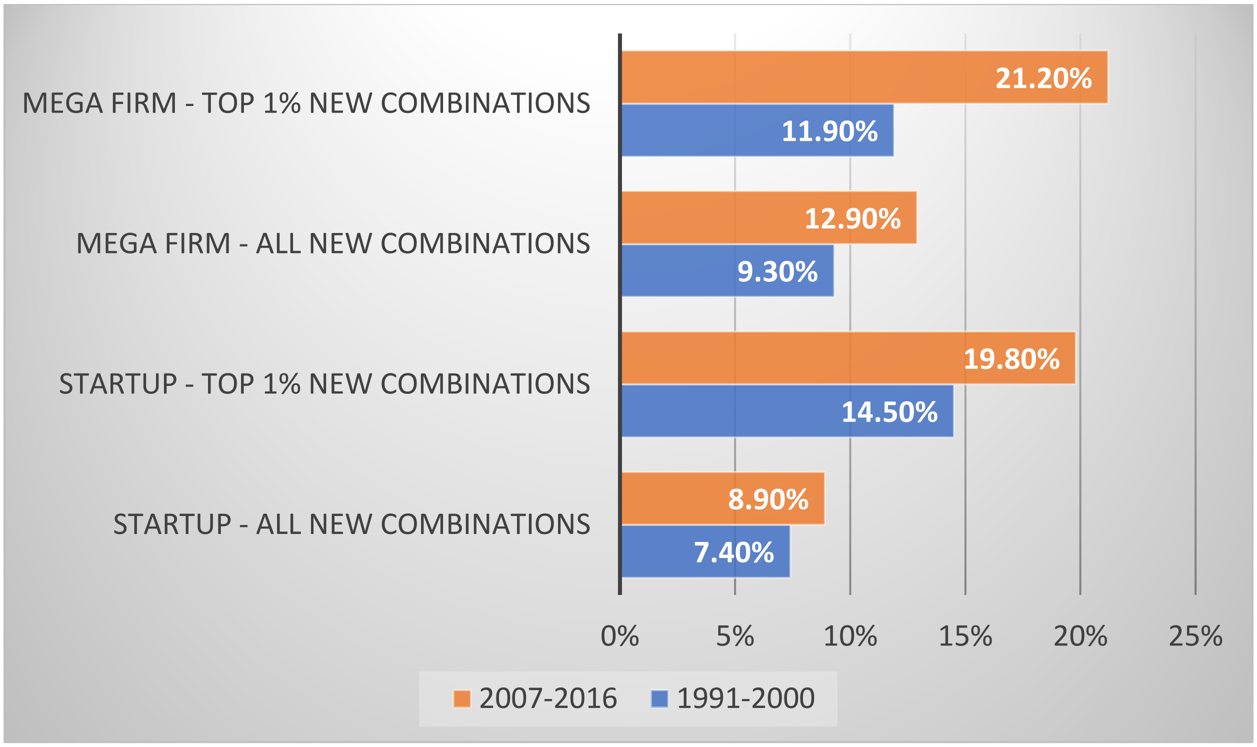
Source: Authors' own calculation using the USPTO matched with Compustat data.

Figure 3. The Number of Novel Patents and Their Share in Total Patent Applications



Source: Authors' own calculation using the USPTO data.

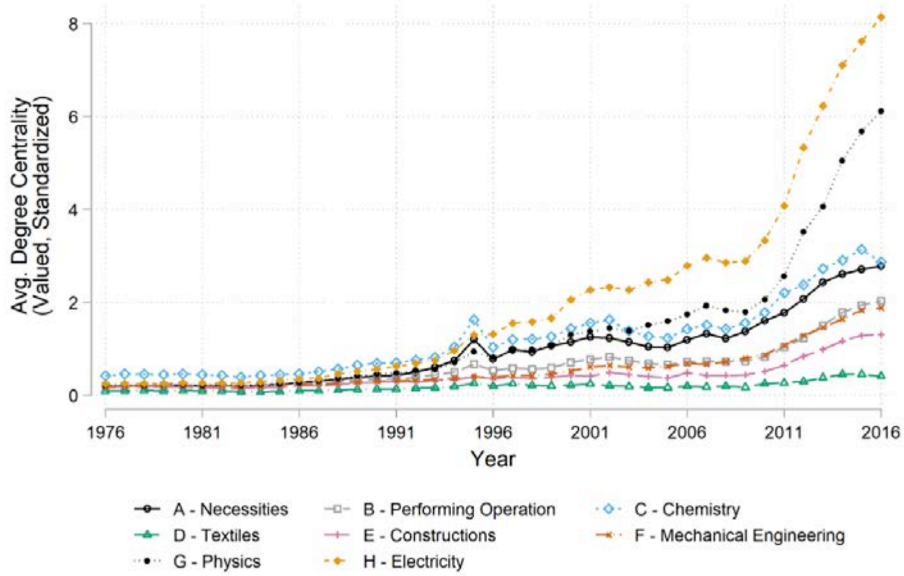
Figure 4. All and Top One Percent New Combinations by Firm Type



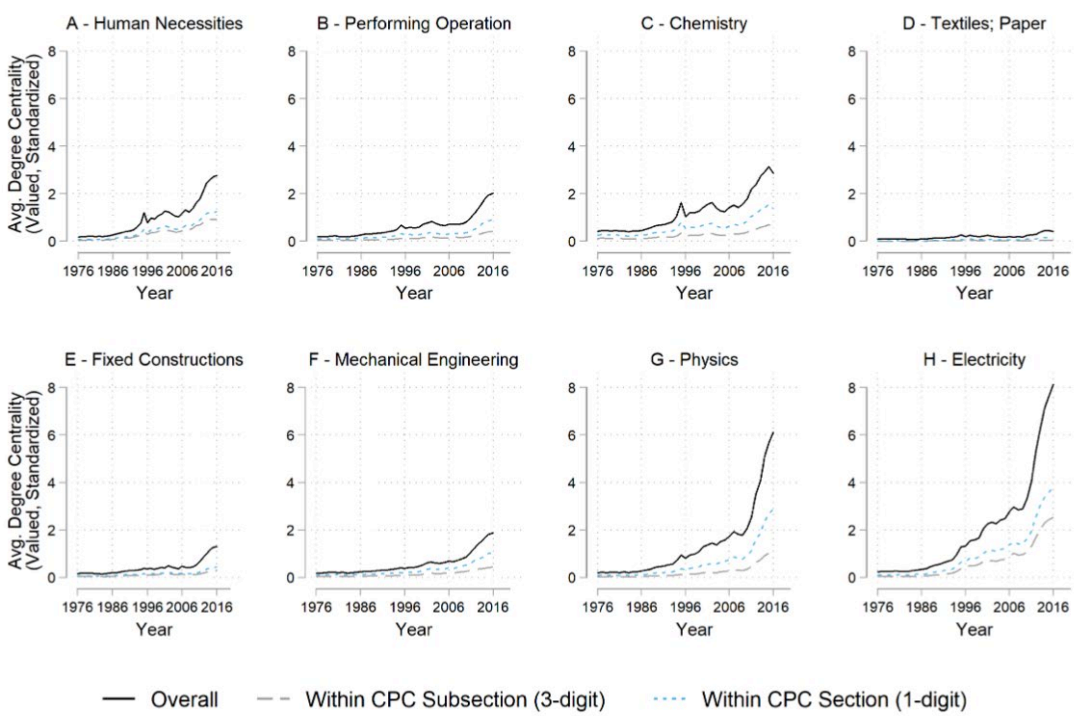
Source: Authors' own calculation using the USPTO matched with Compustat data. Top one percent new combinations: based on the number of follow-on patents using the same technological combinations.

