Mega Firms and New Technological Trajectories in the U.S.[*]

Serguey Braguinsky[†] Joonkyu Choi[‡] Yuheng Ding[§] Karam Jo[¶]
Seula Kim[‖]

April 26, 2025

Abstract

We provide evidence that mega firms have played an increasingly important role in shaping new technological trajectories in recent years. While the share of novel patents—patents introducing new combinations of technological components—produced by mega firms declined until around 2000, it has rebounded sharply since then and the share of novel patents in all patent applications in the U.S. also rebounded accordingly. Additionally, we find that the technological impact and knowledge diffusion from novel patents by mega firms have grown relative to those by non-mega firms after 2001. We also examine potential drivers of this trend, presenting evidence that the rise in novel patenting by mega firms is linked to their disproportionate increase in cash holdings and expansion of their technological scope.

Keywords: Mega Firms, Innovation, Novel Patents, Knowledge Diffusion
JEL Codes: O31, O33, O34, L11, L25

---

[*]Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the World Bank, Korea Development Institute, the Federal Reserve Board of Governors, or its staff. Previously circulated under the title "Mega Firms in Recent Trends in the U.S. Innovation: Empirical Evidence from the U.S. Patent Data." We thank seminar and conference participants at the Workshop on Entrepreneurial Finance and Innovation, 2023 NBER Megafirms and the Economy conference, Ed Snider Center at the University of Maryland, and Hitotsubashi University.

[†]University of Maryland and NBER
[‡]Federal Reserve Board of Governors
[§]The World Bank
[¶]Korea Development Institute
[‖]Pennsylvania State University and IZA

1 Introduction

The concentration of economic activities in the largest businesses, so-called mega firms, has been increasing over the past few decades (Autor et al., 2020; Yeh et al., 2022; Hsieh and Rossi-Hansberg, 2023). Recent literature explores two broad sets of interpretations for this trend. Some studies have emphasized the rise in market power (De Loecker et al., 2020), possibly driven by increasing entry barriers, regulation, and lobbying activities that stifle competition (Gutiérrez and Philippon, 2019; Covarrubias et al., 2020). Other studies have cast doubt on this 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., 2020; Hsieh and Rossi-Hansberg, 2023; Kwon et al., 2024).

A key issue in this debate is the role of mega firms in innovation and knowledge diffusion. Akcigit and Ates (2023) show that, like the increase in market concentration in sales or employment, the share of patents held by the top one percent firms in terms of patent holdings has been on the rise over the past several decades. They suggest that the largest firms may be increasingly building patent thickets that make it difficult for other firms to compete in the technology domain, leading to slower diffusion of knowledge and decline in business dynamism. Alternatively, mega firms may be increasingly becoming leaders in innovation and developing new technologies that could potentially create room for subsequent innovation by other firms. Examining the role played by mega firms in the economy-wide innovation process is important not just from an academic perspective but also because it has major policy implications.

In this paper, we provide new evidence that sheds light on this debate by investigating the trends in innovation activities by mega firms. In particular, we utilize the concept of "novel patents"—the subject of burgeoning research in the technology literature in recent years (e.g., Fleming et al., 2007; Strumsky and Lobo, 2015; Verhoeven et al., 2016). Following previous studies in this literature, we define novel patents as those that introduce new combinations of technological components that had never been utilized together before. Such patents represent economic experimentation and, if successful, may create pathways for new technological trajectories generating new products, adding new features or enhancing quality of existing products. Thus, this concept corresponds to Schumpeter's (1911) definition of innovations as "new combinations" (Epicoco et al., 2022; Pezzoni et al., 2022). We show that our measure of novel patents is strongly correlated with a variety of patent-level measures of novelty and creative destruction (e.g., Akcigit and Kerr, 2018; Arts et al., 2021; Kelly et al., 2021; Kalyani, 2024; Jo and Kim, 2024). Also, focusing on novel patents helps us filter out many patents that may be filed for purely strategic reasons (Noel and Schankerman, 2013; Torrisi et al., 2016).

One example is U.S. patent No. 8,188,868 titled "Systems for activating and/or authenticating electronic devices for operation with apparel" applied for in 2006 by Nike, which combines CPC group G08C17 with CPC groups A43B3 and A41D1 for the first time.[1] 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. Another example of a novel patent as defined in this paper is U.S. patent No. 5,557,518 "Trusted agents for open electronic commerce" applied for in 1994 by Citibank. This patent combined CPC groups G06Q30 and H04L63 for the first time, introducing a system that enables anonymous transaction of electronic merchandise.[2] It has facilitated technological advancement in electronic commerce, solving the joint problem of protecting the privacy of buyers and sellers while ensuring the delivery of merchandise and money. Visual images of these two patents are presented in Figure A1 in the Appendix.

Our main finding is that, while the share of novel patents filed by mega firms had been declining for almost two decades since 1980, there has been a robust turnaround since 2001;

[1] 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."

[2] G06Q30 is "Commerce" and H04L63 is "Network architectures or network communication protocols for network security."

by the mid-2010s, the share reached its highest level since our sample began in 1980. A log-difference decomposition indicates that this rise is driven not only by the faster growth in overall patent applications by mega firms compared to non-mega firms, but also the increasing share of novel patents among mega firms' total patents. Firm-level panel regressions confirm that mega firms became more likely to apply for novel patents than non-mega firms after 2001, even after controlling for firms' total number of patent applications and firm size, indicating that this pattern is not solely driven by mega firms becoming larger or by them producing a larger number of patents. This finding also holds within firms: Firms produce more novel patents than before as they become mega firms—particularly after the early-2000s—suggesting that closing on market leadership has become associated with more, not less, novel innovations in recent decades. Furthermore, we show that the share of mega firms in breakthrough patents—those that combine high novelty with high impact—has also risen significantly since the early 2000s. This reinforces our main findings by demonstrating that mega firms are not only generating more novel patents but are also responsible for a growing share of the most influential technological advancements.

Importantly, the rebound in the share of novel patents filed by mega firms coincided with the rebound in the share of novel patents in all patent applications in the U.S. While the number of novel patent applications had been steady in the 1990s and until the early 2000s, while their share in total patent applications had been steadily declining due to increase in the number of total patent applications, the number of novel patent applications more than doubled from 2005-2016 and their share in total patent applications in 2016 reached the levels not seen since 1995.

To further assess the degree of technological impact of novel patents by mega firms, we adopt the measure suggested by previous studies (e.g., Pezzoni et al., 2022) and track the number of "follow-on patents"—the patents that use the same new technology combination as first introduced by a novel patent. We find empirical patterns consistent with novel patents being highly experimental: 42% of novel patents do not have any follow-on patents in the first five years since their grant year, while a small fraction become "hits," i.e., those that generate many follow-on patents and thus have a large impact on shaping new technological trajectories. We also find that, after (and only after) 2001, novel patents generated by mega firms have on average more follow-on patents and are more likely to become hits than those generated by non-mega firms.

One straightforward way to examine the degree of knowledge diffusion stemming from novel patents is looking at the dissemination of the follow-on patents beyond the focal firm that generated the novel patent. We find that "self-follow-on rate"—the proportion of follow-on patents generated by the same firm that initially produced the novel patent—is similar for mega firms and non-mega firms in both pre-2001 and post-2001 periods, while we find some evidence of a slowdown in knowledge diffusion from novel patents produced by non-mega firms after 2001. Together with mega firms' novel patents generating more follow-on patents in recent years, this finding 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 been understudied in the literature.[3]

An important question is what drives the rise in novel innovation by mega firms since the early-2000s. To shed light on this, we develop two hypotheses and explore the supporting empirical evidence. First, we hypothesize that, given the highly experimental nature of novel innovation, firms with larger cash holdings can afford to engage more in such innovation, and that mega firms have been increasingly holding more cash or equivalent market securities relative to non-mega firms since the early-2000s. This hypothesis is consistent with findings in the finance literature that U.S. firms have been increasing their cash holdings in part to fund risky R&D activities (Bates et al., 2009; Brown et al., 2009).

We find empirical evidence consistent with this hypothesis. To begin with, firms with larger cash holdings are indeed more likely to engage in novel innovation. Using a local

[3]Patent reassignment and acquisitions may be another way for mega firms to defend their technological leadership (Akcigit and Ates, 2023). Subsequent changes in patent ownership are outside the scope of our analysis as we focus on the initial applicants for novel patents.

projection method, we show that a one percent increase in cash holdings (controlling for total assets) leads to a gradual rise in novel patent applications, reaching a five percent increase after two years. While total patent applications also increase, the magnitude of the effect is less than half that observed for novel patents, suggesting that cash holdings are particularly associated with novel innovation.

Examining aggregate trends, we find that the gap in cash holdings between mega and non-mega firms exhibits a U-shaped pattern, with a reversal in the late 1990s. This gap narrowed by as much as 50% of its 1980 value through the late 1990s, but experienced a strong rebound thereafter. By 2016, the final year of our sample, the difference in cash holdings had grown to 15%-30% above its 1980 level. Taken together, these findings support the idea that the rise in cash holdings among mega firms after 2001 facilitated their engagement in novel innovation.

Second, we investigate whether the increased concentration of inventors in the largest firms noted in the literature (Akcigit and Goldschlag, 2023), may be behind the increase in the share of mega firms in novel patents in recent decades. We find that the average inventor team size has indeed been increasing among mega firms relative to non-mega firms, but this trend is observed across all four decades of our observations and thus cannot, by itself, explain the decline in the share of novel patents produced by mega firms in the 1980s-1990s, followed by a turnaround after 2000. We then examine the trend in the relative scope (diversity) of inventor teams' technological expertise between mega and non-mega firms while controlling for inventor team size and find that mega firms experienced a relative decrease in technology scope of its inventor teams in the first two decades of our data, but their technological scope relative to non-mega firms has been on the increase in the more recent decades. Thus, it appears that broadening technological scope, not sheer inventor team size may have given mega firms a new competitive advantage in terms of novel patents, consistent with the definition of novel patents as those combining different technological components that had not been combined together before.

Our findings have important policy implications. If it is true that mega firms are stifling innovation and slowing down knowledge diffusion, there may be a scope for regulatory intervention. If, however, those firms are among the key actors conducting experiments and generating new technological trajectories, then such an approach may backfire. With the U.S. technological dominance facing increasing global challenges, the stakes could not be higher. We provide further discussion in the concluding section.

Relation to the Literature Our paper engages with the ongoing debate surrounding the rise of mega firms and its implications for market dynamics and innovation. While prior research has extensively documented the increasing concentration of economic activities in large firms (Autor et al., 2020; Yeh et al., 2022; Hsieh and Rossi-Hansberg, 2023) and proposed competing explanations—ranging from increased market power (De Loecker et al., 2020; Gutiérrez and Philippon, 2019; Covarrubias et al., 2020) to intensified competition driven by globalization and technology (Autor et al., 2020; Hsieh and Rossi-Hansberg, 2023; Kwon et al., 2024; Foster et al., 2022)—our study shifts the focus to the role of mega firms in innovation, particularly through the lens of novel patents. By analyzing mega firms' contributions to novel technological combinations, our study sheds new light on whether their growing influence enhances or hinders technological progress and whether their market position is driven by innovative capabilities or by market power that may limit competition.

Our paper is also related to the literature on innovation which has recently raised concerns over declining efficiency of R&D investment (Bloom et al., 2020) and the general declining trend in the U.S. innovation creativity (Kalyani, 2024). Since this trend and the afore-mentioned trend toward increasing dominance of mega firms happened concurrently in time, this has led to the examination of the interrelationship between the two trends, with much of the literature focusing on the stifling effect of the rise of mega firms on innovation through strategic patenting and a slowdown in knowledge diffusion (Akcigit and Ates, 2023), increasing concentration of inventors in the largest firms (Akcigit and Goldschlag, 2023), and "killer acquisitions" (Cunningham et al., 2021). Some more recent studies suggest, however, that part of the advantages of mega firms may actually arise from the nature of technology (Gupta et al., 2024). Our main contributions to this body of literature lies in (re-)examining the role of mega firms in innovation through the prism of novel patents and over several decades.

