Technology Life Cycles: Investment, Growth, and Firm Value

David Barkemeyer * Jan Bena† Xiyue (Ellen) Li‡

March 4, 2026

Abstract

We build a new century-long dataset of technology life cycles from USPTO patent titles, tracing 289 prominent technologies from initiation through emergence, prominence, and maturity over 1920–2023, and linking their patent histories to 9,047 global public firms. The earliest patents in prominent technologies are disproportionately influential and are primarily assigned to firms, especially large innovators. Decades before prominence, firms that patent early invest more, grow faster, and command higher valuation ratios. As technologies approach prominence, valuation premia compress even as profit margins and return on invested capital improve. After prominence, innovation becomes less foundational, pioneers' patent-share advantages erode only slowly on average, and the investment, growth, and valuation advantages of early participation fade or reverse. The erosion of early patent-share leadership is fastest when early competition among large innovators is intense.

*The Wharton School, University of Pennsylvania. E-mail: dbarke@wharton.upenn.edu.

†Sauder School of Business, University of British Columbia. E-mail: jan.bena@sauder.ubc.ca.

‡School of Management and Economics, Shenzhen Finance Institute, Chinese University of Hong Kong (Shenzhen). E-mail: lixiyue@cuhk.edu.cn.

1 Introduction

Transformative technologies and equity valuations have long moved together. During the information-technology revolution and the late-1990s "dot-com" episode, markets placed enormous value on firms exposed to new technologies well before diffusion was complete (Hobijn and Jovanovic, 2001; Pástor and Veronesi, 2009). Work on product life cycles documents regularities in entry, exit, and innovation as industries mature (Klepper, 1996), but we know much less about how the life cycle of a technology shapes the corporate policies and market valuations of the firms that develop it. These questions are again front and center as advances in artificial intelligence move from research to broad commercial deployment.[1]

In this paper, we trace 289 *prominent technologies*—technologies initiated in 1920 or later that enter the annual top 100 by patent count at least once by 2023—from initiation to emergence, prominence, and maturity, and we link their patent histories to firm-level financial data for 9,047 global public firms. These technologies account for 11.4% of all patents granted since 1920, and their share rises to 16.8% in 2023. Building on recent work that uses patent data to quantify innovation and its economic significance (Kogan et al., 2017; Kelly et al., 2021), we classify firms by the stage at which they first patent in a technology: before emergence, between emergence and prominence, or only after prominence. We then study how investment (R&D and capital expenditures), growth, profitability, and valuation (asset market-to-book and price-to-earnings) evolve across these entry cohorts and across stages of the technology life cycle.

Three findings stand out. First, firms that patent early invest more and expand faster long before a technology becomes prominent, and they command higher valuation ratios decades in advance, with the strongest evidence in asset market-to-book. Second, as technologies approach prominence, these valuation premia compress even as operating performance improves, especially in profit margins and return on invested capital, consistent with

[1] In a Bloomberg *Odd Lots* interview, Jeffrey Gundlach notes that "the electricity outperformance peaked in 1911" and that "it all gets priced in very quickly and excessively." (Gundlach, 2025)

markets shifting from pricing growth options under uncertainty toward pricing realized fundamentals as diffusion proceeds. Third, after prominence, innovation becomes less foundational and the investment, growth, and valuation advantages of early participation fade and often reverse as technologies mature.[2] Within a technology, pioneers enter the post-prominence phase with larger patent shares, but those leads erode over time—slowly on average, more slowly for large pioneers and for pioneers with broader prior technological scope, and much more quickly when early competition among large innovators is intense.

For these analyses, we construct a new century-long dataset of technologies from the universe of USPTO patent titles, which are available in a consistent format since the establishment of the modern U.S. patent system in 1836. Patent titles are produced under legal and administrative constraints that encourage short, specific phrasing indicative of the invention's subject matter; we exploit this feature to define technologies as recurring multi-word phrases in titles.[3] This transparent, classification-free approach yields technologies that are directly interpretable and can be tracked from inception to maturity. The resulting vocabulary aligns closely with well-known technological eras: nineteenth-century patenting features terms such as *steam engine* and *sewing machine*, *internal combustion engine* dominates much of the twentieth century, and ICT-related terms—including *communication system*, *display device*, *electronic device*, and *machine learning*—rise to prominence in recent decades.