Figure 5. Time Trend in Valued Degree Centrality by CPC Sections

(a) Overall

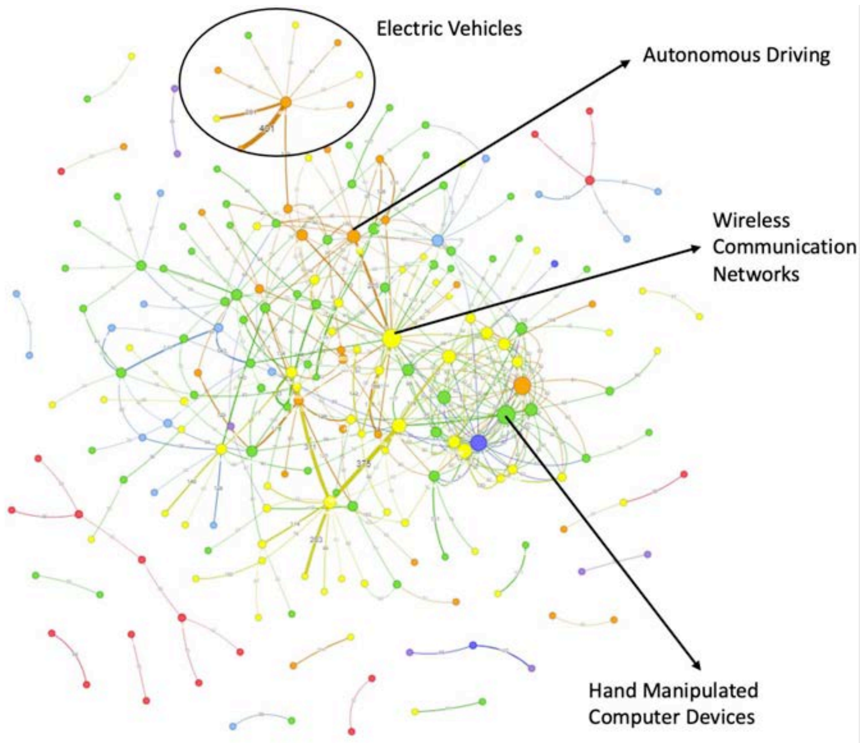


(b) Decomposition into within and between CPC sections



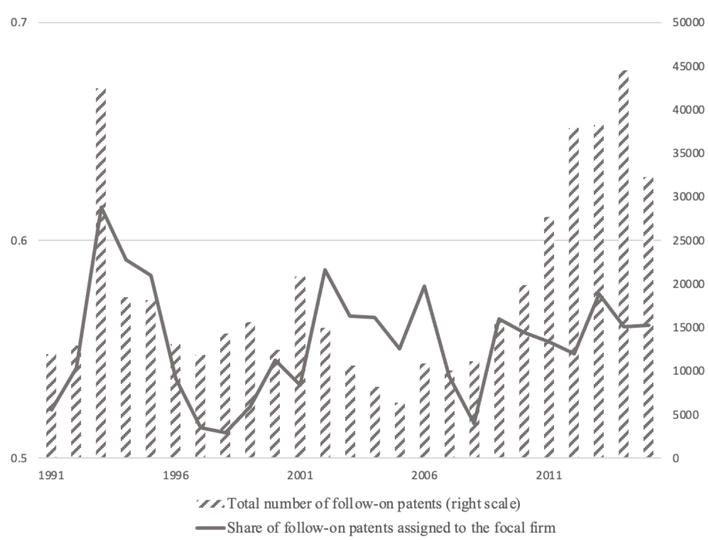
Source: Authors’ calculation using the USPTO data. See the main text for the definition of valued degree centrality.

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). The size of the node indicates the number of other technologies linked to the focal one. Source: created by Yuheng Ding using the USPTO data. Online interactive version: https://www.yuhengding.com/about/vis_most_successful_new_combinations

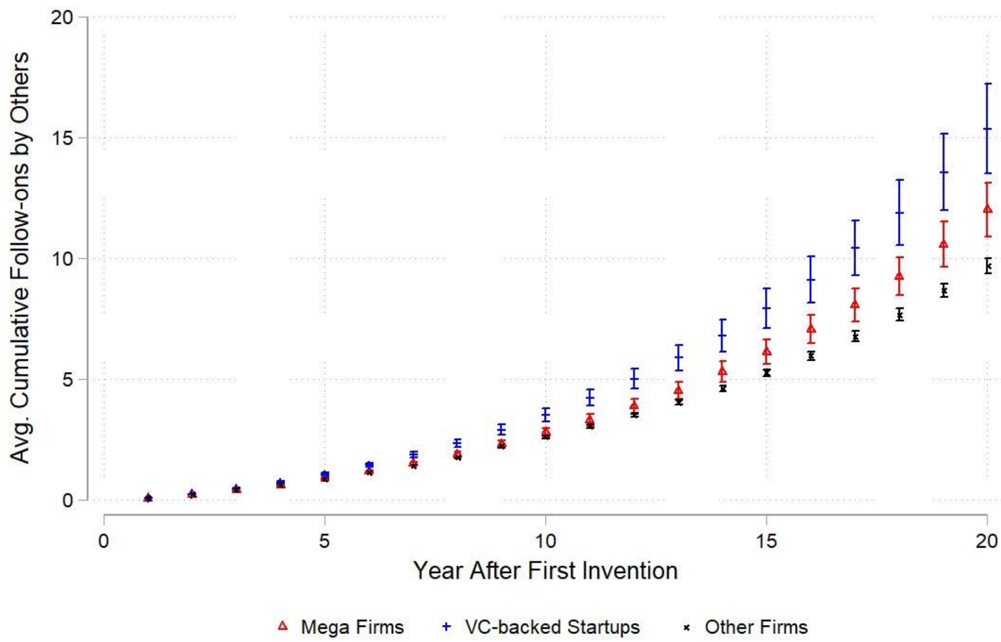
Figure 7. Number of Follow-on Patents and Share Assigned to the Focal Assignee