Our paper is also related to the strand of literature which has documented growing importance of new combinations of technologies in innovation and economic growth. In particular, this literature has examined the technological origins of novel patents that create new knowledge combinations using various measures, most prominently, technological classes assigned to patents by patent examiners (Fleming et al., 2007; Strumsky and Lobo, 2015; Verhoeven et al., 2016; Epicoco et al., 2022). The most closely related paper is Pezzoni et al. (2022) which also looks at follow-on patents (trajectories of new combinations). Additionally, recent studies have leveraged natural language processing (NLP) techniques to develop more nuanced measures of patent novelty and impact, such as those by Arts et al. (2021), Kelly et al. (2021), and Kalyani (2024). These advancements provide valuable benchmarks for assessing the technological significance of patents, which we utilize to validate and compare our measure of novel patents. Most of these papers, however, are limited to examining the technological novelty and the role of recombined knowledge in patents' impact and do not look into who creates new combinations. We link this literature to the above literature strand on the role of mega firms to examine how much such firms contribute to novel patents and what type of new potential technological trajectories they generate to better understand their changing role in the economy-wide innovation process.

The remainder of the paper is structured as follows. Section 2 describes the data and key measurement methods. Section 3 examines the aggregate trend in mega firms' contributions to novel innovation, along with firm-level evidence. Section 4 analyzes the technological impact and knowledge diffusion of novel patents by mega firms compared to non-mega firms over time. Section 5 explores potential drivers of the rise in novel patents by mega firms. Finally, Section 6 concludes.

2 Data and Measurement

2.1 Data Construction and Key Measurement Methods

The primary data sources are the USPTO PatentsView and S&P's Compustat. The USPTO PatentsView tracks all patents ultimately granted by the USPTO from 1976 onward. We collect utility patents granted to U.S. assignees between 1976 and 2023 to track economy-wide innovation activities, and in particular, the creation and trajectories of new technological combinations. We describe detailed matching procedures in the Appendix A.2.[4]

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. 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."

We utilize the Cooperative Patent Classification (CPC) to identify technological components of inventions. The CPC scheme is a hierarchical system with multiple levels of classifications. 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.[5] 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.[6]

[4]While we use our own USPTO-Compustat bridge in this paper, we confirmed that our basic findings are robust to using the DISCERN (Duke Innovation & Scientific Enterprises Research Network) bridge.

[5]Our findings are robust to using different levels of aggregation as well as the IPC classification.

[6]Section Y represents a new addition to patent classifications introduced together with CPC, for general

Following previous studies (Fleming et al., 2007; Strumsky and Lobo, 2015; Verhoeven et al., 2016), 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 defined as 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 first year available in the USPTO PatentsView data is 1976, we do not observe the history of technological combinations created before then. We use the first three years, 1976-1979, as a buffer period and we track novel combinations starting from 1980. In practice, 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.

Note that, by construction, a novel patent may contain multiple new pairwise technological components. While novel patents rely on new pairwise combinations of technological components, we aggregate them to the patent level and use novel patents as our measure of novel innovation, in particular because the number of pairwise new combinations is more likely to overestimate the true number of novel innovations than the measure aggregated to the patent level. Nevertheless, we have verified that all our findings remain robust to using new pairwise combinations as the measure of novel innovation.[7]

To study the diffusion and technological trajectories of new combinations, as well as to develop a measure of novel patents' impact, 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 in years following the appearance of the focal new combination, up to late 2023. Furthermore, to gauge the extent

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.

[7] Results are available upon request.

to which subsequent innovation occurs beyond the boundaries of the focal firm that generated the novel patent, we differentiate follow-on patents that are generated by the focal firm versus those generated by other firms. Occasionally, patents are assigned to multiple assignees. In such cases, we say a follow-on patent is generated by the focal firm if it is one of the assignees.

2.2 Validation of Measurement

While the concept of new technological combinations and the measurement of novel patents based on this idea have been developed in the literature since at least 2007, more recent studies incorporating Natural Language Processing (NLP) techniques have introduced alternative measures of patent novelty and impact, including but not limited to creative destruction. In this section, we briefly examine the relationship between the novel patent measure used in our paper and these alternative measures. Later, we discuss how these measures capture the contributions of mega firms compared to our approach.

Specifically, we first investigate the relationship between our measure of novel patents and three alternative measures. The first two are both based on NLP techniques. The first measure, developed by Arts et al. (2021), extracts the number of new bigrams—compared to all patents filed since 1969—from the title, abstract, and claims of a patent. The second measure is the patent creativity measure due to Kalyani (2024), which is defined as the share of creative technical bigrams—where creativity is determined by these bigrams not appearing in any patent filed in the previous five years—relative to all technical bigrams. The third measure we investigate is backward self-citation ratios, defined as the number of citations to a firm's own previous patents divided by the total number of citations. The endogenous growth literature (e.g., Akcigit and Kerr, 2018; Jo and Kim, 2024) often uses this variable to capture innovation associated with creative destruction—where firms pursue new ideas that expand beyond their existing technological scope to replace incumbents in other markets. A higher self-citation ratio is typically linked to incremental improvements in a firm's existing technology, whereas a lower self-citation ratio suggests a shift toward novel areas.

In Table 1 we use patent-level data from 1980 to 2016 to estimate the relationship between our novel patents measure and those alternative measures. Specifically, we estimate regressions where the dependent variables capture various aspects of patent novelty and creative destruction, and the main independent variable is our novel patent indicator. The results in Column (1) show a strong positive relationship between our measure and that in Arts et al. (2021): we find that while non-novel patents have on average 1.5 new bigrams, novel patents have an additional 0.8 new bigrams.[8] Similarly, the results in Column (2) indicate a strong positive relationship between our measure and the patent creativity measure in Kalyani (2024). Lastly, the results in Column (3) show that novel patents are associated with lower backward self-citation ratios. All regressions include CPC group fixed effects and application year fixed effects, but the results remain robust to their exclusion or to alternative specifications with different combinations of fixed effects, as shown in Appendix A.3.

Table 1: Novel Patents and Other Measures of Novelty and Creative Destruction

	(1)	(2)	(3)
	# New Bigrams	Patent Creativity	Backward Self-citation Ratio
Novel Patent	0.769***	0.020***	-0.027***
	(0.019)	(0.000)	(0.001)
Constant	1.515***	0.068***	0.121***
	(0.003)	(0.000)	(0.000)
CPC Group FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Obs.	2,630,141	2,575,446	2,657,952
R-sq	0.06	0.12	0.06

Notes: The estimates are obtained from OLS regressions with multi-way fixed effects. "# New Bigrams" indicates the number of new bigrams in patent text, obtained from Arts et al. (2021) and "Patent Creativity" is the share of new technical bigrams in patents, obtained from Kalyani (2024). Backward self-citation ratio is defined as the number of citations to a firm's own previous patents divided by the total number of citations. The regression estimates are derived from all patents that were applied between 1980 and 2016. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In the following sections, we will discuss the role of mega firms in generating novel patents

[8] Arts et al. (2021) also extract the number of new words and new trigrams in a patent. The regression coefficients remain similarly large and positive when using these alternative measures. They also develop a measure of backward cosine similarity, for which no meaningful difference is observed. Results are available upon request.

as well as the impact their patents have on follow-on patents. Anticipating this, in Table 2 we also present the results of estimating the relationship between our measure of novel patents and several other measures of patents' impact. Specifically, we once again estimate regressions where the dependent variables capture various measures of a patent's impact while the main independent variable remains our novel patent indicator. Column (1) uses the stock market valuation of patents (expressed as the log of real dollar values) from Kogan et al. (2017), which reflects the patent's contribution to a firm's current and future profits. The results indicate a modest but positive effect, with novel patents valued approximately 2% higher than non-novel patents. Column (2) examines five-year forward citations, the most widely used measure of scientific impact. On average, novel patents receive 0.7 more citations than non-novel patents, which receive an average of 5.4 citations.

Lastly, Column (3) utilizes the "breakthrough" patent indicator from Kelly et al. (2021). This measure combines a measure of patent novelty with a measure of its impact. As such, Kelly et al. (2021)'s measure is conceptually most closely related to the discussion of novel patents by mega firms and their impact in Sections 3 and 4 below, while being constructed in a totally different way. In constructing their measure, Kelly et al. (2021) also leverage NLP techniques to first quantify textual similarity between patent pairs, identifying a patent as novel if its content is distinct from prior patents and impactful if it is similar to future patents. They then compute the ratio of impact to backward similarity (i.e., the inverse of novelty), where higher values indicate distinct advancements at the technological frontier that serve as a foundation for subsequent inventions. The top 10% of patents in the resulting importance distribution are classified as breakthrough patents.[9] The estimation results show that novel patents are 1.5 percentage points (about 15% increase from the mean) more likely to be breakthrough patents than non-novel patents.

[9]Specifically, Kelly et al. (2021) define their measure as the similarity of a patent to all patents filed over the next τ years, divided by its similarity to all patents filed in the previous five years. Their baseline analysis uses $\tau = 10$, though they also construct similar measures for $\tau = 5$ and $\tau = 1$. In our analysis, we use $\tau = 5$ to ensure the measure can be calculated for patents filed through 2016.

Table 2: Impact of Novel Patents

	(1) Market Valuation	(2) 5yr Forward Citation	(3) Breakthrough
Novel Patent	0.021***	0.690***	0.015***
	(0.005)	(0.041)	(0.001)
Constant	1.823***	5.417***	0.099***
	(0.001)	(0.011)	(0.000)
CPC Group FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Obs.	1,504,550	2,838,910	2,836,755
R-sq	0.09	0.04	0.16

Notes: The estimates are obtained from OLS regressions with multi-way fixed effects. "Market Valuation" is the stock market value of a patent obtained from Kogan et al. (2017), and "Breakthrough" is an indicator of whether a patent is a breakthrough patent, obtained from Kelly et al. (2021). The regression estimates are derived from all patents that were applied between 1980 and 2016. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

We also examine how the results in Table 2 vary under different combinations of CPC group and year fixed effects. Controlling for year fixed effects is essential for revealing the positive relationship in Column (1) of Table 2, as it accounts for the time trend in the overall stock prices that are driven by macroeconomic factors such as changes in long-term interest rates or equity risk premium. Similarly, controlling for CPC group fixed effects is necessary to uncover the positive relationships in Columns (2) and (3), as it adjusts for systematic differences in the number of patents across technology groups. Further details are provided in Appendix A.4.