We validate the dataset by reproducing three long-run patterns documented in prior work. First, inventive activity becomes increasingly combinatorial (Figure 1), consistent

[2] A classic practitioner view is that technological breakthroughs can create enormous social value without guaranteeing value capture for early investors or producers. Discussing airlines, Buffett writes that "if a farsighted capitalist had been present at Kitty Hawk, he would have done his successors a huge favor by shooting Orville down." (Buffett, 2008)

[3] Patent titles are intended to identify the invention's subject matter as precisely as possible; see, e.g., the WIPO "ST.15 – Guidelines for the wording of titles of inventions in patent documents" and 37 C.F.R. §1.72. In brief, we clean and standardize titles, segment them into short phrases using a dictionary of separators (punctuation, conjunctions, and prepositions), generate candidate multi-word n-grams within each phrase, and retain those that recur across patents and frequently appear as stand-alone phrases; see Sections 2.1 and 2.2 and Appendices B and C for details. The resulting vocabulary contains 715,788 technologies. Coverage is broad: at least one technology is identified in 69% of all patents granted between 1836 and 2023; coverage exceeds 50% after 1870 and reaches about 80% in the most recent decade. Coverage is also similar across one-digit CPC sections.

with theories and evidence on recombinant invention (Weitzman, 1998; Fleming, 2001). Second, patenting becomes increasingly concentrated in a small set of dominant technologies (Figure 2), echoing evidence that U.S. innovative activity has increasingly tilted toward a narrow set of high-growth technology domains (Autor et al., 2020a). Third, the rise of dominant technologies coincides with a growing role of large innovators (Figure 3), in line with recent evidence on rising patenting concentration among market-leading firms (Akcigit and Ates, 2023). Recovering these stylized facts within a single unified framework provides external validation for our technology measures and motivates their use in the corporate-finance analysis that follows.

We then date three milestones for each technology. The *initiation year* is the grant year of the technology's first patent. The *emergence year* is the first year in which the technology has at least five granted patents, and the *prominence year* is the first year in which it ranks among the annual top 100 technologies by patent count. Prominent technologies typically take decades to develop: the median technology reaches emergence about 10 years after initiation and reaches prominence after 29 years, with substantial heterogeneity in diffusion speed and ultimate scale. The earliest phase is small but potentially pivotal: the average prominent technology has about 20 pre-emergence patents and about 483 pre-prominence patents. Some technologies exhibit long gestation and sustained growth, while others rise rapidly and later decline, underscoring the importance of a life-cycle approach.[4]

A central feature of the technology life cycle is that the earliest stage of a prominent technology is small in patent counts, yet those patents are unusually influential. Patents associated with prominent technologies are systematically higher quality than the average USPTO patent: they are more likely to receive forward citations and receive more citations, are more likely to be designated as breakthroughs (Kelly et al., 2021), and are more original in the sense of drawing on a broader set of prior technologies (Hall et al., 2001). This quality

[4]Table C1 reports the five largest technologies in each decade, ranked by the number of patents. Table A1 lists the earliest and latest prominent technologies in our sample, ordered by their year of initiation. Figure 5 visualizes the life-cycle trajectories of each prominent technology.

premium is concentrated early in the life cycle. Patents granted before emergence—and, more broadly, before top-100 prominence—are especially likely to receive citations and to qualify as breakthroughs, consistent with a small set of foundational inventions anchoring subsequent technological development. Large assignees produce higher-quality patents on average, but within prominent technologies their relative advantage is concentrated in the earliest stage and attenuates after prominence, suggesting a shift from foundational invention toward more incremental or strategic patenting as technologies mature.

Motivated by this concentration of influence in the earliest patents, we examine who holds the foundations of prominent technologies. Across the 289 technologies, roughly half of pre-emergence patents are assigned to large innovators (top 1% of firm assignees by grant-year patent counts), and the corresponding share is even larger for pre-prominence patents.[5] Most of the remainder is held by other firms, especially diversified non-top-1% innovators, while universities, government entities, and individual inventors account for relatively small shares, particularly in recent cohorts.[6] Taken together, these patterns indicate that the foundational patents underlying prominent technologies are typically held by firms—especially large innovators—rather than by universities or individual inventors. This finding motivates our next question: whether firms that enter a technology early also exhibit distinctive corporate policies and valuation dynamics as the technology approaches prominence.

We ask whether early innovators—firms that patent early in a technology—exhibit distinctive corporate policies and valuation dynamics in the run-up to prominence. Relative to other firms in the same technology-year, they invest more in capital expenditures and R&D, grow faster in assets, sales, and employment, and command higher valuation ratios well

[5]To construct long-run assignee histories, we combine assignee information from Gross and Sampat (2023) for 1920–1979 with PatentsView for post-1976 patents, harmonize assignee identities across sources, and match assignees to Compustat firms using a new release of the Global Corporate Patent Dataset; see Appendix D.

[6]A founder-centric narrative need not imply that early *assignees* are individuals. Our evidence is based on patent *assignees* (owners), whereas many canonical "founder" inventions are assigned to firms, or reassigned to firms shortly after invention, as commercialization proceeds. Illustrative examples include the telephone (US 174465), the Apple II microcomputer (US 4136359A), and PageRank (US 6285999B1), alongside inventions developed within large incumbents such as the transistor (US 2569347A), the hard disk drive (US 3134097), and the microprocessor (US 3753011A).

before prominence, with the strongest evidence in asset market-to-book. These differences are most pronounced in the earliest stage of the technology life cycle. As prominence approaches, valuation premia compress even as operating performance improves, especially in profit margins and return on invested capital, consistent with markets shifting from pricing growth options under uncertainty toward pricing realized fundamentals as diffusion proceeds. We establish these results in regressions with firm fixed effects and technology-by-year fixed effects, controlling for firm size and age. The estimates therefore capture how the corporate-policy and valuation consequences of early participation evolve over a technology's life cycle, net of time-invariant firm heterogeneity and technology-year conditions common to all firms.