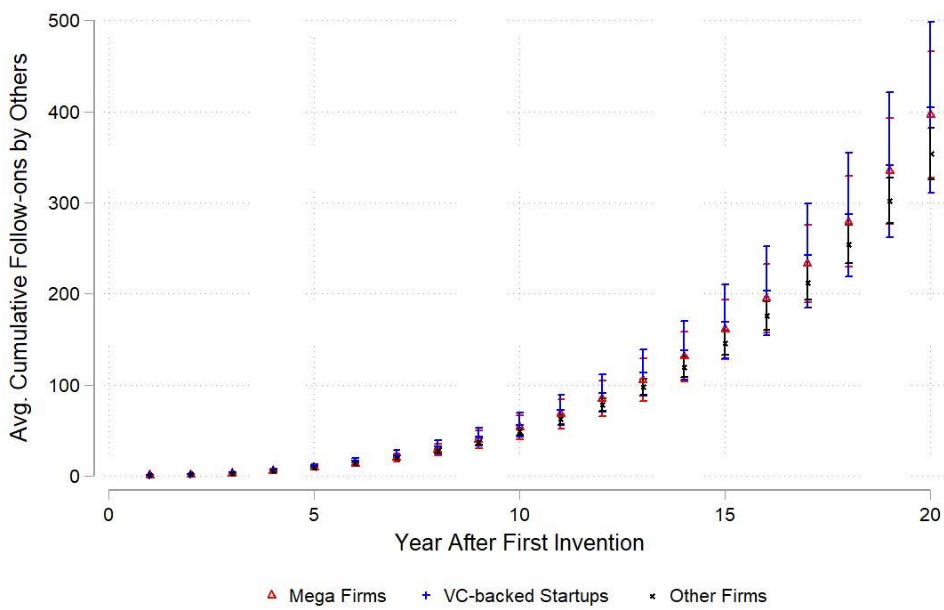
Source: Authors' own calculation using the USPTO data.

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 the 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. Bars represent the 95th percentile confidence intervals.

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	Industry-year	Industry-year	Industry-year	Industry-year
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 & Industry-year	Firm & Industry-year	Firm & Industry-year	Firm & Industry-year
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, ** p < 0.05, * p < 0.1. 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-year		none	Firm & Industry-year		none	Firm & Industry-year	
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, ** p < 0.05, * p < 0.1.

Table 3. Follow-on Patents on New Combinations by Mega Firms and VC-backed Startups

	(1)	(2)	(3)
DV:	IHS # follow-on patents	"Failed"	Logged # follow-on patents
VARIABLES			
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***	0.731***
	(0.002)	(0.001)	(0.003)
Year FE	Included	Included	Included
Observations	228,831	228,831	117,720
R-squared	0.086	0.114	0.025

Note: Estimation method: OLS, absorbing year fixed effects. Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1. Follow-on patents within the 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 the USPTO merged with Compustat data.

Table 5. Share of Follow-on Patents Assigned to the Focal Assignee

over the First Five Years in 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						
Dummy equal to one if for the 2007-2016 period	0.001	0.002	-0.008***	0.226***	0.218***	0.158***
	(0.002)	(0.002)	(0.003)	(0.013)	(0.012)	(0.014)
Dummy equal to one if mega firm		-0.023***	-0.036***		-0.083***	-0.055***
		(0.004)	(0.006)		(0.016)	(0.019)
Mega firm X 2007-2016 period			0.023***			-0.028
			(0.008)			(0.031)
Dummy equal to one if VC-backed startup		0.047***	0.008		0.209***	0.041**
		(0.004)	(0.006)		(0.017)	(0.021)
VC-backed startup X 2007-2016 period			0.072***			0.298***
			(0.008)			(0.032)
Constant	0.556***	0.553***	0.558***	0.169***	0.138***	0.168***
	(0.002)	(0.002)	(0.002)	(0.008)	(0.009)	(0.009)
Observations	142,733	142,733	142,733	2,652	2, 652	2, 652
R-squared	0.000	0.001	0.002	0.102	0.171	0.201

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 an 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 stem names (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 an 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. Figure A1 shows the match rates in our bridge over time.