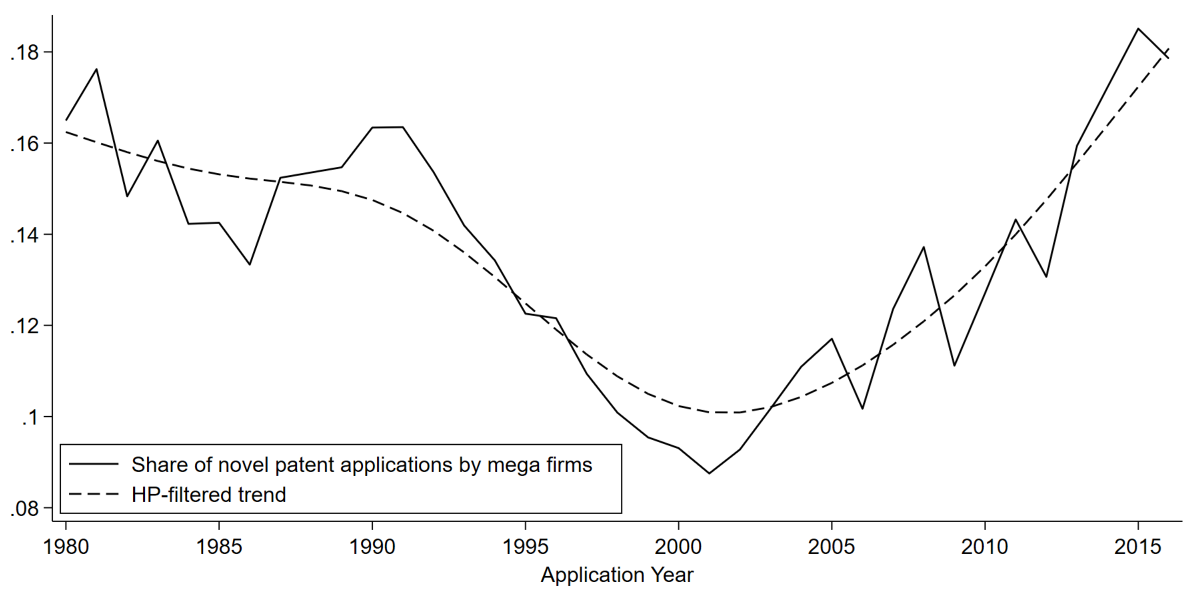
3 Contribution of Mega Firms to Novel Innovation

3.1 Aggregate Trend

We define mega firms as the top 50 firms by sales among all publicly listed firms in the Compustat dataset for each year. Although publicly traded firms represent a selected subset of all firms, this sample selection is unlikely to introduce meaningful bias, as the vast majority of large firms are publicly listed. While these 50 firms constitute only a small fraction of the millions of firms in the U.S., our calculations indicate that, on average, they account for 8% of employment, 14% of sales, and 17% of patent applications annually in the U.S. economy.[10] Thus, they represent a sufficiently large share of business activity and innovation to have a meaningful aggregate impact.

Figure 1 shows the dynamics of the contribution of mega firms to novel innovation. The solid line indicates the share of novel patent applications by mega firms out of all novel patent applications applied by U.S. patent assignees and the dashed line is its smoothed trend. The share displays a U-shaped trend: There had been a prolonged decline from 16 percent in 1980 to nine percent in 2001, followed by a steep increase afterward reaching its historical high level of 18 percent in 2016.

Figure 1: Share of Novel Patent Applications By Mega Firms



Source: Author's own calculation using the USPTO patent data matched with Compustat data.
Notes: The figure presents the share of patent applications filed by mega firms as a proportion of all patent applications in the patent database (solid line), along with its smoothed trend estimated using the Hodrick-Prescott filter with a smoothing parameter of 100 (dashed line). Mega firms are defined as the top 50 firms by sales in Compustat for each year. Note that the vertical axis does not start at zero.

To better understand the sources behind the trend, we decompose the share of novel

[10] Aggregate employment data are obtained from the Business Dynamics Statistics of the Census Bureau, and aggregate business receipts from the Statistics of Income of the Internal Revenue Service. Additionally, within the Compustat dataset, mega firms account for 23% of employment, 31% of sales, and 26% of patent applications.

patents produced by mega firms as follows:

$$\underbrace{\frac{N_{m,t}}{N_{m,t}+N_{o,t}}}_{Y_t\,:\,\text{Mega firm share (novel)}} = \underbrace{\frac{N_{m,t}/T_{m,t}}{(N_{m,t}+N_{o,t})/(T_{m,t}+T_{o,t})}}_{X_{1,t}\,:\,\text{Relative tendency}} \times \underbrace{\frac{T_{m,t}}{T_{m,t}+T_{o,t}}}_{X_{2,t}\,:\,\text{Mega firm share (total)}} \tag{1}$$

where $N_{m,t}$ and $N_{o,t}$ are the number of novel patents applied by mega firms and other firms, respectively, and $T_{m,t}$ and $T_{o,t}$ are the total number of patents applied by mega firms and other firms, respectively. The first term ($X_{1,t}$) in the decomposition represents the relative tendency of mega firms engaging in novel innovation compared to all assignees, and the second term ($X_{2,t}$) represents the share of total patent applications accounted for by mega firms. Taking logs on both sides of equation (1) and time-differencing, we can quantify the contribution of each term in explaining the changes in the share of novel patents by mega firms over time. We find that between 1980 and 2000, the share of novel patents by mega firms (Y_t) declined by 7.2 percentage points, with $X_{1,t}$ and $X_{2,t}$ explaining 77 percent and 23 percent of the decline, respectively. Between 2000 and 2016, Y_t increased by 8.5 percentage points, with $X_{1,t}$ and $X_{2,t}$ explaining 35 percent and 65 percent of the increase, respectively. Therefore, the rise in the share of novel patents by mega firms in recent decades is driven not only by the faster increase in the number of overall patent applications by mega firms (compared to non-mega firms), but also the rise in the share of novel patents among mega firms' total patents. The time series of $X_{1,t}$ and $X_{2,t}$ are shown in Appendix A.5.

We also assess the robustness of our findings in Figure 1 using alternative definitions of mega firms. In Appendix A.6, we demonstrate that the U-shaped trend remains evident when we expand the definition of mega firms to the top 100 firms by sales in each year, or the top 10 firms in each two-digit NAICS sector, totaling 190 firms annually across 19 sectors. The trend also holds when mega firms are defined as the top 50 firms by sales in each year, but only among those that file at least one patent, accounting for the fact that patenting firms represent a highly selected subset.

3.2 Firm-level Evidence

In this section, we employ a firm-year level panel regression framework to assess the robustness of the increase in novel innovation by mega firms since 2000, controlling for various firm characteristics. Specifically, we run a set of Poisson regressions of the following form:

$$E(Y_{it}|X_{it}) = \lambda_{it} = exp(\beta_1 \cdot Mega_{i,t} + \beta_2 \cdot Mega_{i,t} \times Post_t + X'_{it}\gamma + \alpha_i + \eta_t) \qquad (2)$$

where the dependent variable Y_{it} is the number of novel patent applications by firm i in year t, $Mega_{i,t}$ is an indicator whether firm i is a mega firm in t, $Post_t$ is an indicator whether $t \geq 2001$, X_{it} is a vector of firm-level characteristics, α_i is firm fixed effect and λ_t is industry by year fixed effect (industry notation omitted). The independent variable of interest is β_2. Naturally, this regression framework uses a subset of patent assignees that are publicly-listed firms (i.e., exclude privately-owned firms and research institutions) in order to measure their firm characteristics and it compares these firms' novel patenting behavior between 1980-2000 period and 2001-2016 period. Table 3 shows the estimation results.

Table 3: Firm-level Regressions of Novel Patents by Mega Firms in 1980-2016

	(1)	(2)	(3)	(4)
	# Novel Patents	# Novel Patents	# Novel Patents	# Novel Patents
Mega firm	2.909***	0.185***	0.109***	0.041
	(0.062)	(0.037)	(0.037)	(0.050)
Mega firm x Post 2001	0.421***	0.291***	0.347***	0.217***
	(0.123)	(0.053)	(0.053)	(0.056)
Lagged IHS(# total patents)		0.896***	0.826***	0.657***
		(0.005)	(0.008)	(0.014)
Lagged log sales			-0.023	-0.049*
			(0.019)	(0.026)
Lagged log employment			0.119***	0.089***
			(0.022)	(0.030)
Constant	0.355***	-3.014***	-2.905***	-1.560***
	(0.025)	(0.028)	(0.035)	(0.089)
NAICS4 x Year FE	Yes	Yes	Yes	Yes
Firm FE	No	No	No	Yes
Obs.	49,461	38,819	32,879	24,136
R-sq	0.45	0.73	0.74	0.76

Notes: The estimates are obtained from Poisson pseudo-maximum likelihood regressions (PPML) with multi-way fixed effects. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. IHS stands for Inverse-Hyperbolic Sine transformation.

Column (1) is the benchmark specification that only controls for industry (four-digit NAICS) by year fixed effects, which enables us to verify whether the increasing trend of novel patents by mega firms is driven by their industry composition.[11] The estimated coefficients indicate that mega firms produced 18.3 times more novel patents than non-mega firms prior to 2001, while mega firms produce an additional 49 percent more novel patents after 2001.[12] In Columns (2) and (3), we examine how much of these differences are explained by observable firm characteristics. Column (2) shows that controlling for the firms' engagement in overall patenting, measured by the lagged value of total number of patents (applying inverse-hyperbolic sine transformation to accommodate zero patents), reduces the

[11] The results are very similar when we do not include these fixed effects.
[12] The interpretation of the coefficients are derived from $\exp(2.909) = 18.3$ and $\exp(0.421)-1 = 0.49$.

vast majority of the difference in the outcome variable between mega firms and non-mega firms prior to 2001, while a large and significant difference still exists after 2001. In Column (3), we additionally control for firm sales and employment and find that mega firms produce relatively more novel patents after 2001, suggesting that the trend toward increasing novel patents by mega firms holds independently of the increase in their total number of patents or size. Finally, in Column (4), we include firm fixed effects to also account for time-invariant unobservable factors. We find that mega firms produce 24 percent more novel patents than non-mega firms after 2001 even after controlling for firm fixed effects, suggesting that firms are more (less) likely to produce novel patents as they pass the size threshold from (to) a non-mega firm to (from) a mega firm in recent decades.

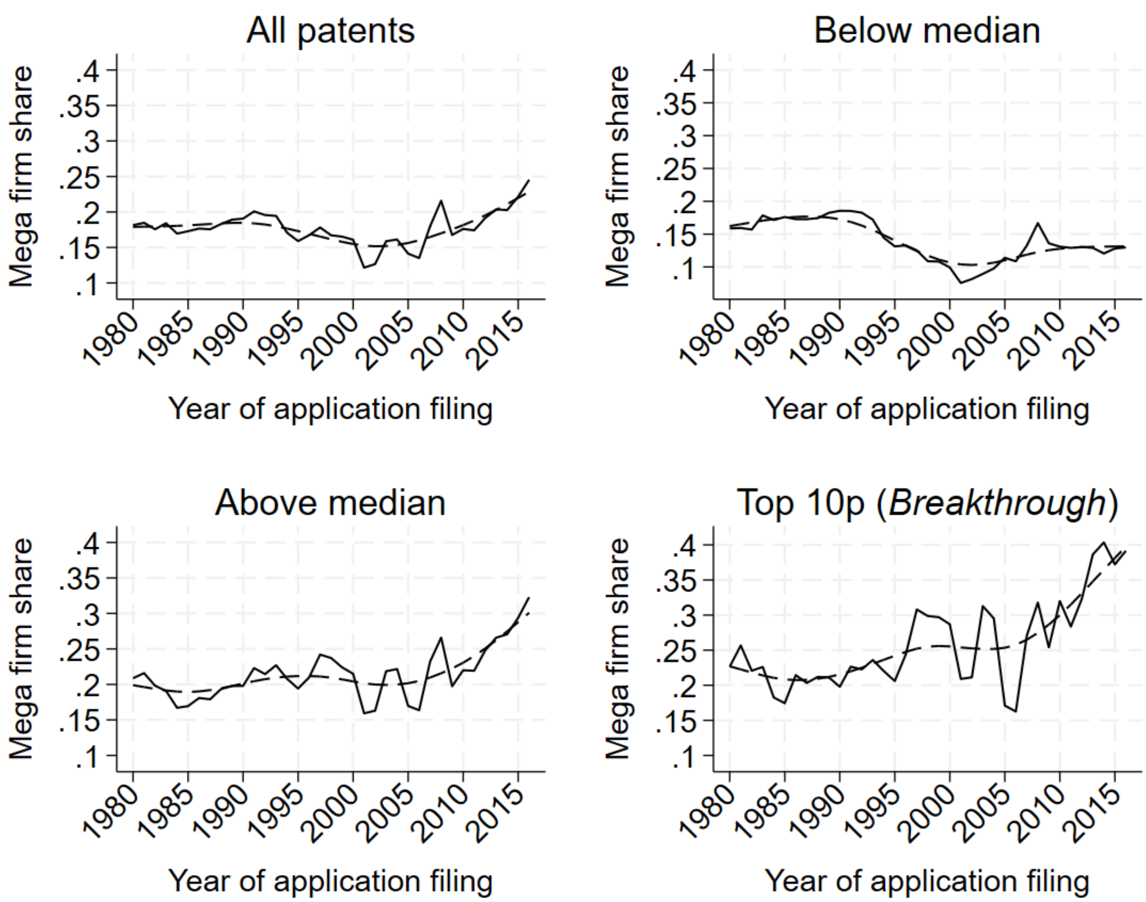
3.3 Trends in Mega Firms' Share of Novel and Impactful Innovation: An Alternative Measure