Having documented strong pre-prominence advantages for early innovators, we next ask how durable those advantages are once the technology diffuses broadly and matures. Because technologies differ substantially in diffusion speed and patenting intensity, calendar time is not a comparable measure of maturity across technologies. We therefore organize each technology's post-prominence history into equal-sized patent batches, where the initial batch contains all patents granted up to and including the prominence year. This volume-based "technology clock" maps very different diffusion paths into comparable maturity stages. We find that, across these stages, patent quality declines systematically, consistent with innovation becoming less foundational as technologies mature.

We ask whether early assignees retain technological leadership as a technology evolves. Early assignees begin the post-prominence period with substantially higher within-technology patent shares. This advantage erodes only slowly on average, but its persistence is highly heterogeneous: large pioneers start with much larger shares and retain them more durably than non-large pioneers, broader prior technological scope slows the erosion of early leads, and intense early competition among large innovators both reduces the initial lead and speeds its decay. Together, these results indicate that early entry alone does not guarantee sustained dominance after prominence: persistence is concentrated among large, capable innovators and is shaped by the competitive environment at the onset of the technology.

These post-prominence shifts in innovation type and technological leadership are mirrored in firm outcomes. As technologies mature, early innovators invest less and grow more slowly than other firms in the same technology-year, with the largest shortfalls in later maturity stages. Their earlier valuation premia eventually turn into discounts, especially for pre-emergence innovators, while capital efficiency and profit margins also become less favorable relative to later entrants in the same technology-year. More broadly, the corporate-finance benefits of early participation fade and often reverse as technologies move deeper into maturity.

Taken together, our results point to a systematic life-cycle pattern in how transformative technologies translate into firm value. The earliest patents that seed prominent technologies are unusually influential and are disproportionately held by large, established innovators and other firms rather than universities or individual inventors. Early innovators invest more, grow faster, and command valuation premia long before a technology becomes prominent, but those premia compress as uncertainty resolves and technologies diffuse. After prominence, innovation becomes less foundational, early patent-share leadership becomes harder to sustain, and the corporate advantages of early participation fade or reverse as capital efficiency, margins, and valuation become less favorable relative to later entrants. These findings provide new evidence on when—and for whom—technological leadership translates into sustained rents and offer a unified empirical framework for studying how corporate investment, growth, and valuation evolve during periods of major technological change.

These findings connect to several strands of literature. First, our paper builds on work that uses patent data to measure innovation and to quantify its economic significance.[7] Recent finance and macro studies leverage these data to link innovation to reallocation and growth (Kogan et al., 2017) and to measure technological change over long horizons using patent text (Kelly et al., 2021). We complement this work by constructing an interpretable, patent title-based technology dictionary with consistent coverage since 1836, dating tech-

[7]Classic references include Griliches (1990), Trajtenberg (1990), Hall et al. (2005), and the harmonized citation database of Hall et al. (2001).

nologies' life-cycle milestones, and linking those dynamics to firms' investment, growth, profitability, and market valuation across a broad set of prominent technologies.

A second strand studies how technologies emerge, diffuse, and mature. Work on general purpose technologies emphasizes that broad technologies can reshape downstream innovation and growth (Bresnahan and Trajtenberg, 1995), while evidence on diffusion highlights substantial heterogeneity in adoption dynamics across technologies and settings (Comin and Hobijn, 2010; Kalyani et al., 2025). In parallel, the product life-cycle literature documents systematic evolution in entry, exit, and innovation as markets mature (Klepper, 1996), and theories of recombinant growth emphasize that much invention reflects new combinations of existing building blocks (Weitzman, 1998; Fleming, 2001). We contribute by dating initiation, emergence, and prominence for hundreds of technologies and by tracing how patent characteristics and firm outcomes evolve across these life-cycle stages.

A third strand examines competition in innovation and whether pioneering translates into persistent rents. Schumpeterian growth models formalize creative destruction (Aghion and Howitt, 1992), and classic innovation-race and strategic-patenting models show how early leads can be defended or displaced depending on incentives and the competitive environment (Reinganum, 1981; Gilbert and Newbery, 1982); work on appropriability and complementary assets underscores why inventors may fail to capture value even when they innovate first (Teece, 1986). A related literature documents long-run changes in the organization of production and innovation, including the rise of "superstar" firms (Autor et al., 2020b), higher markups and market power (De Loecker and Eeckhout, 2020), slower idea production (Bloom et al., 2020), and declining business dynamism (Akcigit and Ates, 2023). We contribute evidence at the innovator–technology level that links these themes by showing how innovation becomes less foundational after prominence, how early assignees' within-technology patent-share advantages erode, and how the persistence of those advantages is concentrated among large early innovators and shaped by early competition intensity.