Figure A1. UPSTO-Compustat Match Rates

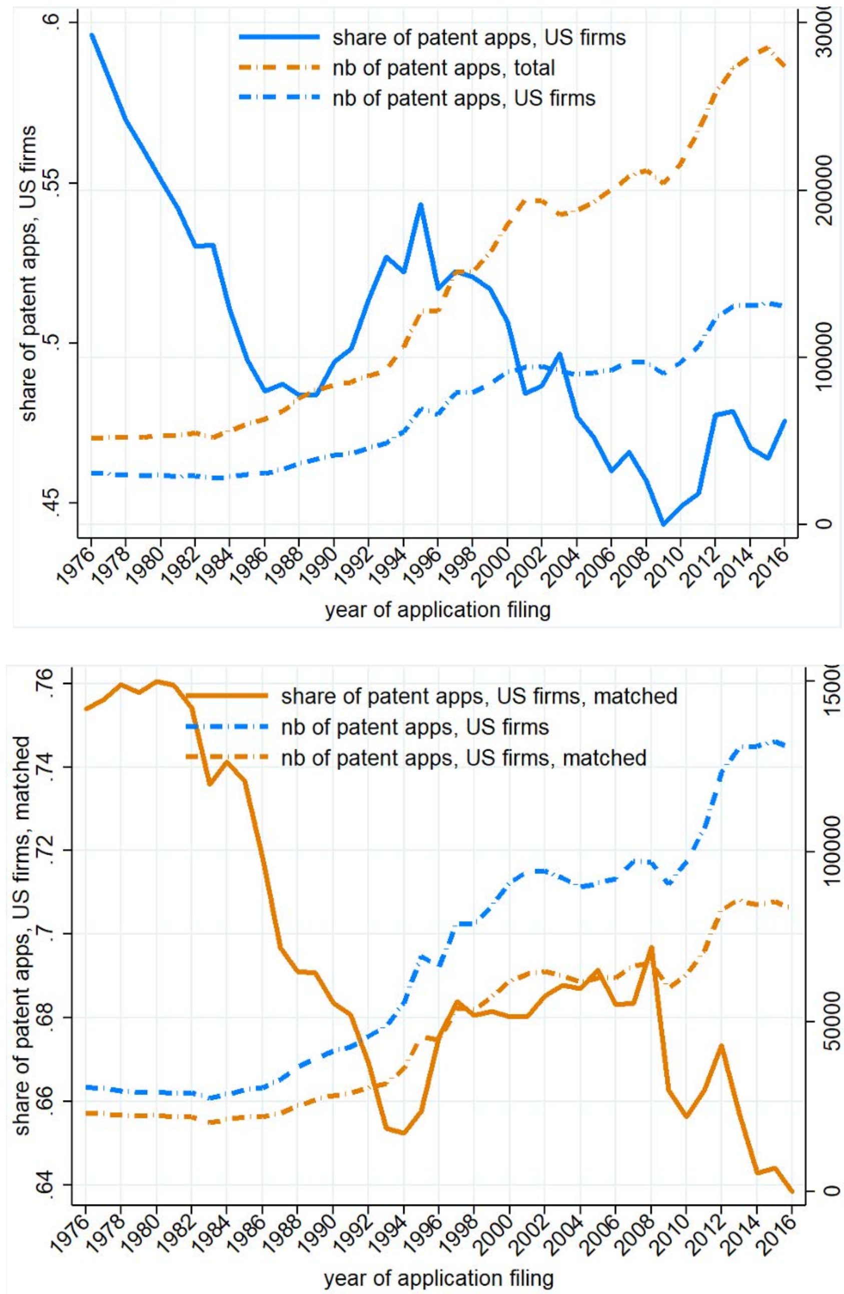
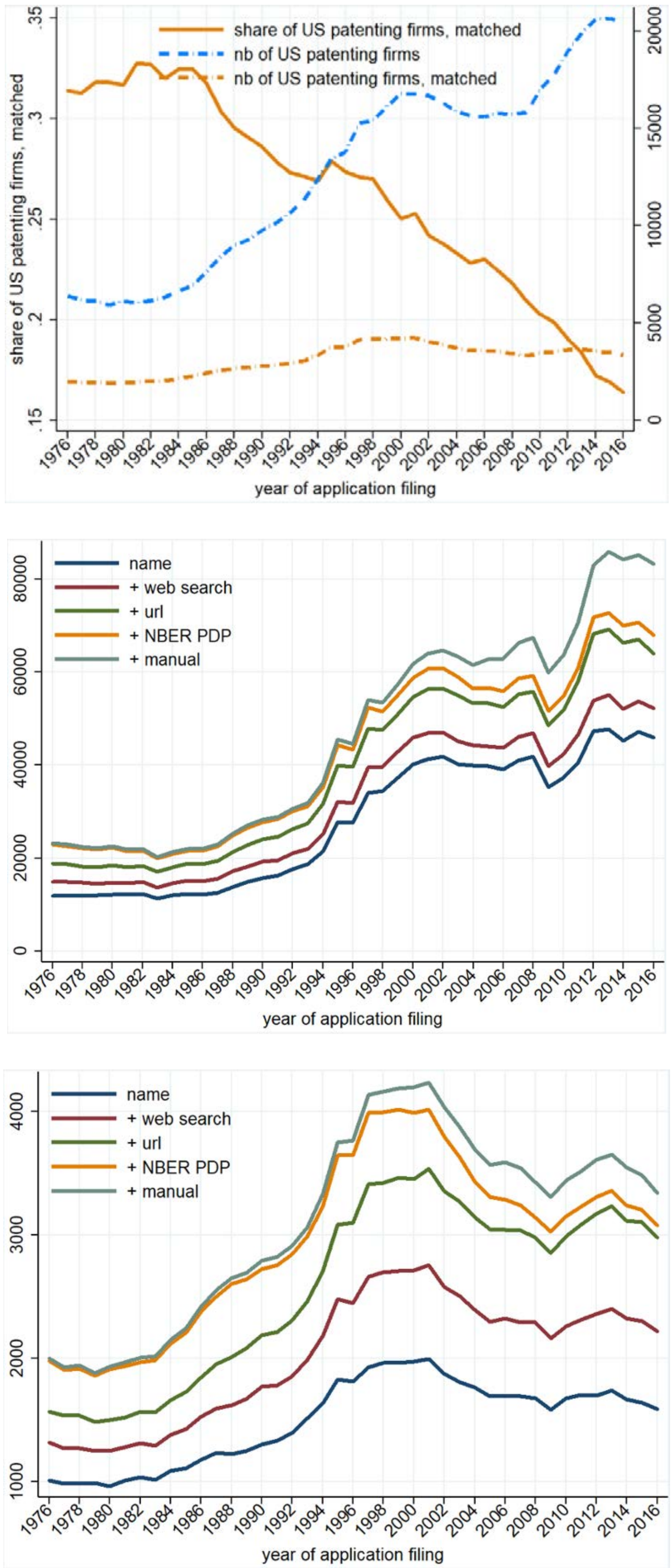


Figure A1. UPSTO-Compustat Match Rates (continued)



A.2 USPTO-VentureXpert Crosswalk and VC-backed Startups

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

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 if they file patent applications between the founding year and the year of exit (i.e., IPO, M&A, or bankruptcy, etc.), or if 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 ICT Industries and Technologies

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 (2020)

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

A.4 Compositional Changes Among Top One Percent Patenting Firms

As mentioned in the main text, Akcigit and Ates (2023, forthcoming, Figure 9) present suggestive evidence of a secular trend toward increasing concentration of patent ownership in the United States as measured by the share of the top one percent of patenting firms. In Figure A2 we present a similar figure, constructed using our data and also adding the time trend in novel patents. The solid line depicts the share of the top one percent of patent assignees in the cumulative number of all patents filed with the USPTO since 1976. Consistent with Akcigit and Ates' Figure 9, this line shows a pronounced increase over time.[20] The dotted line (not in Akcigit and Ates, Figure 9) depicts the same share in cumulative novel patent applications. The time trend is very similar to the overall trend until about 2000 but after that, the two lines diverge—the concentration of the stock of all patent applications among the top one percent of assignees keeps increasing, but their share in novel patents starts decreasing.

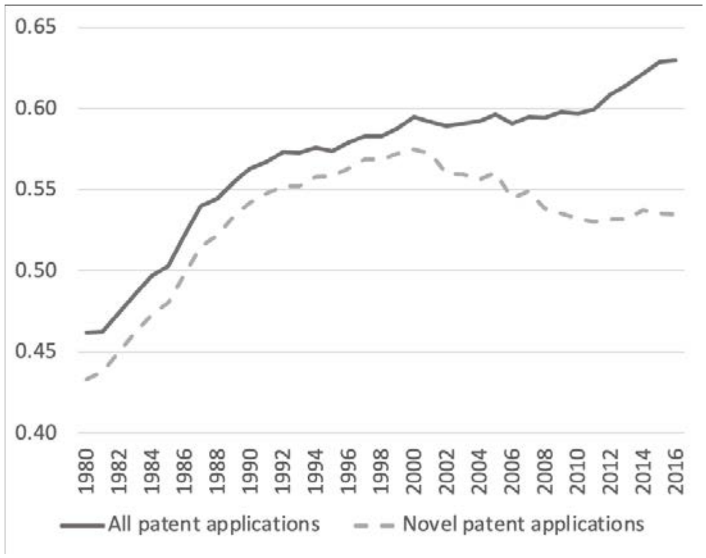
Figure A2 is indeed consistent with the notion that the use of patents in the U.S. is becoming more concentrated; furthermore, the fact that the top one percent assignees account for a smaller stock of novel patents over time appears to vindicate the commonly held view that there is indeed perhaps some "abuse" of patents in the United States. However, as we show in the main text, if we focus our attention on mega firms, we do not observe a secular trend toward increasing concentration of patents in those firms; instead, we observe the decline in their share in patent applications in the 1990s-early 2000s, followed by a robust rebound from around 2007 (Figure 2 in the main text). Furthermore, and most importantly from the vantage point of this paper, the share of mega firms in the total number of novel patent applications exhibits the same trend, so the divergence we see in Figure A2 is not driven by mega firms either.