We further examine whether the rise in mega firms' contributions to novel innovation since 2001 remains evident when using alternative measures of novelty. Specifically, we employ the widely recognized indicator of patent importance developed by Kelly et al. (2021), which leverages natural language processing techniques to establish links between new patents and both existing and subsequent patents. As mentioned in Section 2.2, Kelly et al. identify a patent as important if its content is distinct from prior patents (is novel) but similar to future patents (is impactful). Thus, the indicator of patent importance identifies advancements at the technological frontier that serve as a foundation for subsequent inventions. Since Kelly et al. adopt a fundamentally different measurement approach from ours, applying this metric enables us to assess the robustness of our findings regarding trends in mega firms' role in creating novel innovation. Additionally, it sheds light on whether mega firms' contributions to both novel *and* impactful patents have increased over time.[13]

[13]We present further evidence on the impact of novel patents produced by mega firms using the follow-on patents measure in the next section.

We categorize patents into groups based on their importance, as identified by Kelly et al., ranging from the bottom 10 percent (the least novel and least impactful) to the top 10 percent (the most novel and impactful), with the latter classified as breakthrough patents. Figure 2 presents the share of patent applications filed by mega firms as a proportion of all patent applications (solid line) within each group, alongside its smoothed trend (dashed line), estimated using the HP filter.

Figure 2: Share of Patents Filed by Mega Firms Across Novelty and Impact Groups



Source: Author's own calculation using the USPTO patent data matched with Compustat data.
Notes: The figure displays the share of patent applications filed by mega firms as a proportion of all patent applications (solid line) within each novelty and impact group, along with its smoothed trend (dashed line) estimated using the Hodrick-Prescott filter with a smoothing parameter of 100. The title in each panel indicates the respective novelty and impact group, based on the measure developed by Kelly et al. (2021). For instance, the top-left panel represents the bottom 10% group (i.e., the least novel and least impactful patents), while the bottom-right panel corresponds to the top 10% group (i.e., the most novel and impactful patents, labeled as breakthrough patents by Kelly et al. (2021)). Mega firms are defined as the top 50 firms by sales in Compustat in each year.

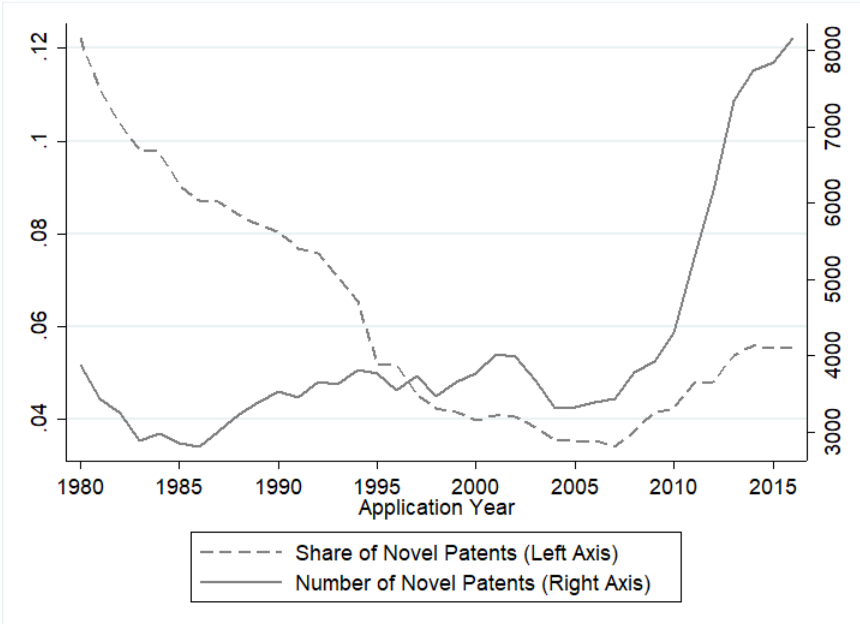
We find that the share of patent applications filed by mega firms among the most novel and impactful patents has risen rapidly since the early 2000s. Specifically, the share of breakthrough patents accounted for by mega firms nearly doubled from 21% in 2001 to 39% in 2016, with a large increase observed among patents in the above-median group. By contrast, mega firms' presence in the less novel and less impactful patents (below median) declined notably in the 1990s and has not rebounded in any meaningful way since. These findings suggest that the growing importance of mega firms extends beyond novel innovation and encompasses both novel and impactful innovation.[14]

3.4 Trends in the Share of Novel Patents in All Patent Applications

The recent increase in novel patents by mega firms shown above is correlated with the changing dynamics of novel patents in the U.S. economy overall. Figure 3 shows that the absolute number of novel patent applications assigned to U.S. entities had been basically flat, while the share of novel patents in total had been declining steadily, from 12% in 1980 to 8% at the start of the 1990s to less than 4% in the early- to mid-2000s, reflecting a rapid increase in the total number of patent applications. After the mid-2000s, 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 almost 6 percent, the level last seen in the mid-1990s. Comparing Figures 1 and 3, we can also see that the increase in the share of novel patents in total patent applications by mega firms started increasing already from around 2001, that is, several years earlier than the share of novel patents applications among all U.S. entities. Thus, the turnaround in the propensity to generate novel patents by mega firms preceded in time the similar turnaround in the broader economy.

[14] We also examined the trend in the share of mega firms using alternative text-based novelty only measures from Arts et al. (2021) and Kalyani (2024). The findings using those alternative measures are, once again, similar to the findings using our measure of novel patents, although the degree of similarity varies according to how many "new bigrams" are used to define novel patents. Details are available upon request.

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



Source: Author's own calculation using the USPTO patent data.

4 Technological Impact and Knowledge Diffusion

4.1 Follow-on Patents

Having established that mega firms are playing a growing role in producing novel and impactful patents, as measured by Kelly et al. (2021), we now examine the technological impact of these novel patents more directly. While Kelly et al.'s approach provides a valuable assessment of a patent's novelty and its impact, it captures these dimensions through a broad lens of linguistic similarity. To complement this perspective, we shift our focus to how these novel patents shape new technological trajectories through a more direct reuse of their innovations. To this end, we adopt the approach of Pezzoni et al. (2022), who measure the technological impact of a novel patent by the number of follow-on patents it generates—patents that re-use the same new technological combination introduced by the initial novel patent. In the baseline, we track follow-on patents over a five-year window to ensure consistent measurement across all patents, including the most recent ones. We also conduct robustness checks using alternative time windows, which are discussed in the appendix. This measure allows us to capture the extent to which a novel patent serves as a foundation for subsequent innovations, offering a more targeted assessment of its success in driving technological progress.[15]

Using patent-level data on novel patents, we compare the number of follow-on patents generated from their novel patents to those from non-mega firms and examine how this relationship evolves over time. We estimate regressions where the outcome variables capture different aspects of the follow-on patent distribution. Table 4 presents the results.

Table 4: Follow-on Patents on Novel Patents by Mega Firms

	(1)	(2)	(3)
	# Follow-on Patents	No Follow-on	Hit
Mega firm	-0.070**	0.005	-0.001
	(0.027)	(0.005)	(0.001)
Mega firm x Post 2001	0.190***	-0.015**	0.005***
	(0.037)	(0.007)	(0.002)
Constant	0.909***	0.419***	0.012***
	(0.007)	(0.001)	(0.000)
Section x Year FE	Yes	Yes	Yes
Obs.	147,318	147,318	147,318
Pseudo R-sq	0.07		
R-sq		0.04	0.00

Notes: Column (1) shows the result from a Poisson pseudo-maximum likelihood regression (PPML) with multi-way fixed effects, where the outcome variable is the number of follow-on patents. Columns (2) and (3) are results from linear probability OLS regressions. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In Column (1), the outcome variable is the number of follow-on patents applied for within the first five years after the novel patent, while in Column (2), it is an indicator for whether no follow-on patents were applied for within this period. Column (1) reports estimates from a Poisson pseudo-maximum likelihood regression, and Column (2) presents results from a linear probability model. The results in Column (1) indicate that, before 2001, novel patents

[15] Forward citations, a widely-used measure of a patent's impact, is less appealing in our context because a novel patent may be cited for reasons unrelated to the new combination they introduce. Nevertheless, our findings remain robust when using citation-based measures.

by mega firms generated 7% fewer follow-on patents than those by non-mega firms, but after 2001, they produced 13% more.[16] The estimates in Column (2) suggest that the likelihood of having no follow-on patents within five years was similar for mega and non-mega firms before 2001 but was 1.5 percentage points lower for mega firms thereafter.

To assess the role of mega firms in producing not just any novel patents but the most successful ones, we follow Pezzoni et al. (2022) and identify "hit" novel patents. In the baseline specification, these are defined as novel patents that rank in the top one percentile of follow-on patent counts within their main CPC section over the first five years (results are robust to alternative thresholds, such as the top five percentile). Column (3) of Table 4 presents estimates from a linear probability model where the outcome variable is an indicator for whether a novel patent becomes a hit. The results indicate no significant difference between mega and non-mega firms before 2001, but after 2001, novel patents by mega firms were 0.5 percentage points more likely to become hits, compared to a baseline probability of 1.2% for non-mega firms.

4.2 Self-follow-on Patents

Pioneering a new combination is inherently experimental and can be viewed as risk-taking reducing the uncertainty in viability of a new technological space and thus facilitating follow-on innovation. It has been argued that a possible reason for declining business dynamism in the U.S. and elsewhere may be a slowdown in new knowledge diffusion from leading to laggard firms (e.g., Akcigit and Ates, 2021). One rather straightforward way to examine knowledge spillover from leading to laggard firms is by looking at the dissemination of the follow-on patents beyond the focal firm that generated the novel patent—whether follow-on patents are assigned to the same or to different firms. If the share of follow-on patents generated by the same firm that came up with the new combination (self-follow-on rate, hereafter) is increasing over time, it can perhaps be interpreted as evidence of a slowdown

[16] exp(0.190-0.070)-1=0.127.

in new knowledge diffusion.