In finance, a closely related literature studies how technological change and uncertainty

are capitalized into stock prices and valuation ratios. Classic episodes such as the IT revolution illustrate how market valuations can move far ahead of realized diffusion (Hobijn and Jovanovic, 2001), and theory links technological revolutions and uncertainty about future profitability to valuation dynamics (Pástor and Veronesi, 2009, 2006). Investment-based asset-pricing models emphasize that shocks to technology can differentially affect assets in place versus growth opportunities (Berk et al., 1999; Papanikolaou, 2011; Kogan and Papanikolaou, 2014), and a large empirical literature studies how R&D and intangible capital relate to market value and Tobin's q (Chan et al., 2001; Peters and Taylor, 2017). Our paper contributes cross-technology evidence on the timing and evolution of valuation premia around technology milestones, while simultaneously linking those valuation dynamics to the underlying evolution of investment, growth, and operating performance.

Finally, a broad corporate-finance literature examines how financing conditions, information frictions, governance, and organizational form shape innovative investment and outcomes. Work documents the role of internal funds and equity issuance in financing innovation (Brown et al., 2009), how financial development affects innovation (Hsu et al., 2014), and how the information environment and control rights influence innovative activity (He and Tian, 2013; Atanassov, 2013). Other research emphasizes the role of firm boundaries, corporate venture capital, and M&A in shaping innovative output (Seru, 2014; Chemmanur et al., 2014; Bena and Li, 2014). We contribute to this literature by conditioning corporate-finance outcomes on both where a firm sits in a technology's life cycle and when it entered the technology, thereby connecting canonical innovation-finance mechanisms to the life-cycle dynamics of transformative technologies.

2 Technology Construction and Firm Matching

2.1 Sources of Patent Data

To construct technologies from patent titles, we start from the universe of the United States Patent and Trademark Office (USPTO) utility patents granted between 1836 and 2023 and assemble a unified patent-title file spanning the full period. Patent titles are a natural basis for identifying technologies and tracking their life cycles because they are produced under legal and administrative constraints that encourage short, specific phrasing that is indicative of the invention's subject matter. In particular, USPTO regulations require a "title of the invention" that is "as short and specific as possible" and "reasonably indicative of the subject matter of the invention" (37 C.F.R. §1.72). Together with the disclosure and claiming requirements of U.S. patent law (35 U.S.C. §112), these rules incentivize applicants to foreground the key technological elements embodied in the claimed invention while minimizing narrative detail. This structure makes titles comparatively stable and interpretable over long horizons, allowing us to define technologies as recurring multi-word expressions and to date initiation, emergence, and prominence using consistent text-based definitions rather than technology classification systems that evolve over time.

For patents granted from 1976 onward, we use PatentsView,[8] a USPTO-supported platform that provides structured bibliographic records, including grant dates, titles, CPC classifications, and assignee identifiers. For patents granted before 1976, structured title metadata are not systematically available. We therefore collect titles from Google Patents,[9] which hosts scanned PDFs of original patent documents and reports bibliographic fields on each patent's HTML page, often derived via optical character recognition (OCR). OCR-derived titles contain systematic noise (e.g., non-title text captured as titles, character-level errors, and truncation).

Appendix B details our correction pipeline: (i) dictionary-based normalization of re-

[8] Available at https://patentsview.org/.
[9] Available at https://patents.google.com/.

curring OCR artifacts; (ii) NLP-based screening rules to flag implausible or incomplete titles; (iii) supplementation of flagged/missing titles using MTurk transcription and targeted ChatGPT-assisted extraction from the PDF header; and (iv) iterative re-screening. A stratified manual audit of 100 patents per grant-decade indicates accuracy above 98% in each decade. We then merge the cleaned pre-1976 records with PatentsView to obtain titles for 11,804,242 granted patents from 1836–2023 (title-record coverage of 99.56% through 2023); the remaining missing records correspond primarily to patents subject to secrecy orders or otherwise unavailable in digitized records.[10]

2.2 Extraction of Technologies from Patent Titles

Using the unified patent title file, we construct a dictionary of title-based technologies and assign each patent the set of technologies mentioned in its title. Our objective is an interpretable and time-consistent representation of technological content that does not rely on administrative classification systems. Because titles often reference multiple components, our approach allows a single patent to map to multiple technologies.

We implement a text-processing pipeline that converts each raw title into a set of candidate multi-word expressions and then filters these candidates to retain terms that behave like stand-alone technological concepts (Appendix C). We first normalize titles to ensure consistent tokenization across time and data sources (e.g., lowercasing, removing parenthetical content, standardizing punctuation and whitespace, and lemmatizing nouns using a standard lemmatizer augmented with patent-specific dictionaries). We then segment each title into short phrases using an iteratively constructed list of separators—primarily prepositions and conjunctions that typically link distinct components (e.g., "with," "for," "and")—so that extracted terms do not combine distinct components of the invention. Within each phrase, we remove a curated set of general and patent-specific stopwords.