The differences between Figure A2 and Figure 2 in the main text appear to be due to compositional changes among the top one percent of patenting firms over time. More specifically, as can be seen from Table A3, the decline in the share of mega firms among the top one percent of assignees in the 1990s-first half of 2000s is offset by the increasing share of non-mega firms that are among the top 5% in terms of sales in their 2-digit or 4-digit NAICS industries. However, the share of those ("second-tier-large") firms in the number of novel patent applications increases

[20] We were unable to replicate Figure 9 in Akcigit and Ates exactly because we suspect we apply a somewhat different data cleaning procedure compared to them. Figure A2 also plots the share of the top one percent of all assignees in the USPTO data, not just "patenting firms." Despite these differences, the overall picture is very similar.

much less than their share in all patent applications, and then it starts falling precipitously since 2000, explaining the divergence between the two lines in Figure A2 (see also Figure A3 for a visual depiction).

Figure A2. Share of Patents of the Top 1% Patent Assignees



Source: our calculations using the USPTO data.

One possible explanation for these findings is the widening gap between market leaders ("superstar mega firms") and their immediate followers (large but non-mega firms) who try to secure competitive edge by resorting to strategic patenting. The leaders, in contrast, can afford to do risky experiments.[21] Note that we also found positive association between firm performance and novel patents even after controlling for firm fixed effects (Table 1 in the main text). While we cannot claim any causality from those findings, the fact that mega firms have regained their share among the top one percent patenting firms, especially in novel patents, at the expense of their immediate followers in recent years also suggests that innovation could be key for some mega firms to gain and consolidate their positions, while strategic patenting is being increasingly employed by large but non-mega firms trying to defend.

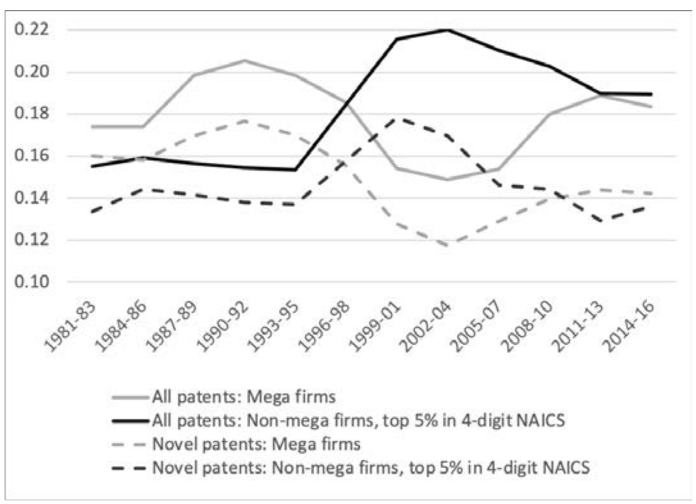
[21] We thank Sina Ates for this suggestion.

Table A3.
Breakdown of Shares in Total Patent Applications by Firm Types

Years	Mega firms		Non-mega firms, top 5% in 2-digit NAICS industries		Non-mega firms, top 5% in 4-digit NAICS industries		Other public firms		Non-public entities	
Patents	All	Novel	All	Novel	All	Novel	All	Novel	All	Novel
1981-83	0.174	0.160	0.122	0.100	0.155	0.134	0.045	0.049	0.030	0.035
1984-86	0.174	0.158	0.143	0.125	0.159	0.144	0.046	0.043	0.030	0.037
1987-89	0.199	0.170	0.144	0.130	0.156	0.142	0.057	0.061	0.034	0.044
1990-92	0.205	0.177	0.146	0.131	0.155	0.138	0.053	0.057	0.044	0.058
1993-95	0.198	0.170	0.152	0.136	0.154	0.137	0.053	0.062	0.048	0.062
1996-98	0.186	0.155	0.181	0.162	0.184	0.158	0.040	0.051	0.054	0.070
1999-01	0.154	0.128	0.221	0.194	0.216	0.178	0.026	0.025	0.058	0.082
2002-04	0.149	0.117	0.204	0.177	0.220	0.170	0.040	0.035	0.062	0.095
2005-07	0.154	0.129	0.197	0.168	0.210	0.146	0.056	0.041	0.052	0.083
2008-10	0.180	0.140	0.187	0.162	0.203	0.144	0.050	0.038	0.052	0.087
2011-13	0.189	0.144	0.180	0.143	0.190	0.129	0.052	0.037	0.060	0.098
2014-16	0.184	0.142	0.181	0.147	0.189	0.136	0.064	0.041	0.064	0.097

Source: our calculations using the USPTO data. "Other pubic firms" are not mega firms and not among the top 5% in terms of sales in 2-digit or 4-digit NAICS industries. The total shares do not sum up to 1 because the top 5% firms in 2-digit and 4-digt NAICS industries are not mutually exclusive categories.

Figure A3. Dynamics of Total/Novel Patent Shares of Top 1% Patent Assignees that are Mega Firms and Non-mega Firms-Top 5% by Sales in 4-digit NAICS Industries



Most patents in the USPTO-Compustat matched data are generated in just a few industries. At the two-digit NAICS classification level, those are 21 (Mining, quarrying, and oil and gas extraction), 32 and 33 (most manufacturing industries, except food and light industries), 51 (Information), and 54 (Professional and technical services). To focus specifically on the picture in those high-patenting industries, we have reconstructed the top one percent of patenting firms as those with cumulative number of applications in the top percentile only among patent assignees in the above two-digit NAICS industries, excluding all other patent applicants (public firms from other industries as well as all non-public patent assignees). All the other variables were constructed in the same way as in the main text—for instance the share of mega firms was constructed by limiting the top one percent patenting firms in the above two-digit NAICS industries to mega firms (top 50 percent in terms of sales overall), the share of top firms in two- and four-digit NAICS industries, by limiting the top one percent of patenting firms in the above two-digit NAICS industries to top percent five percent of firms by sales in those industries, and so on.