Table 5 presents regression results where the dependent variable is the self-follow-on rate within the first five years. Column (1) includes an indicator for mega firms as the key independent variable, while Column (2) adds an interaction term between the post-2001 dummy and the mega firm indicator. Columns (3) and (4) replicate this analysis for "hit" novel patents, defined as before.[17]

Table 5: Self-follow-on Rates of Novel Patents over the First Five Years

	(1) Self-follow-on Rate	(2) Self-follow-on Rate	(3) Self-follow-on Rate	(4) Self-follow-on Rate
Mega firm	-0.003	0.004	0.015	0.016
	(0.006)	(0.006)	(0.030)	(0.035)
Mega firm x Post 2001	-0.011	-0.002	-0.036	-0.002
	(0.008)	(0.008)	(0.038)	(0.042)
Post 2001	0.015***		-0.010	
	(0.003)		(0.015)	
Constant	0.354***	0.361***	0.220***	0.208***
	(0.002)	(0.002)	(0.011)	(0.007)
Section x Year FE	No	Yes	No	Yes
Condition on Hits	No	No	Yes	Yes
Obs.	85,656	85,656	1,763	1,725
R-sq	0.00	0.02	0.00	0.23

Notes: All columns show the result from linear probability OLS regressions, where the outcome variable is the share of follow-on patents that are produced by the firm that initially produced the corresponding novel patent. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The estimation results in Column (1) do provide some evidence of a slowdown in knowledge diffusion for novel patents overall, as indicated by the positive and significant coefficient on the post-2001 dummy variable. However, mega firms exhibit a self-follow-on rate comparable to that of all firms, with no significant change over time. Also, as can be seen from Column (3), there is no evidence of any decline in knowledge diffusion from most important "hit" novel patents, either for all firms or for the subset of mega firms, before and after

[17]Novel patents without any follow-on patents are excluded from these regressions since the dependent variable is the share of self-follow-on among all follow-on patents.

2001. Thus, at this level of analysis, we find no evidence supporting a decline in knowledge diffusion from novel combinations pioneered by mega firms.

5 Potential Driving Factors

As mentioned, the time concurrence between the declining trend in the efficiency of R&D investment and business dynamism, combined with the rise of "superstar" firms, has prompted economists to consider the possible relation between the two. Our empirical examination of the most recent trends in novel patents has, however, produced evidence that at least largest among the largest, mega firms have been contributing more, not less to the generation of novel patents compared to other firms, with the turnaround happening sometime in the early 2000s. A key question is what drove this rise in novel innovation by mega firms since the early 2000s. In this section, we propose two hypotheses that can potentially shed light on this issue and present some empirical evidence to support those.

5.1 Cash Holdings by Mega Firms

Several studies in the finance literature have documented a rise in U.S. firms' cash holdings relative to assets since the 1980s, largely driven by R&D-intensive firms and those becoming more R&D-intensive over time (Bates et al., 2009; Pinkowitz et al., 2016). Consistent with this trend, previous studies also find a strong relationship between R&D spending and cash holdings (e.g., He and Wintoki, 2016). Such cash holdings could be especially important for R&D spending aimed at creating new technological combinations because new technological combinations are inherently risky and require some financial "slack" to make them feasible (as mentioned, 42% of novel patents do not have any follow-on patents in the first five years). Hence, one possible mechanism behind increasing share of mega firms in producing novel patents in recent years could be the increase in their cash holdings relative to non-mega firms. We are thus led to examine the trend in differences in cash holdings between mega and non-mega firms and whether increased cash holdings predict greater novel innovation.

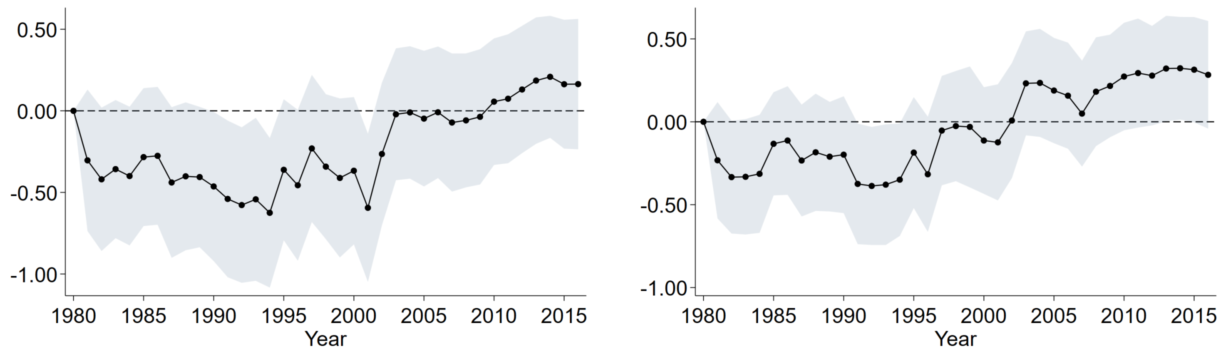
To examine the trend in cash holdings differences, we estimate a difference-in-difference style regression using our Compustat firm panel data. Specifically, we estimate the regression

$$Y_{it} = \alpha + \beta \cdot Mega_{it} + \sum_{\tau=1980}^{2016} \eta_\tau \cdot \mathcal{I}_{\tau=t} + \sum_{\tau=1980}^{2016} \delta_\tau \cdot \mathcal{I}_{\tau=t} \cdot Mega_{it} + X_{it}'\gamma + \theta_i + \epsilon_{it} \quad (3)$$

where Y_{it} represents the log of cash holdings for firm i in year t, and $Mega_{it}$ is an indicator for mega firms. $\mathcal{I}_{\tau=t}$ are year dummies and the term η_τ denotes year fixed effects, while δ_τ is the coefficient of interest, capturing the evolution of differences in cash holdings between mega firms and non-mega firms. The control variables X_{it} include log assets as the baseline control, aligning with the literature that measures cash holdings using cash-to-asset ratios. The results remain robust when additionally controlling for employment or sales.

Figure 4 presents the estimated δ_τ using 1980 as the base year. The left panel once again displays the results without firm fixed effects (θ_i), while the right panel includes them. We find that mega firms' cash holdings, relative to non-mega firms, declined from 1980 to the mid-1990s but have steadily increased since. The sharp rebound since the early 2000s is evident regardless of whether we isolate within-firm changes in cash holdings (i.e., firms holding more cash as they become mega firms) or also account for compositional shifts (i.e., firms with more cash becoming mega firms).

Figure 4: Log Difference in Cash Holdings Between Mega Firms and Non-mega Firms



(a) Without Firm Fixed Effects

(b) With Firm Fixed Effects

Notes: The figure shows the estimated difference in cash holdings between mega firms and non-mega firms over time using 1980 as the base year, that is, δ_τ in Equation (3). The shaded area displays 95% confidence intervals around the estimated coefficients, constructed using heteroskedasticity robust standard errors.

To examine the relationship between change in cash holdings and subsequent novel innovation, we employ a local projection method. Specifically, we estimate the following equation:

$$\frac{Y_{i,t+k} - Y_{i,t-1}}{(Y_{i,t+k} + Y_{i,t-1})/2} = \alpha + \sum_{j=0}^{3} \beta_j \ln(cash)_{i,t-j} + \sum_{j=0}^{3} \gamma_j' X_{i,t-j} + \theta_i + \eta_t + \epsilon_{t+k} \tag{4}$$

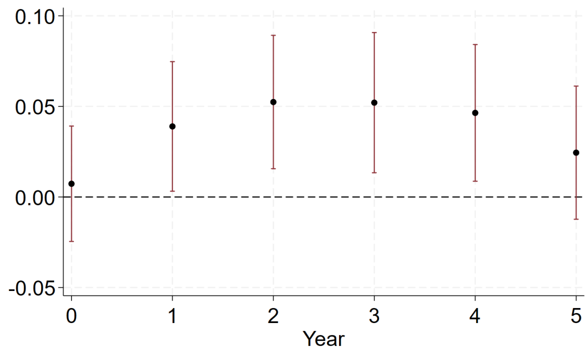
where i indexes firm, t denotes years, Y_{it} represents the number of novel patent applications, and $\ln(cash)_{i,t}$ is the log of cash holdings. The vector $X_{i,t}$ includes control variables such as log assets and total patent applications. Firm and year fixed effects are denoted as θ_i and η_t, respectively. The dependent variable captures the percent change in novel patent applications between year t-1 and t+k. The coefficient of interest is β_0, which indicates the percent change in novel patent applications in year $t + k$ following a one percent increase in cash holdings in year t.

The dependent variable is a simple relative change measure developed by Tornqvist et al. (1985) and popularized by Davis et al. (1996). This measure offers several advantages. First, it is symmetric for increases and decreases, similar to the log-difference measure, but unlike the log difference, it accommodates zeros.[18] This property is particularly relevant to this analysis, as many firm-year observations in our dataset have zero novel patent applications. Second, this measure aligns with the recommendations of Chen and Roth (2024) and Mullahy and Norton (2024) who advocate for scale-independent change measures. In contrast, transformations such as $log(1+x)$ or inverse hyperbolic sine are scale-dependent and therefore less suitable for outcome measures. Lastly, this measure provides a better model fit when dealing with highly skewed variables compared to the traditional percent change measure.

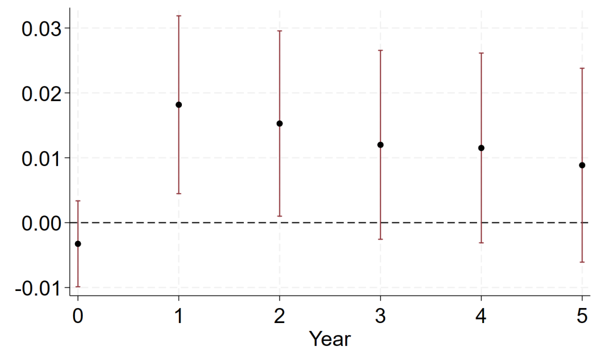
Figure 5 displays the estimated β_0 for $k = 0, 1, 2, 3, 4$, and 5. The left and right panels display results where Y_{it} represents novel patent applications and total patent applications, respectively. The left panel shows that following a one percent increase in cash holdings in year t (holding total assets constant), the number of novel patent applications gradually rises, reaching a 5% increase by year t+2. In contrast, the right panel indicates a more immediate increase in total patent applications, though the peak effect is small (2%) compared to novel patent applications. Thus, increases in cash holdings are more strongly associated with novel patents than with total patent applications.

[18]This measure is closely related to the standard measure of relative change (e.g., $(y-x)/x$) which lacks symmetry in increases and decreases (Tornqvist et al., 1985). It closely approximates the log-relative change measure for positive outcomes, as discussed in Davis et al. (1996) and Vartia (1976).

Figure 5: Percent Change in Patent Applications Following An Increase In Cash Holdings



(a) Novel Patent Applications



(b) Total Patent Applications

Notes: The figures show estimated values of β_j in Equation (4), where the left panel and right panel use the percent change in notal patent applications and total patent applications, respectively. Whisker bands represent 95% confidence intervals. Robust standard errors are used.

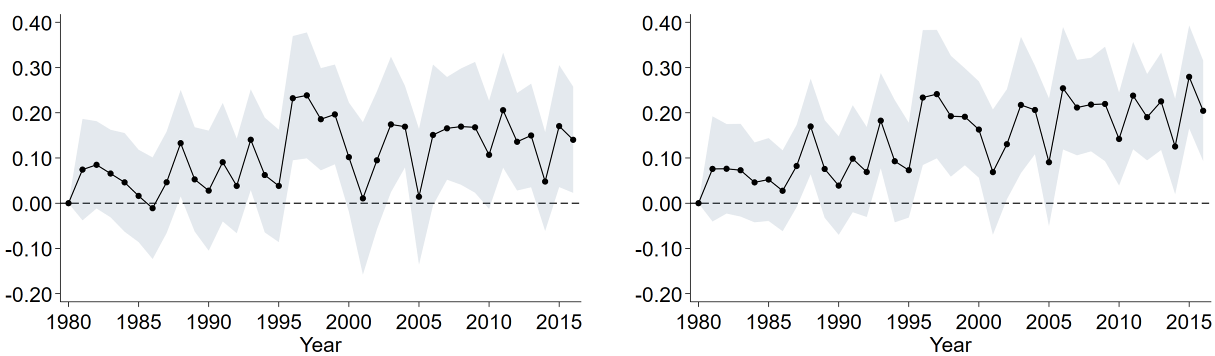
Combined, these findings support the notion that the relative increase in cash holdings by mega firms after 2001 has enabled them to engage in greater novel innovation compared to non-mega firms.