From each resulting phrase, we generate all contiguous n-grams and drop single-word

[10]For example, certain inventions are subject to secrecy orders under 35 U.S.C. §181.

terms, which are typically too imprecise to identify distinct technologies. We then standardize common formatting differences, most importantly hyphenation, by mapping hyphenated variants to a single canonical form (e.g., "type-writing machine" and "typewriting machine"). To limit noise from OCR artifacts and idiosyncratic fragments, we impose minimum-frequency requirements and remove generic invention-type descriptors (e.g., "method," "process"). Finally, we apply an additional usage-based filter that drops n-grams that rarely appear as complete phrases in titles and instead primarily occur as modifiers or partial fragments (e.g., "high speed," "internal combustion"). Appendix C provides the formal definitions and thresholds used in these filters.

2.3 Technology Coverage and Descriptive Statistics

The output of the procedure in Section 2.2 is a vocabulary of 715,788 multi-word n-grams, which we interpret as discrete technologies. Each patent is assigned the set of these technologies that appear in its processed title. Because titles are available in a consistent format since the start of the modern U.S. patent system, this approach produces a single, transparent technology layer spanning 1836–2023. The resulting technology labels are directly interpretable, which makes it straightforward to inspect face validity and to track technologies from initiation to maturity without relying on classification systems whose definitions change over time.[11]

Coverage is broad in both the long run and the modern period that anchors our empirical analysis. At least one retained technology is assigned to 69% of granted patents between 1836 and 2023. Coverage exceeds 50% after 1870 and rises to roughly 80% in the most recent decade (Figure C1a). This pattern indicates that the title-based vocabulary captures the

[11]For context: the International Patent Classification (IPC) was established under the Strasbourg Agreement (1971), with the first edition becoming effective in 1968, and is revised over time; see https://www.wipo.int/en/web/classification-ipc/faq. The Cooperative Patent Classification (CPC) is a joint USPTO–EPO system in use since January 1, 2013; see https://www.uspto.gov/patents/search/classification-standards-and-development. CPC updates are implemented via periodic releases and associated reclassification of documents; see, e.g., https://www.cooperativepatentclassification.org/CPCRevisions (CPC "Notices of Changes").

key technological content for the majority of patents, including in historical periods where structured classification systems are unavailable or not comparable.

Coverage is also balanced across fields of technology. Figure C1b reports coverage by one-digit CPC section and shows a relatively narrow range (58%–79%), suggesting that the method is not concentrated in a small subset of domains but instead captures technological content across diverse areas of invention. This broad coverage is important for our life-cycle analysis because it limits the risk that measured technology dynamics reflect field-specific measurement artifacts.

Finally, the technology labels are easy to interpret and align closely with well-known technological eras. Table C1 lists the five most prevalent technologies by decade. Early decades are dominated by mechanically intuitive terms such as *steam engine* and *sewing machine*; much of the twentieth century is dominated by *internal combustion engine*; and recent decades feature ICT-related terms such as *communication system*, *display device*, and *electronic device*. This interpretability is central to our approach: it facilitates transparent validation, supports case-based intuition for life-cycle patterns, and enables technology trajectories to be tracked consistently from initiation through emergence and prominence using a stable, text-based definition.

2.4 Firm-level Data and Firm–Technology Matching

A central contribution of our data is that we can follow technologies—defined from patent titles and tracked from initiation through emergence and prominence—and link the associated patent histories to firm identities and financial outcomes. To do so, we assemble long-run assignee information and then map assignees to public firms in Compustat.

Patent assignee data. For the historical period, we use the assignee name strings (and assignee types) constructed by Gross and Sampat (2023), which provide harmonized assignee information for U.S. patents granted from 1920 through 1979. For 1976 onward,

assignee name strings are available from USPTO's structured patent data (accessed through PatentsView), which provides cleaned, disambiguated assignee identifiers and names for granted patents from 1976 onward.[12]

Matching assignees from 1976 onward to Compustat. We first build the patent–firm linkage for the 1976-onward period using PatentsView assignee organizations and firm records from Compustat North America and Compustat Global. This step follows the methodology in Appendix D and yields a new release of the Global Corporate Patent Dataset (GCPD). In brief, we (i) normalize firm and assignee names; (ii) generate candidate links using fuzzy string similarity and web-identifier overlap; (iii) aggregate candidates into matches using pre-specified acceptance criteria, supplemented with targeted manual validation; and (iv) apply timing corrections so that matches respect corporate events and name changes. The output is a high-coverage mapping from PatentsView assignees to Compustat firm identifiers, enabling us to link patenting activity to firm balance-sheet and market outcomes.

Extending the linkage to pre-1976 assignees. We then extend this mapping to the pre-1976 assignee strings from Gross and Sampat (2023). We begin with automated linkage for assignee names that also appear in the 1976-onward data: after standardizing name strings using the same normalization rules as in Appendix D, we directly inherit the established Compustat match for any historical assignee that can be reliably mapped to a PatentsView-era assignee identity. This step covers the majority of economically relevant historical assignees and ensures continuity of firm identifiers across the pre-1976 and post-1976 regimes.