In Figure A4 (a, b, c), we show the dynamics of overall concentration of patent applications (all patents and novel patents among them) in the top one percent of patenting firms, the dynamics of the shares of mega firms and non-mega but top five percent of firms by sales in their four-digit NAICS industries in such applications, respectively, within the sample of firms in two-digit NAICS industries 21, 32, 33, 51, and 54 as above.[22]

The dynamics presented in Figure A4 are broadly similar to Figures A2 and A3 but there are some important nuances. First, the overall concentration of patents among the top one percent of patenting firms in those industries remains flat since about 2000 at slightly below 50 percent. Also, Figure A4 (b) shows that the share of mega firms in novel patent applications in these industries is much closer in levels and follows the dynamics of their share in total patent applications even closer than in Figure A2. It thus appears that mega firms in the sample limited to most actively patenting industries conduct relatively even more experimentation with novel technologies than they do in the whole sample. In contrast, as can be seen by comparing Figure A4 (c) with the corresponding lines for non-mega top five percent firms in four-digit NAICS industries in Figure A3, the decline in the share of non-mega top five percent firms in four-digit NAICS industries in novel patents is even more pronounced in high-patenting industries than in the whole sample, and the gap between their total patent applications and novel patent applications in the recent decades is also larger. Thus, once again, we cannot reject the hypothesis that these "also-ran" firms are indeed trying to defend their market positions by resorting to strategic patenting, while curtailing their experimentation with new combinations.

[22] We also include Compustat industry classification 99 (unclassified). Ninety percent of patents in this classification are assigned to General Electric (a mega firm throughout our sample), and the remaining patents also seem to be assigned to firms in manufacturing and/or information sectors.

Figure A4 (a)

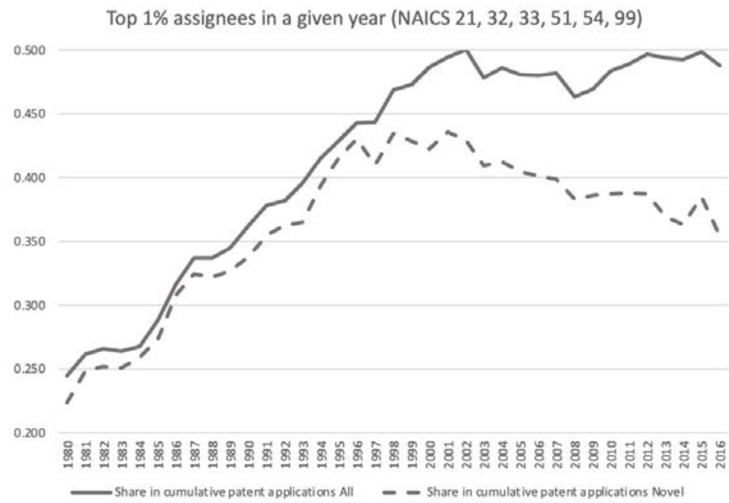


Figure A4 (b)

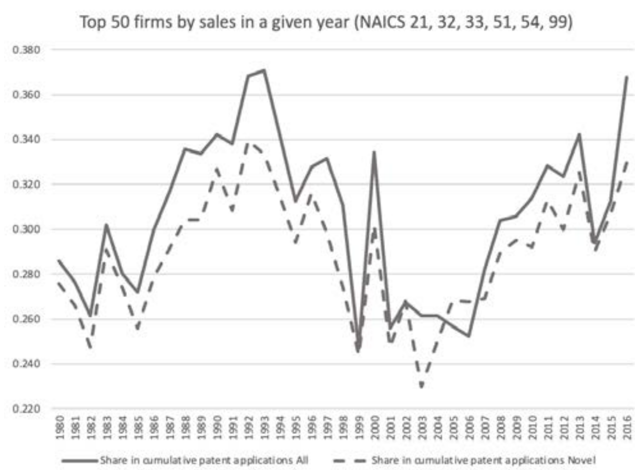


Figure A4 (c)

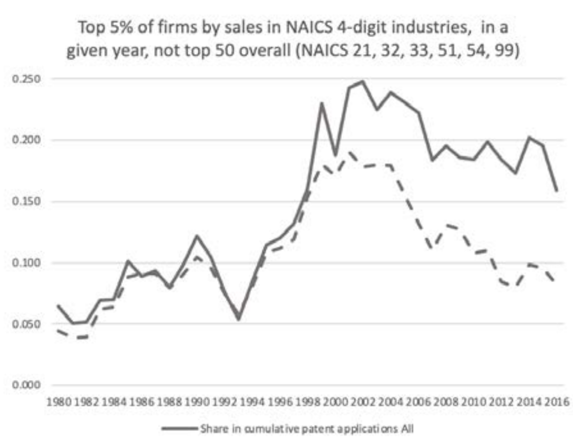


Table A4. Novel Patents and Firm Employment

	(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$_{t-1}$			0.396***						
			(0.015)						
Logged sales$_{t-2}$						0.329***			
						(0.015)			
Logged sales$_{t-3}$									0.280***
									(0.015)
Logged total patents$_{t-1}$			0.099***						
			(0.007)						
Logged total patents$_{t-2}$						0.094***			
						(0.008)			
Logged total patents$_{t-3}$									0.082***
									(0.008)
Constant	0.402***	0.513***	-0.149***	0.430***	0.553***	0.019	0.470***	0.606***	0.180***
	(0.045)	(0.005)	(0.023)	(0.047)	(0.005)	(0.024)	(0.048)	(0.005)	(0.026)
Fixed effects	none	Firm & Industry-year		none	Firm & Industry-year		none	Firm & Industry-year	
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, ** p<0.05, * p<0.1.

Table A5. 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

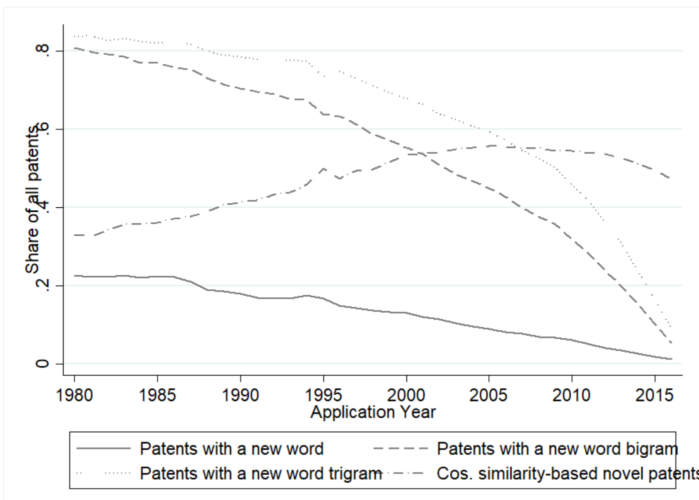
Note: Estimation method: OLS. Robust standard errors clustered at the firm level in parentheses. *** p < 0.01, ** p<0.05, * p<0.1.

A.5. Comparison with NLP-based Measures of Novel Innovation

In this section, we describe the comparison between our measure of novel innovation and the NLP-based measures of novel innovation (Arts et al., 2021; Kalyani, 2022) in more detail. Given data availability, we use the measure developed by Arts et al. (2021) for this comparison.

Arts et al. (2021) identify novel patents using a new word (unigram), new two consecutive words (bigram), new three consecutive words (trigram), and keyword combinations appearing for the first time in the title, abstract, or claims of a patent, as well as calculate the cosine similarity between the technical content of a focal patent and all prior patents. Figure A5 displays the share of novel patents out of all patents using the various NLP-based measures. With the exception of the cosine similarity-based measure, the shares of all NLP-based novel patents have been continuously declining over time. This pattern is also consistent with the trend reported in Kalyani (2022).