5.2 Size versus Technological Scope of Inventor Teams

Recent studies find that inventors are becoming increasingly concentrated in large firms (Akcigit and Goldschlag, 2023). One may think that this may also contribute to the increase in the share of mega firms in novel patents as the likelihood of generating new combinations may be increasing almost mechanically with the larger size of inventor teams. To assess whether the evidence aligns with this simple mechanism, we analyze the trends in the differences in inventor team size between mega firms and non-mega firms. Specifically, we once again estimate Equation (3) where Y_{it} now represents the average number of inventors associated with patent applications by firm i in year t, while other variables are as defined in the previous subsection. The coefficient of interest, δ_τ captures the evolution of differences in the average number of inventor team size between mega firms and non-mega firms.

Figure 6 presents the evolution of differences in the average size of inventor teams (i.e., δ_τ in Equation (3)) using 1980 as the base year. The left panel shows results without firm fixed effects, while the right panel includes them. We find that the relative size of inventor teams in mega firms, compared to non-mega firms, has grown steadily since 1980. This increasing trend appears more pronounced when focusing on within-firm variations. The relative increase in inventor team size in mega firms, however, is observed throughout the time period, so it is not consistent with the decline in novel patents generated by mega firms from 1980 to 2000. In other words, if increasing the relative size of the inventor teams were all what it takes to increase the output of novel patents, we would expect to see such increase uniformly over all the decades. Instead, the trend in the share of mega firms in novel patents is U-shaped (Figure 1), so clearly, the increase in the team size is not the (whole) story.

Figure 6: Log Difference in Inventor Team Size Between Mega Firms and Non-mega Firms



(a) Without Firm Fixed Effects

(b) With Firm Fixed Effects

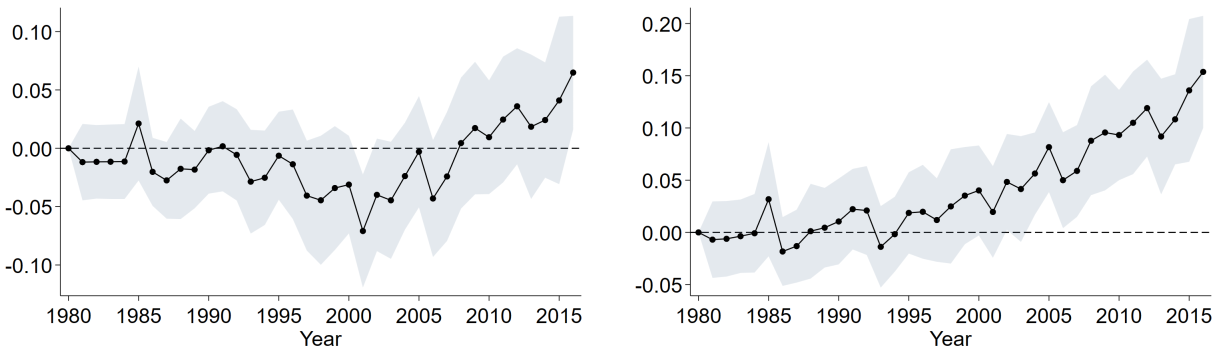
Notes: The figures present estimated values of δ_τ in Equation (3), where the outcome variable is the log of the average number of inventors associated patent applications at the firm-year level. The left panel excludes firm fixed effects, while the right panel includes them. Shaded area represents 95% confidence intervals, and robust standard errors are used.

Now, by definition, novel innovation involves combining technological components that have not previously been used together. This concept is closely aligned with recombinant innovation, which relies on new recombinations of existing knowledge (Weitzman, 1998). We thus conjecture that what matters for novel patents is not merely the size of the inventor team but, more importantly, the breadth (scope) of the technological expertise embodied in these teams. To probe this conjecture, we replace the dependent variable Y_{it} with the average number of technological areas in which inventor teams have expertise when estimating Equation (3). Since team size and technological scope may be mechanically correlated, we control for firms' average inventor team size, though results remain robust to excluding this control. To construct the variable capturing an inventor's technological expertise in year t, we identify the CPC technology group in which they have filed the largest number of patents over their entire cumulative patenting history up to year t.[19] We then aggregate this technological expertise across all members of the inventor team associated with patent applications by firm i in year t.

Figure 7 presents the evolution of differences in the technological scope of inventor teams between mega and non-mega firms. The left panel, once again, presents results without firm fixed effects, capturing both within-firm changes and compositional shifts. The relative technological scope of inventor teams at mega firms follows a U-shaped pattern, with a trough around 2000. The right panel incorporates firm fixed effects, isolating within-firm variations. From the mid-1990s, firms began broadening the technological scope of their inventor teams as they became mega firms.

[19]See more details in Appendix A.7.

Figure 7: Log Difference in Technology Scope Between Mega Firms and Non-mega Firms



(a) Without Firm Fixed Effects

(b) With Firm Fixed Effects

Notes: The figures present estimated values of δ_τ in Equation (3), where the outcome variable is the log of the average number of CPC technological groups in which inventor teams have expertise at the firm-year level. The left panel excludes firm fixed effects, while the right panel includes them. Shaded area represents 95% confidence intervals, and robust standard errors are used.

The picture in the left panel exhibits a declining trend until the early 2000s and an increasing trend thereafter. In contrast, in the right panel where we control for firm fixed effects, the trend is flat until about 2000, with a rather steep increase after that. Thus, comparing the two panels, we can see that the decline in technological scope in the two last decades of the 20th century was driven by compositional shifts—an increasing share of firms with narrower technological scope among mega firms, while after 2000, mega firms with a broader technological scope took the center stage.

Thus, the estimation results suggest that the competitive advantage of mega firms in producing novel innovations may be accruing to those of them that assembled not just larger but more diverse, in terms of their technological expertise, inventor teams (recall that we control for firms' average inventor team size when estimating Equation (3)). The findings are also robust to including additional firm size measures, such as employment and sales. Overall, these results support the hypothesis that an increasing scope of technological expertise at mega firms has facilitated the rapid reversal in their contribution to novel innovations since the early 2000s.

6 Conclusion

Our empirical analysis reveals a U-shaped trend in the share of novel patents produced by mega firms, declining from 1980 to the early 2000s before rebounding sharply to reach its highest level by 2016. This resurgence reflects not only an increase in mega firms' overall patenting activity but also a heightened propensity to engage in novel innovation, as further evidenced by firm-level analysis. Since 2001, novel patents by mega firms have demonstrated greater technological impact, generating more follow-on patents and being more likely to become "hits" compared to those by non-mega firms. We also found that the share of novel patents in U.S. patent applications overall had declined over several decades until the mid-2000s, followed by a rebound since then, starting several years after the start of the rebound in the share of novel patents by mega firms. Together, these findings suggest that mega firms might be increasingly shaping new technological trajectories in the U.S. economy.

Importantly, our findings on knowledge diffusion indicate that the self-follow-on rate—the proportion of follow-on patents generated by the same firm—has remained stable for mega firms over time, showing no evidence of impeding knowledge spillover specifically for their novel patents. This stability, combined with the higher volume of follow-on patents post-2001 for these novel innovations, suggests that mega firms are facilitating knowledge diffusion to other firms through this selected subset of innovations. While mega firms may restrict knowledge diffusion tied to other types of patents, our analysis highlights that this pattern does not extend to their novel innovations. We also explore potential drivers of this trend, finding suggestive evidence that the relative increase in cash holdings since the early 2000s may have enabled mega firms to fund risky, experimental R&D, while the broadening technological scope of their inventor teams may have enhanced their capacity to pioneer new combinations.

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. While mega firms may protect their technological superiority in certain dimensions, our evidence suggests that, especially in recent years, they are leading technological experimentation through novel innovations that enable follow-on innovation by others. If mega firms were predominantly hindering knowledge spillover, there might be a case for considering regulatory intervention. However, given their role as key actors in generating novel technologies, the impact of such measures requires careful consideration to avoid unintended consequences. Understanding the balance between these countervailing forces remains a critical research agenda for informing policy in this debate.

References

Akcigit, Ufuk and Sina T Ates (2021) "Ten facts on declining business dynamism and lessons from endogenous growth theory," *American Economic Journal: Macroeconomics*, 13 (1), 257–298.

——— (2023) "What happened to US business dynamism?" *Journal of Political Economy*, 131 (8), 2059–2124.

Akcigit, Ufuk and Nathan Goldschlag (2023) "Where have all the" creative talents" gone? Employment dynamics of us inventors," Working Paper No. 31085, National Bureau of Economic Research.

Akcigit, Ufuk and William R Kerr (2018) "Growth through heterogeneous innovations," *Journal of Political Economy*, 126 (4), 1374–1443.

Arts, Sam, Jianan Hou, and Juan Carlos Gomez (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, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen (2020) "The fall of the labor share and the rise of superstar firms," *The Quarterly Journal of Economics*, 135 (2), 645–709.

Bates, Thomas W, Kathleen M Kahle, and René M Stulz (2009) "Why do US firms hold so much more cash than they used to?" *The Journal of Finance*, 64 (5), 1985–2021.

Bloom, Nicholas, Charles I Jones, John Van Reenen, and Michael Webb (2020) "Are ideas getting harder to find?" *American Economic Review*, 110 (4), 1104–1144.

Brown, James R, Steven M Fazzari, and Bruce C Petersen (2009) "Financing innovation and growth: Cash flow, external equity, and the 1990s R&D boom," *The Journal of Finance*, 64 (1), 151–185.

Chen, Jiafeng and Jonathan Roth (2024) "Logs with zeros? Some problems and solutions," *The Quarterly Journal of Economics*, 139 (2), 891–936.

Covarrubias, Matias, Germán Gutiérrez, and Thomas Philippon (2020) "From good to bad concentration? US industries over the past 30 years," *NBER Macroeconomics Annual*, 34 (1), 1–46.

Cunningham, Colleen, Florian Ederer, and Song Ma (2021) "Killer acquisitions," *Journal of Political Economy*, 129 (3), 649–702.

Davis, Steven, John Haltiwanger, and Scott Schuh (1996) *Job Creation and Destruction*, Cambridge, MA: MIT Press.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger (2020) "The rise of market power and the macroeconomic implications," *The Quarterly Journal of Economics*, 135 (2), 561–644.

Epicoco, Marianna, Magali Jaoul-Grammare, and Anne Plunket (2022) "Radical technologies, recombinant novelty and productivity growth: a cliometric approach," *Journal of Evolutionary Economics*, 32 (2), 673–711.

Fleming, Lee, Santiago Mingo, and David Chen (2007) "Collaborative brokFerage, generative creativity, and creative success," *Administrative Science Quarterly*, 52 (3), 443–475.

Foster, Lucia S, John C Haltiwanger, and Cody Tuttle (2022) "Rising markups or changing technology?" Working Paper 30491, National Bureau of Economic Research.

Gupta, Abhinav, Naman Nishesh, and Elena Simintzi (2024) "Big Data and Bigger Firms: A Labor Market Channel," *Working Paper*.

Gutiérrez, Germán and Thomas Philippon (2019) "The failure of free entry," Working Paper No. 26001, National Bureau of Economic Research.

He, Zhaozhao and M Babajide Wintoki (2016) "The cost of innovation: R&D and high cash holdings in US firms," *Journal of Corporate Finance*, 41, 280–303.