For assignees that cannot be linked reliably through this automated bridge, we focus manual effort where identification quality is most consequential. In particular, for the set of *prominent technologies* defined in Section 4.1—the core sample for our analysis of technology

[12]PatentsView is built from USPTO bulk patent data and provides disambiguated assignee entities for granted patents since 1976.

life cycles and firm outcomes—we manually match every assignee appearing on patents in these technologies. This targeted hand matching yields high-precision links to Compustat for the firms involved in the initiation, emergence, and rise to prominence of major technologies, allowing us to connect technology-specific patent histories to accounting and market outcomes for 9,047 global public firms.

3 Stylized Facts on Technological Change since 1870

This section uses our patent-title-based technology measures to recover three long-run patterns documented in prior work: invention becomes more recombinant, patenting becomes more concentrated in a small set of dominant technologies, and large innovators play an increasing role within those technologies. We use the full patent record from 1870 onward for the first two facts. Because reliable assignee coverage begins in 1920, the third fact focuses on 1920–2023. The period from 1870 onward corresponds to the conventional start of the Second Industrial Revolution and the rise of more systematic invention and denser patenting (Mokyr, 1998; Meisenzahl and Mokyr, 2011).

Fact 1: Rising technological recombination. Figure 1 shows that patents increasingly combine multiple technologies within a single invention. Panel (a) plots the distribution of patents by the number of distinct retained title-based technologies they contain. Before the mid-twentieth century, most patents contain only one such technology. Since then, the share of multi-technology titles rises steadily, and by 2023 a majority of patents contain multiple technologies, with a sizable right tail of titles containing five or more technologies.[13] Interpreting co-occurrence in titles as evidence of combining technological components is consistent with theories of recombinant growth and search (Weitzman, 1998; Fleming, 2001). Panel (b) shows that recombination also increases along the extensive margin: the number

[13] Figure C2 shows that the average number of technologies per title increases from roughly 1.5 in 1950 to about 3 by 2023.

of distinct pairwise technology combinations observed in a year, scaled by the number of patents granted that year, rises sharply after 1980 and peaks in the late 2010s at roughly one distinct pairwise combination for every three patents. The analogous series for *new* pairwise combinations follows the same time profile but remains substantially smaller, indicating that rising recombination reflects both reuse of established components and continued introduction of new combinations.[14]

Fact 2: Growing concentration in dominant technologies. Figure 2 shows that patenting has become increasingly concentrated in a small set of dominant technologies. Panel (a) ranks technologies each year by *fractionalized* patent counts, where a patent that lists N technologies contributes $1/N$ to each of them. This adjustment prevents the rise of multi-technology titles documented in Fact 1 from mechanically inflating measured concentration. The figure groups technologies into four annual rank tiers (top 0.1%, top 0.1–1%, top 1–10%, bottom 90%) and shows a pronounced U-shaped pattern: concentration declines through the mid-twentieth century and then rises sharply in recent decades, with the top 0.1% and top 0.1–1% tiers accounting for an increasingly large share of fractionalized patent mass.

Panel (b) provides a complementary view by fixing the number of dominant technologies and plotting the fraction of patents that contain at least one of the top 10 or top 100 technologies in a given year. This share likewise falls through the mid-century period and then reverses into a strong upward trend, especially after the early 2000s. Together, Panels (a) and (b) show that rising concentration is not an artifact of multi-technology titles. Instead, patenting has increasingly tilted toward a relatively small set of dominant technologies, consistent with broader evidence on rising concentration in innovation and market leadership (Autor et al., 2020a; Akcigit and Ates, 2023).

[14]Figure C3 shows that both patent volume and the number of observed combinations increase over time. Figure 1, Panel (b) isolates the fact that combinations grow faster than patents, so combinations per patent rises.

Fact 3: Large innovators' growing role within dominant technologies. Finally, we show that the rising concentration of patenting in dominant technologies is mirrored by a growing role for large innovators within those technologies. Using assignee data, which are available with reliable coverage from 1920 onward, we define *large innovators* each year as the top 1% of firm assignees by grant-year patent counts.[15] Figure 3 plots the share of patents assigned to these large innovators, both in the full patent universe and within dominant technologies. The benchmark share for all patents trends upward, but the large-innovator share is systematically higher within top technologies than in the patent universe as a whole. Moreover, the gap widens markedly after the late twentieth century: by the 2000s–2020s, large innovators account for more than three-quarters of patents in the top technology tiers, compared with roughly one-half across all patents. This pattern is consistent with evidence that innovative activity and patenting have become increasingly concentrated among large incumbents (Akcigit and Ates, 2023).[16]

Taken together, Figures 1–3 show that our title-based technology measures recover three canonical long-run patterns: invention becomes more recombinant, patenting becomes more concentrated in a small set of dominant technologies, and large innovators account for a rising share of patents—especially within those dominant technologies. Recovering these stylized facts in a unified, interpretable, classification-free framework provides external validation for our technology construction and motivates its use in the life-cycle and firm-outcome analyses that follow.

4 Prominent Technologies

So far, we have shown that patenting has become increasingly concentrated in a small set of dominant technologies and that large innovators account for a growing share of patenting

[15] Figure A1 reports the annual number of distinct firm assignees that fall in the top 1% by grant-year patent counts.