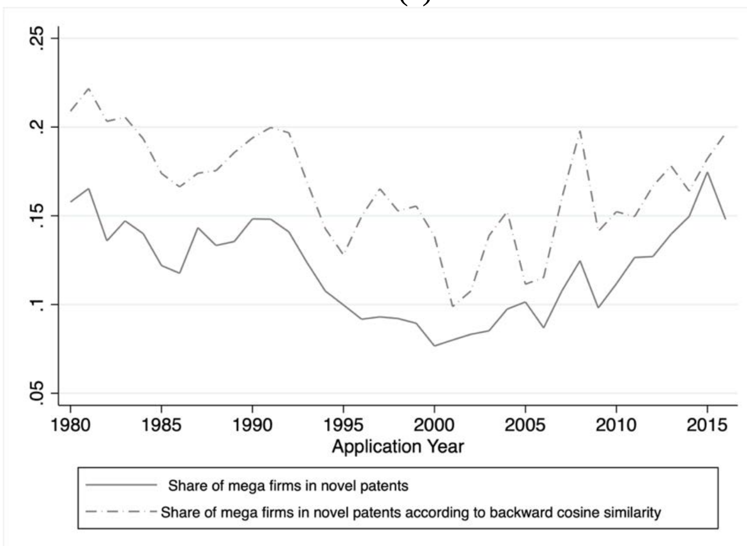
Figure A5. Share of NLP-based Novel Patents out of All Patents



Nonetheless, interestingly, the shares of mega firms in all NLP-based novel patents exhibit time trends very similar to the share of mega firms in our CPC-based measure of novel patents as shown in Figure A6. Therefore, regardless of which measure we use, mega firms have been contributing more to novel patents since the mid-2000s. As mentioned in the main text, the overall trend in our measure of novel patents is different from the overall trends in the NLP-based measures of novel patents as these measures capture different aspects of patent novelty. For example, it is possible that ICT and non-ICT new combinations do not qualify as dissimilar enough from previous patents using the NLP methodology because new combinations are likely to combine existing knowledge. However, these novel combinations are still important from the economic point of view, as they are most related to Schumpeter's (1911) definition of innovations as "new combinations."

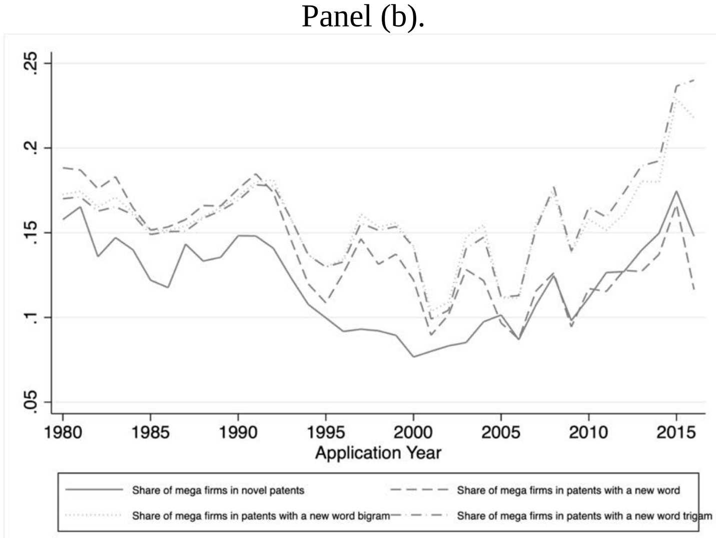
Figure A6. Share of Mega Firms in Novel Patents Identified Using NLP Methodology

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 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 matched with Compustat data.

A.6. Novel Patents by "Old" and "New" Mega Firms

In this section, we further investigate whether the reversal trends in post 2007 are driven by a specific group of mega firms. First, we divide the mega firms into the following two groups: i) those who have ever been identified as mega firms before 2007 (pre-2007 mega firms) and ii) those who have not appeared as mega firms before 2007 but newly identified as mega firms only since 2007 (post-2007 mega firms). Then, we decompose the previous set of new combinations by mega firms into those generated by the pre- and post-2007 mega firms.

We extend the regression model in section 3.3 by including $I_{\{pre-2007\ mega\ firm\}t}$, which indicates new combinations created by a pre-2007 mega firm (a mega firm at t who has also been identified as a mega firm before 2007). Note that this leads β_2 to capture the impact of new combinations generated by a post-2007 mega firm (a mega firm at t who has never been a mega firm before 2007) and β_3 to indicate the impact of those made by a pre-2007 mega firm. [23]

$$\begin{aligned} y_t = \alpha + \beta_1 I_{\{mega\ firm\}t} + \beta_2 I_{\{mega\ firm\}t} X\{2007 - 2016\ period\} \\ + \beta_3 I_{\{pre-2007\ mega\ firm\}t} X\{2007 - 2016\ period\} + \beta_4 I_{\{VC\}t} \\ + \beta_5 I_{\{VC\}t} X\{2007 - 2016\ period\} + \delta_t + \varepsilon_t \end{aligned}$$

Table A6 shows the estimation results indicating that the reversals of the trends seen in Table 3 in the main text for the three dependent variables are mainly driven by patents created by post-2007 mega firms, while pre-2007 mega firms play a much lesser role in this reversal. Indeed, the coefficient β_3 on the interaction term between the pre-2007 mega firm and the 2007-2016 period is negative and its magnitude offsets about two-thirds of the magnitude of the coefficient β_2. This suggests that recent changes in the composition of mega firms were an important factor behind the increase in the impact of new combinations generated by those firms.

[23] Note that $I_{\{pre-2007\ mega\ firm\}t}$ by itself is eventually omitted due to its multicollinearity with (the combination of) $I_{\{mega\ firm\}t}$, $I_{\{mega\ firm\}t} X\{2007 - 2016\ period\}$, and $I_{\{pre-2007\ mega\ firm\}t} X\{2007 - 2016\ period\}$.

Table A6. Follow-on Patents on New Combinations by the Pre-2007, Post-2007 Mega Firms and VC-backed Startups

	(1)	(2)	(3)
DV:	IHS # follow-on patents	"Failed"	Logged # follow-on patents
VARIABLES			
Dummy equal to one if mega firm	-0.115***	0.049***	-0.068***
	(0.011)	(0.006)	(0.014)
Mega firm X 2007-2016 period	0.359***	-0.151***	0.242***
	(0.022)	(0.011)	(0.027)
Pre-2007 Mega firm X 2007-2016 period	-0.220***	0.098***	-0.133***
	(0.020)	(0.010)	(0.024)
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	0.045***	-0.024***	0.049***
	(0.015)	(0.008)	(0.017)
Constant	0.773***	0.493***	0.731***
	(0.002)	(0.001)	(0.003)
Year FE	Included	Included	Included
Observations	228,831	228,831	117,720
R-squared	0.086	0.115	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: $y = ln\left(x + \sqrt{x^2+1}\right)$.