Hsieh, Chang-Tai and Esteban Rossi-Hansberg (2023) "The industrial revolution in services," *Journal of Political Economy: Macroeconomics*, 1 (1), 3–42.

Jo, Karam and Seula Kim (2024) "Heterogeneous Innovations and Growth Under Imperfect Technology Spillovers," IZA Discussion Paper No. 17581, IZA.

Kalyani, Aakash (2024) *The Creativity Decline: Evidence from US Patents*: Working Paper.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy (2021) "Measuring technological innovation over the long run," *American Economic Review: Insights*, 3 (3), 303–320.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman (2017) "Technological innovation, resource allocation, and growth," *The Quarterly Journal of Economics*, 132 (2), 665–712.

Kwon, Spencer Y, Yueran Ma, and Kaspar Zimmermann (2024) "100 years of rising corporate concentration," *American Economic Review*, 114 (7), 2111–2140.

Mullahy, John and Edward C Norton (2024) "Why transform y? The pitfalls of transformed regressions with a mass at zero," *Oxford Bulletin of Economics and Statistics*, 86 (2), 417–447.

Noel, Michael and Mark Schankerman (2013) "Strategic patenting and software innovation," *The Journal of Industrial Economics*, 61 (3), 481–520.

Pezzoni, Michele, Reinhilde Veugelers, and Fabiana Visentin (2022) "How fast is this novel technology going to be a hit? Antecedents predicting follow-on inventions," *Research Policy*, 51 (3), 104454.

Pinkowitz, Lee, René M Stulz, and Rohan Williamson (2016) "Do US firms hold more cash than foreign firms do?" *The Review of Financial Studies*, 29 (2), 309–348.

Strumsky, Deborah and José Lobo (2015) "Identifying the sources of technological novelty in the process of invention," *Research Policy*, 44 (8), 1445–1461.

Tornqvist, Leo, Pentti Vartia, and O. Yrjo Vartia (1985) "How Should Relative Changes Be Measured?" *The American Statistician*, 39 (1), 43–46.

Torrisi, Salvatore, Alfonso Gambardella, Paola Giuri, Dietmar Harhoff, Karin Hoisl, and Myriam Mariani (2016) "Used, blocking and sleeping patents: Empirical evidence from a large-scale inventor survey," *Research Policy*, 45 (7), 1374–1385.

Vartia, Yrjo O. (1976) "Relative Changes and Index Numbers," unpublished licentiate thesis, University of Helsinki, Department of Statistics.

Verhoeven, Dennis, Jurriën Bakker, and Reinhilde Veugelers (2016) "Measuring technological novelty with patent-based indicators," *Research Policy*, 45 (3), 707–723.

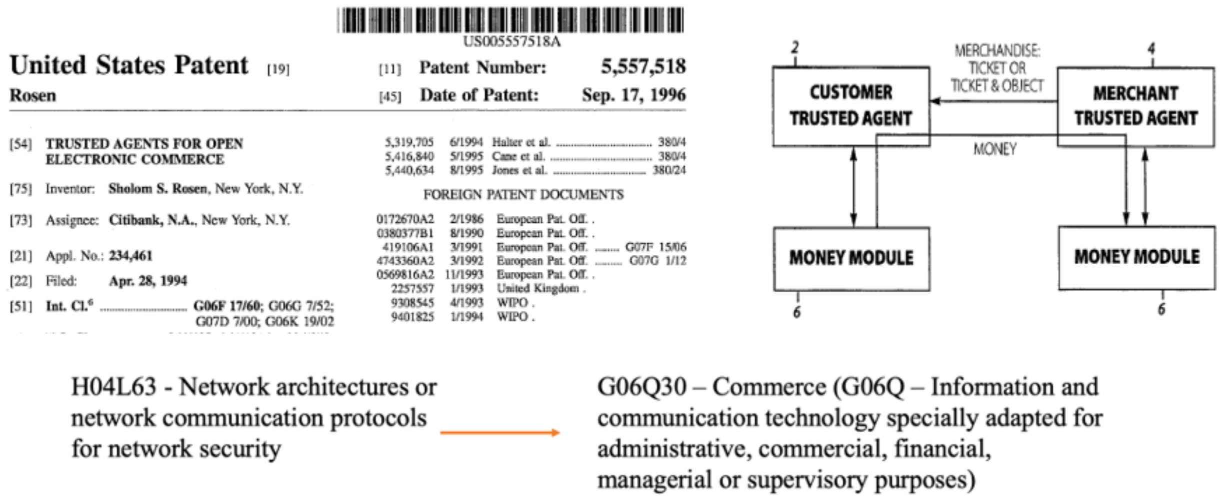
Weitzman, Martin L (1998) "Recombinant growth," *The Quarterly Journal of Economics*, 113 (2), 331–360.

Yeh, Chen, Claudia Macaluso, and Brad Hershbein (2022) "Monopsony in the US labor market," *American Economic Review*, 112 (7), 2099–2138.

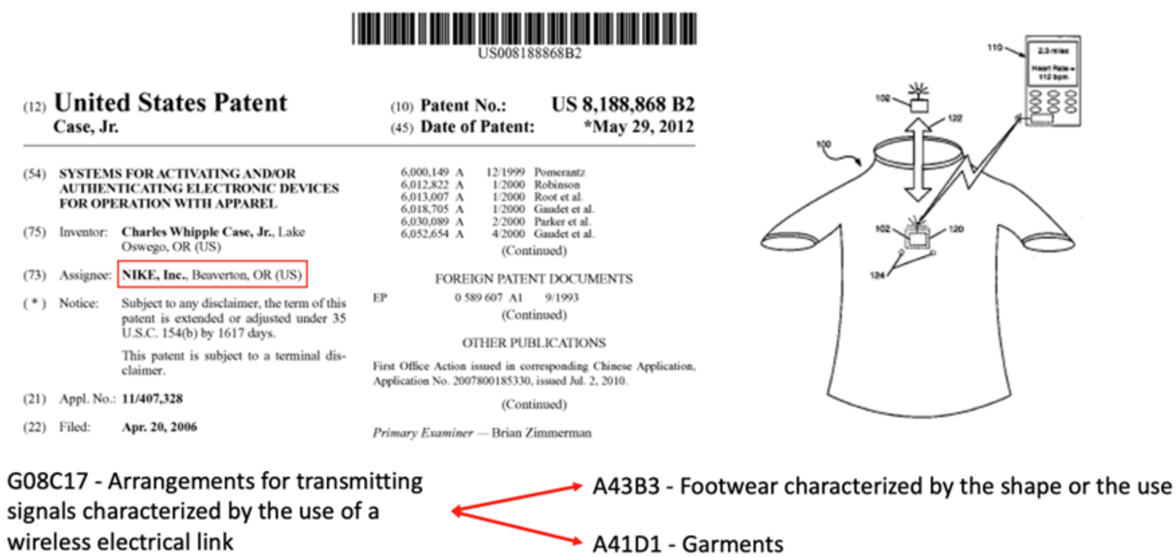
A Online Appendix to Mega Firms and New Technological Trajectories in the U.S.

A.1 Examples of Novel Patents

Figure A1: Visual images of novel patents examples in the Introduction



(a) A Novel Patent by Citibank Combining Two ICT-related Components



(b) A Novel Patent by NIKE Combining ICT-related and Non-related Components

Source: USPTO.

A.2 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 68.1% of utility patent applications (granted until Sep. 30, 2023) filed by U.S. patenting firms, and 25.6% of U.S. patenting firms to Compustat firms from 1976 to 2016.

A.3 Novel Patent Indicator and Alternative Measures of Novelty

This section presents regression results where the dependent variables capture various alternative measures of patent novelty and creative destruction, with the novel patent indicator as the main independent variable. The tables below report results under different specifications, varying the inclusion of CPC group fixed effects and year fixed effects. The findings remain robust across all fixed-effect combinations.

Table A1: Novel Patent & Number of New Bigrams

	(1) # New Bigrams	(2) # New Bigrams	(3) # New Bigrams	(4) # New Bigrams	(5) # New Bigrams
Novel Patent	0.777***	0.852***	0.559***	0.769***	0.715***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Constant	1.516***	1.510***	1.528***	1.515***	1.518***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
CPC Group FE	No	Yes	No	Yes	No
Year FE	No	No	Yes	Yes	No
CPC Group x Year FE	No	No	No	No	Yes
Obs.	2,633,463	2,630,141	2,633,463	2,630,141	2,629,129
R-sq	0.00	0.04	0.03	0.06	0.08

Notes: The estimates are obtained from OLS regressions with multi-way fixed effects. "# New Bigrams" indicates the number of new bigrams in patent text, obtained from Arts et al. (2021). Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2: Novel Patent and Patent Creativity

	(1) Patent Creativity	(2) Patent Creativity	(3) Patent Creativity	(4) Patent Creativity	(5) Patent Creativity
Novel Patent	0.022***	0.022***	0.018***	0.020***	0.019***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.068***	0.068***	0.068***	0.068***	0.068***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CPC Group FE	No	Yes	No	Yes	No
Year FE	No	No	Yes	Yes	No
CPC Group x Year FE	No	No	No	No	Yes
Obs.	2,578,919	2,575,446	2,578,919	2,575,446	2,574,407
R-sq	0.00	0.06	0.06	0.12	0.16

Notes: The estimates are obtained from OLS regressions with multi-way fixed effects. "Patent Creativity" is the share of new technical bigrams in patents, obtained from Kalyani (2024). Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3: Novel Patent and Backward Self-citation Ratio

	(1) Backward Self-citation Ratio	(2) Backward Self-citation Ratio	(3) Backward Self-citation Ratio	(4) Backward Self-citation Ratio	(5) Backward Self-citation Ratio
Novel Patent	-0.031***	-0.026***	-0.032***	-0.027***	-0.026***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.121***	0.121***	0.121***	0.121***	0.121***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CPC Group FE	No	Yes	No	Yes	No
Year FE	No	No	Yes	Yes	No
CPC Group x Year FE	No	No	No	No	Yes
Obs.	2,661,756	2,657,952	2,661,756	2,657,952	2,656,902
R-sq	0.00	0.06	0.00	0.06	0.08

Notes: The estimates are obtained from OLS regressions with multi-way fixed effects. Backward self-citation ratio is defined as the number of citations to a firm's own previous patents divided by the total number of citations. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.4 Novel Patent Indicator and Measures of Impact

This section presents the results of regressions in which the dependent variables represent various measures of patent impact, while the key independent variable is the novel patent indicator. Table A4 illustrates the relationship between stock market valuation and the novel patent indicator.