[16] As a robustness check, Figure A2 reproduces the same trend using PatentsView assignee data, yielding closely similar dynamics.

within them. We now move from aggregate patterns to technology life cycles by focusing on *prominent technologies*—those that enter the annual top 100 by patent count at least once. We study how these technologies emerge and diffuse, who their early innovators are, how often early innovators are large innovators, and whether large-innovator presence reflects early entry and persistent leadership or later-stage entry as technologies mature.

4.1 Sample Construction and Life-Cycle Milestones

Our sample consists of title-based technologies (n-grams) that meet two criteria: (i) the initiation year—the grant year of the first patent—is 1920 or later, which ensures reliable assignee coverage over the full observed life cycle; and (ii) the technology enters the annual top 100 by patent count in at least one year through 2023, indicating that it attains substantial prominence in the patent record. To ensure that the resulting set of technologies is distinct and interpretable, we manually review candidate n-grams. If two n-grams refer to the same underlying technology but differ only in phrasing (e.g., "catalytic conversion" vs. "catalytic converter"), we map them to a common canonical form. If one n-gram strictly contains another (e.g., "machine learning" vs. "machine learning model"), we retain the more general term to avoid redundant nesting.[17]

These criteria yield 289 prominent technologies.[18] Table A1 lists the ten prominent technologies with the earliest initiation years and the ten with the most recent initiation years. Five technologies begin in 1920, the first year in our sample, whereas four begin in the twenty-first century. These endpoints include technologies as varied as "automotive vehicle" and "heat transfer" at the beginning of the sample and "additive manufacture," "organic light emit display," "social network," and "deep learn" at the recent end.

To characterize technology life cycles, we date three milestones for each technology. The *initiation year* is the grant year of the technology's first patent. The *emergence year* is the

[17] We manually verify all n-grams that ever rank among the top 150 to ensure comprehensive coverage of those that appear in the top 100.

[18] Figure A3 plots the share of patents accounted for by the prominent technologies in our sample. These technologies account for 11.4% of patents granted after 1920, and their share rises to 16.8% in 2023.

first year in which the technology attains at least five granted patents. The *prominence year* is the first year in which the technology ranks among the annual top 100 technologies by patent count. Figure 4 shows substantial heterogeneity in the speed with which technologies reach these milestones. Some become prominent almost immediately after initiation (e.g., "combine typewrite" and "aerosol insecticide"), whereas others exhibit very long gestation (e.g., "wireless device" reaches emergence 72 years after initiation and prominence 84 years after initiation). The median prominent technology reaches emergence after 10 years and prominence after 29 years (Table 1, Panel A). The earliest phase is small but potentially pivotal: the average technology has 20 pre-emergence patents and 483 pre-prominence patents (Table 1, Panels B–C).

Figure 5 visualizes these life-cycle trajectories by plotting each technology's patent share—relative to all patents granted in a given year—against technology age. Technologies are grouped by initiation cohort; within each panel, each line traces one technology, and the highlighted line marks the representative technology with the highest patent share in that cohort. The figure highlights pronounced heterogeneity in diffusion dynamics. For example, "electronic device" follows a slow trajectory, reaching emergence after roughly 15 years and prominence about 65 years after initiation, while continuing to grow to about 1.5% of all patents by 2023. In contrast, "semiconductor device" and "liquid crystal" rise to prominence within about five years but later decline in patent share. These patterns motivate our subsequent analysis of early innovators and how participation by large innovators evolves over the technology life cycle.

4.2 Patent Quality Across Life-Cycle Stages

We next examine whether patents associated with prominent technologies—and especially those granted early in a technology's life cycle—are systematically higher quality.[19] We

[19]This analysis uses patents granted from 1947 onward because the USPTO did not consistently record patent-to-patent citations prior to 1947. We compute citations using the EPO Worldwide Patent Statistical Database (PATSTAT), which provides standardized, machine-readable patent citation links at scale; see

estimate

$$Y_k = \beta \, ProminentTech_k + \delta_{c,t} + \alpha_{i,t} + \epsilon_k, \tag{1}$$

where k indexes patents, c denotes 3-digit CPC technology class, i denotes assignees, and t denotes grant year. $ProminentTech_k$ is an indicator equal to one if patent k contains at least one prominent technology. The outcome Y_k is one of four patent-level measures: (i) $\log(1 + Citations)$, where *Citations* is the count of forward citations received within five years of grant; (ii) $\mathbb{1}(Citations > 0)$, an indicator for receiving at least one citation within five years; (iii) $\mathbb{1}(Breakthrough)$, an indicator for patents classified as breakthroughs by Kelly et al. (2021); and (iv) *Originality*, which measures the diversity of technologies cited by the patent following Hall et al. (2001).[20] Table 2 (Panel A) reports summary statistics for these variables. We include 3-digit CPC-by-year fixed effects $\delta_{c,t}$ and assignee-by-year fixed effects $\alpha_{i,t}$ to absorb time-varying differences across technology classes and assignees, and we cluster standard errors by grant year.