A.7 Alternative Measures of the Impact of Novel Patents

In the main text, we studied the impact of new combinations by examining follow-on patents, that is, inventions using the same pair of technology combinations as the novel patent that first introduced it. As noted there, it is the most straightforward way of doing so, but here we briefly examine what happens if we employ another way of measuring the patents’ impact widely adopted in the literature, namely forward citations.[24]

Forward Citations of New Combinations by Mega Firms and VC-backed Startups

To check the robustness of the findings in the main text related to how successful mega firms were in generating follow-on innovation and how that changed over time, we replace the number of follow-on patents—those that are using the same combinations as the novel patent—by the number

[24] This is a much less direct method for our purposes, as it does not distinguish between the citations that are related to the new combinations first introduced by the novel patent or something else that might be unrelated to the new combinations itself.

of forward citations received by each new combinations and estimate otherwise the same regressions as presented in Table 3 in the main text. The results are presented in Table A7.

Table A7. Forward Citations on New Combinations by Mega Firms and VC-backed Startups

	(1)	(2)	(3)	(4)	(5)	(6)
	All Forward Citations			Of which: From Follow-on Patents		
DV:	IHS # Forward Citations	No Forward Citations	Logged # Forward Citations	IHS # Forward Citations	No Forward Citations	Logged # Forward Citations
VARIABLES						
Dummy equal to one if mega firm	-0.085***	0.006	-0.080***	-0.116***	0.069***	-0.150***
	(0.015)	(0.005)	(0.015)	(0.007)	(0.005)	(0.020)
Mega firm X 2007-2016 period	0.097***	-0.018***	0.071***	0.080***	-0.029***	0.269***
	(0.018)	(0.006)	(0.019)	(0.008)	(0.006)	(0.025)
Dummy equal to one if VC-backed startup	0.502***	-0.015***	0.497***	0.171***	-0.082***	0.214***
	(0.017)	(0.005)	(0.016)	(0.008)	(0.005)	(0.017)
VC-backed startup X 2007-2016 period	-0.043**	-0.103***	-0.105***	-0.030***	-0.001	-0.054***
	(0.020)	(0.007)	(0.020)	(0.009)	(0.006)	(0.021)
Constant	1.575***	0.327***	1.592***	0.250***	0.807***	0.477***
	(0.003)	(0.001)	(0.003)	(0.001)	(0.001)	(0.004)
Year FE	Included	Included	Included	Included	Included	Included
Observations	228,831	228,831	155,784	228,831	228,831	44,566
R-squared	0.297	0.242	0.139	0.029	0.027	0.032

Note: Estimation method: OLS, absorbing year fixed effects. Robust standard errors in parentheses. *** p < 0.01. Forward citations within first five years after grant year. IHS is inverse-hyperbolic sine transformation: $y = ln\left(x + \sqrt{x^2 + 1}\right)$.

The outcome variable in column (1) is the total number of citations received by the focal new combination within five years since the novel patent was granted. The results are qualitatively similar to those in column (1), Table 3 in the main text, except that new combinations created by VC-backed startups on average seem to become somewhat less cited between 2007-2016 as compared to the 1990s. The coefficient estimate, however, is only marginally statistically significant. The results are not sensitive to the IHS transformation and simple log transformation as shown in column (3). Also consistent with the findings in the main text, new combinations created by both mega firms and VC-backed startups are less likely to not being cited by any patents (column (2)). Further restricting the outcome to forward citations from follow-on patents (those that also re-use the same new combination(s)) produces qualitatively similar results, as shown in columns (4)–(6).

The Shift of "Hit" New Combinations from ICT & ICT to ICT & Non-ICT

In Table 4 in the main text, we showed a big shift in technology components underlying the most successful ("hit") novel patents from the 1990s to the late 2000s from ICT & ICT to ICT & non-ICT new combinations. What happens if we redefine "hit" combinations as those receiving the most forward citations within 5 years since the novel patent had been granted? To see this, we now define a new combination to be a "hit" if the 5-year forward citations are ranked among the top one percent among all new combinations created in the same year. As before, we then categorize new combinations into three categories including ICT & ICT, ICT & non-ICT, and non-ICT & non-ICT, and compare the compositional changes between 1991-2000 and 2007-2016.

Table A8 shows that while 21.2% of all "hits" in the 1990s combined ICT and ICT components, this share goes down to 5.5% among the most highly cited ones between 2007-2016. Meanwhile, the share of ICT & non-ICT combinations among the top one percent most highly cited novel patents increased from 41.1% in the 1990s to 44.4% between 2007-2016. This shift in the type of highly cited new combinations is less than observed in Table 4 in the main text but it is notably more pronounced among mega firms, with 21.4% and 45.2% of hits being ICT & ICT and ICT & non-ICT in the 1990s, to 5.6% and 53.4% between 2007-2016, respectively.

Table A8. Types of 1% Most Cited New Combinations: 1991-2000 and 2007-2016

		Period			
		1991-2000		2007-2016	
Mega firms	ICT & ICT	9	21.43%	18	5.56%
	ICT & non-ICT	19	45.24%	173	53.40%
	non-ICT & non-ICT	14	33.33%	133	41.05%
	Total	42	100.00%	324	100.00%
All assignees	ICT & ICT	170	21.17%	96	5.48%
	ICT & non-ICT	330	41.10%	778	44.38%
	non-ICT & non-ICT	303	37.73%	879	50.14%
	Total	803	100.00%	1753	100.00%

Source: Authors' calculation using the USPTO and Compustat data.

We can further restrict forward citations received by the novel patent to those that also use at least one of its new combinations. That is, we redefine "hits" as those ranked top 1% among all new combinations that are created in the same year in terms of the number of forward citations received in the first five years, that **also** re-use at least one of its new combinations. As can be seen from Table A9, the share of ICT & non-ICT combinations increases dramatically among "hit" novel patents in this definition, from 20.5% in the 1990s to 31.1% between 2007-2016, and the changes among hits created by mega firms are even more pronounced, from only 6.5% to 59.6%.

Table A9. Types of 1% Most Cited and Re-used New Combinations: 1991-2000 and 2007-2016

		Period			
		1991-2000		2007-2016	
Mega firms	ICT & ICT	4	8.70%	12	4.90%
	ICT & non-ICT	3	6.52%	146	59.59%
	non-ICT & non-ICT	39	84.78%	87	35.51%
	Total	46	100.00%	245	100.00%
All assignees	ICT & ICT	74	7.31%	91	4.79%
	ICT & non-ICT	208	20.53%	591	31.09%
	non-ICT & non-ICT	731	72.16%	1219	64.12%
	Total	1013	100.00%	1901	100.00%

Source: Authors' calculation using the USPTO and Compustat data.

Figure A7.