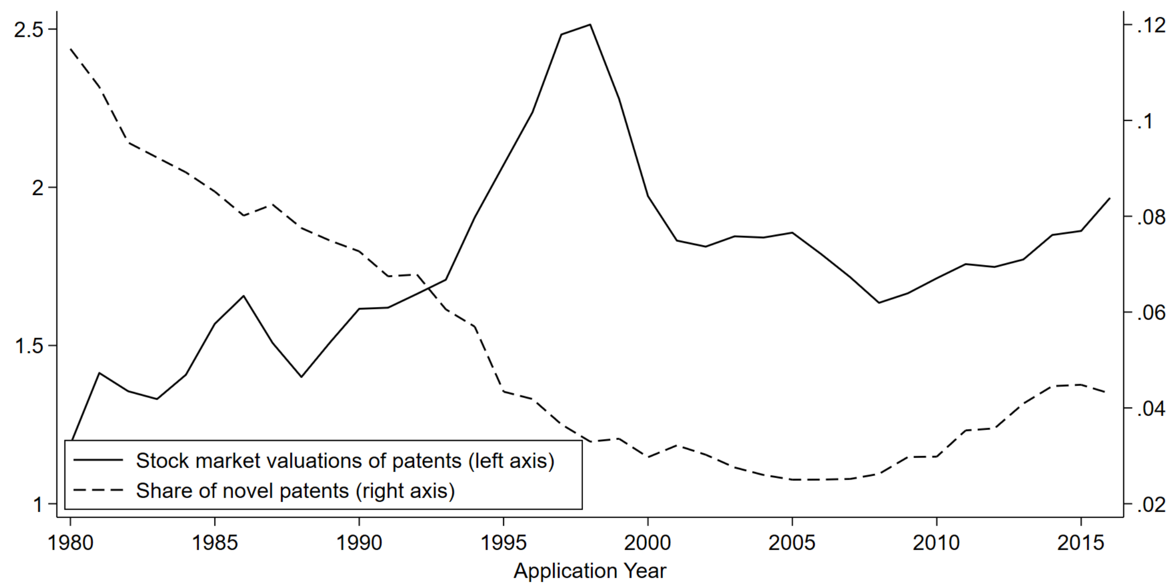
Table A4: Novel Patent and Market Valuation

	(1) Market Valuation	(2) Market Valuation	(3) Market Valuation	(4) Market Valuation	(5) Market Valuation
Novel Patent	-0.025***	-0.025***	0.043***	0.021***	0.006
	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)
Constant	1.825***	1.825***	1.822***	1.823***	1.823***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
CPC Group FE	No	Yes	No	Yes	No
Year FE	No	No	Yes	Yes	No
CPC Group x Year FE	No	No	No	No	Yes
Obs.	1,506,012	1,504,550	1,506,012	1,504,550	1,502,624
R-sq	0.00	0.05	0.04	0.09	0.14

Notes: The estimates are obtained from OLS regressions with multi-way fixed effects. "Patent Creativity" is the share of new technical bigrams in patents, obtained from Kalyani (2024). Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

We find that the relationship is negative when year fixed effects are excluded, as shown in the first two columns. To understand this result, we plot the time series of the average stock market value of patents (solid line) and the share of novel patents (dashed line) in Figure A2. The two series exhibit opposite trends until around 2008, explaining the negative coefficient when time fixed effects are omitted. The share of novel patents has been declining during this period, a trend that is also broadly consistent with the findings of Arts et al. (2021) and Kalyani (2024). Meanwhile, average stock prices have been rising, largely due to factors unrelated to patent novelty, such as the decline in long-term interest rates and the equity risk premium. Therefore, we deem it more appropriate to control for time fixed effects to better isolate the relationship between the two variables. We also find that the estimated relationship is positive but becomes very small and statistically insignificant when controlling for CPC group-by-year fixed effects. This result suggests that the impact of novel patents on firms' profits may be limited and weak.

Figure A2: Trend in the Average Stock Market Valuations and Share of Novel Patents



Source: Author's own calculation using the market valuation data from Kogan et al. (2017) and UPSTO patent database.
Notes: Stock market valuations are log of valuations deflated to 1982 (million) dollars using the CPI.

Tables A5 and A6 present the relationship between the novel patent indicator and two outcomes: the five-year forward citation count and the breakthrough patent indicator developed by Kelly et al. (2021), respectively. In these cases, controlling for CPC group fixed effects is crucial for uncovering the positive relationship.

Table A5: Novel Patent and 5-year Forward Citations

	(1) 5-year Forward Citation	(2) 5-year Forward Citation	(3) 5-year Forward Citation	(4) 5-year Forward Citation	(5) 5-year Forward Citation
Novel Patent	-0.675***	0.528***	-0.428***	0.690***	0.575***
	(0.041)	(0.042)	(0.041)	(0.041)	(0.042)
Constant	5.482***	5.425***	5.469***	5.417***	5.424***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
CPC Group FE	No	Yes	No	Yes	No
Year FE	No	No	Yes	Yes	No
CPC Group x Year FE	No	No	No	No	Yes
Obs.	2,842,894	2,838,910	2,842,894	2,838,910	2,837,936
R-sq	0.00	0.03	0.01	0.04	0.07

Notes: The estimates are obtained from OLS regressions with multi-way fixed effects. Backward self-citation ratio is defined as the number of citations to a firm's own previous patents divided by the total number of citations. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A6: Novel Patent and Breakthrough Patents

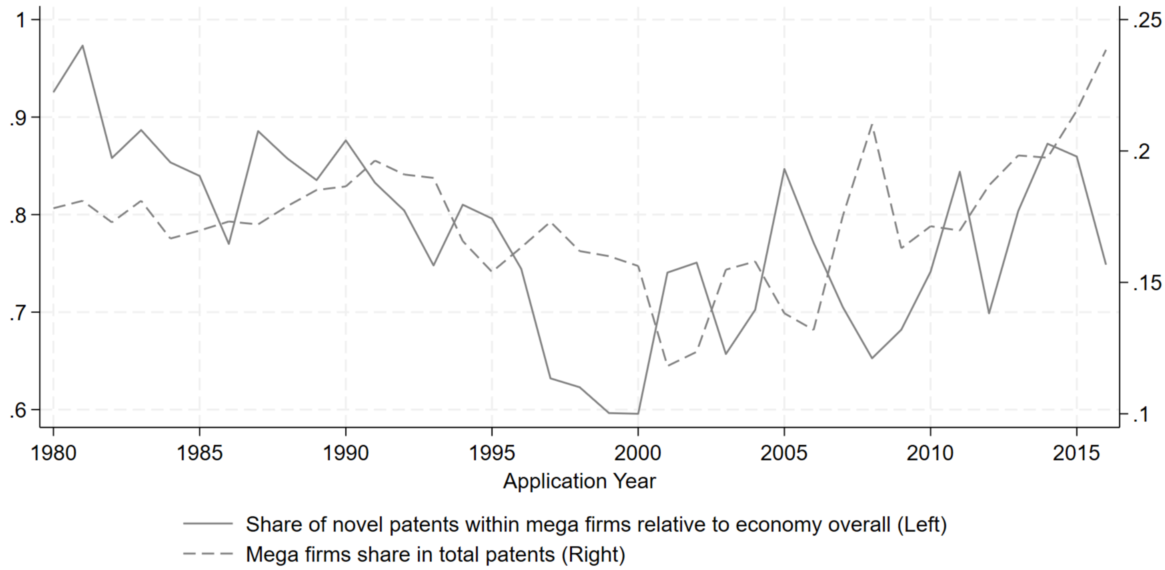
	(1) Breakthrough	(2) Breakthrough	(3) Breakthrough	(4) Breakthrough	(5) Breakthrough
Novel Patent	-0.039***	0.015***	-0.039***	0.015***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.102***	0.099***	0.102***	0.099***	0.100***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CPC Group FE	No	Yes	No	Yes	No
Year FE	No	No	Yes	Yes	No
CPC Group x Year FE	No	No	No	No	Yes
Obs.	2,836,794	2,836,755	2,836,794	2,836,755	2,835,781
R-sq	0.00	0.11	0.02	0.16	0.29

Notes: The estimates are obtained from OLS regressions with multi-way fixed effects. Backward self-citation ratio is defined as the number of citations to a firm's own previous patents divided by the total number of citations. Note that the breakthrough measure in Kelly et al. (2021) is already residualized with respect to year fixed-effects, so including year fixed effects does not change the coefficients. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.5 Decomposition of Mega Firms' Share in Novel Patents

Figure A3 below displays the decomposition of novel patent applications filed by mega firms into (i) the share of novel patents within mega firms relative to that share among all firms (solid line) and (ii) the share of total patents accounted for by mega firms (dashed line). Both time series exhibit a U-shaped pattern with the reversal occuring around 2000.

Figure A3: Decomposition of Mega Firms' Share in Novel Patents

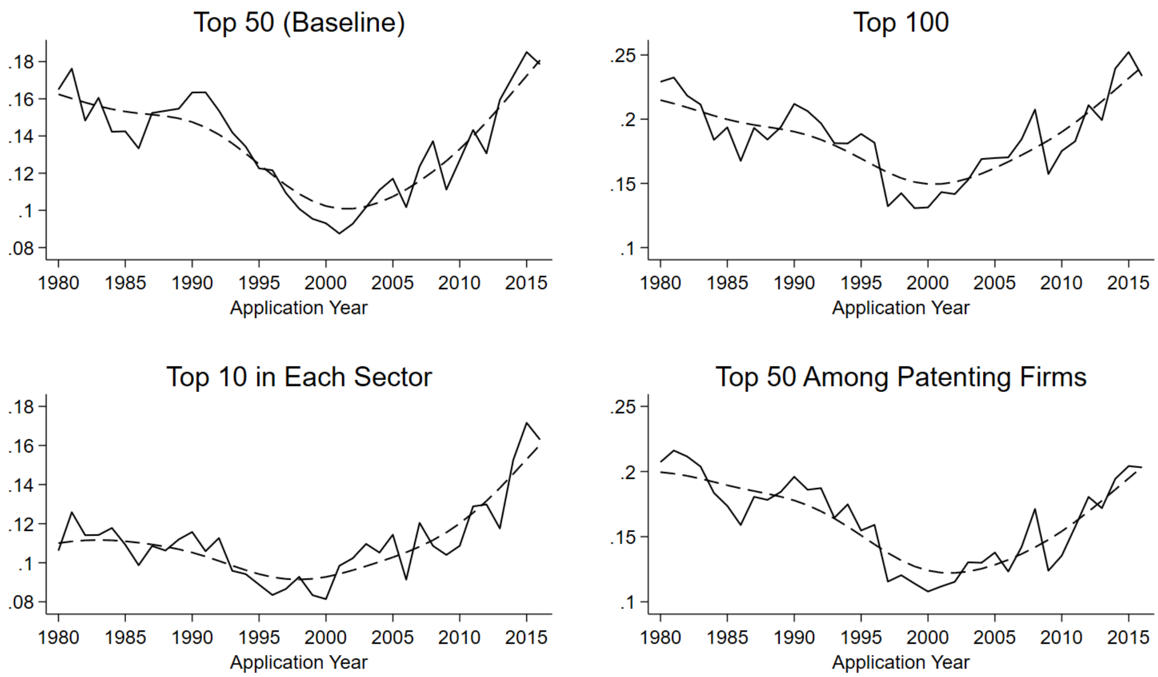


Source: Author's own calculation using UPSTO patent database and COMPUSTAT.

A.6 Trends in Mega Firms' Share of Novel Patents Under Alternative Definitions of Mega Firms

Figure A4 illustrates the share of novel patent applications filed by mega firms under various definitions. The top-left panel reproduces Figure 1 as a benchmark. The top-right panel expands the definition of mega firms to the top 100 firms by sales in each year. The bottom-left panel defines mega firms as the top 10 firms by sales in each two-digit NAICS sector, totaling 190 firms annually across 19 sectors. The bottom-right panel restricts mega firms to the top 50 firms by sales in each year, but only among COMPUSTAT firms that file at least one patent, accounting for the fact that patenting firms represent a highly selected subset. We find that the U-shaped trend reported in Figure 1 remains evident across all definitions.

Figure A4: Trends in Mega Firms' Share in Novel Patents Under Alternative Definitions of Mega Firms



Source: Author's own calculation using UPSTO patent database and COMPUSTAT.

A.7 Inventor Diversity Measures

To measure the diversity of inventor teams in the USPTO, we define an inventor's technological expertise based on the technology class information associated with their patenting history since 1963. For each year, we identify an inventor's core CPC technology group as the one in which they have filed the most patents up to that point.[20] This measure serves as a proxy for the inventor's technological expertise in each period. We then aggregate this information across inventors listed on team patents filed by a firm in each year.

[20]For patents associated with multiple CPC groups, we select the primary CPC group based on the "sequence" variable provided by the USPTO.