Table 4 reports estimates of equation (1). In Panel A, the coefficient on $ProminentTech$ is positive and statistically significant at the 1% level across all outcomes, indicating that patents associated with prominent technologies are higher quality on average: they are more likely to receive citations, receive more citations, are more frequently classified as breakthroughs, and are slightly more original. In economic terms, relative to sample means, prominent-technology patents are, for example, about 7% more likely to receive at least one citation within five years and about 16% more likely to be classified as breakthroughs.

To pinpoint where in the technology life cycle this premium arises, Panel B augments equation (1) by interacting $ProminentTech$ with indicators for early-stage patents: $PreEmergence$ ($PreProminence$) equals one if a patent is granted on or before the emer-

https://www.epo.org/en/searching-for-patents/business/patstat. To our knowledge, there is no comparable comprehensive source that provides consistent USPTO citation links at scale before 1947.

[20] We construct *Originality* using title-based technologies rather than CPC patent classifications, whose definitions and boundaries change over time, as pointed out by Balsmeier et al. (2017).

gence (prominence) year of the earliest prominent technology referenced in its title. The interaction coefficients are again highly statistically significant for key outcomes and show that the quality premium is concentrated early: relative to sample means, pre-emergence (pre-prominence) patents are about 52% (26%) more likely to be cited and about 154% (64%) more likely to be classified as breakthroughs.

We then ask whether the early-stage quality premium is driven by large innovators. Table 5 reports estimates of equation (1) augmented with an indicator for innovator size, $Assignee^{\text{Large}}$, which equals one if the patent's assignee ranks in the top 1% of firm assignees by grant-year patent counts (our definition of *large innovators* in Section 3). In Panel A, the coefficient on $Assignee^{\text{Large}}$ is positive and statistically significant at the 1% level for key outcomes, implying that large innovators produce higher-quality patents on average. For example, relative to sample means, their patents are about 4% more likely to receive at least one citation within five years and about 12% more likely to be classified as breakthroughs. However, this advantage is substantially weaker within prominent technologies: the interaction term $ProminentTech \times Assignee^{\text{Large}}$ is negative and statistically significant at the 1% level across citation, breakthrough, and originality outcomes, largely offsetting the unconditional large-innovator premium among patents associated with prominent technologies.

Panel B shows that this attenuation masks sharp life-cycle heterogeneity. To gauge magnitudes, consider the unconditional means in Table 2 (Panel A): the probability a patent receives at least one five-year citation is about 0.65, and the breakthrough rate is about 0.05. Against these baselines, the *unconditional* large-innovator premium in Panel A of Table 5 is economically modest—about a 4% higher citation likelihood and a 12% higher breakthrough likelihood—corresponding to roughly a +2.4 percentage point increase in $\mathbb{1}(Citations > 0)$ (0.65×0.04) and a +0.6 percentage point increase in $\mathbb{1}(Breakthrough)$ (0.05×0.12). In contrast, the *pre-emergence* large-innovator advantage is much larger: the estimates imply roughly a 21% higher citation likelihood and a 52% higher breakthrough likelihood relative to the unconditional average, which translates into about a +13.5 percentage point increase

in the citation indicator (0.65×0.21) and a +2.6 percentage point increase in breakthrough incidence (0.05×0.52). Put differently, large innovators are only slightly more likely than average to produce cited or breakthrough patents overall, but they are disproportionately responsible for the most influential patents in the pre-emergence phase. After emergence and prominence, this relative advantage declines and can reverse, consistent with a shift from foundational invention toward more incremental or strategic patenting as technologies mature.

4.3 Composition of Early Assignees

We next characterize who holds patents in the earliest stages of prominent technologies, distinguishing between two early windows. *Pre-emergence assignees* are assignees with at least one patent granted on or before a technology's emergence year, and *pre-prominence assignees* are assignees with at least one patent granted on or before its prominence year.

We begin with technology-level illustrations. Figure A4 shows the pre-emergence assignees of four representative technologies and highlights substantial institutional heterogeneity. Individual inventors account for a sizable share of early activity in *electronic device*, while universities such as Carnegie Mellon and MIT appear prominently among early assignees in *machine learning*. Yet large innovators (top 1% of firm assignees by grant-year patent counts) appear among the early assignees in all four examples. Consistent with this pattern, Table A2 shows that large, high-volume innovators such as AT&T and RCA rank among the leading early assignees, both by the number of pre-emergence patents and by the number of distinct prominent technologies in which they appear early.[21]

Figure 6 summarizes these patterns across all 289 prominent technologies by reporting the assignee composition of pre-emergence patents (Panel (a)) and pre-prominence patents (Panel (b)) by initiation cohort.[22] Panel (a) shows that the examples above reflect broad

[21] Figure A5 illustrates the set of prominent technologies in which IBM, RCA, AT&T, and GE are pre-emergence assignees.

[22] Table 1 reports summary statistics on the number of patents and assignees in the pre-emergence and

regularities: large innovators account for roughly half of pre-emergence patents on average, and their share rises for technologies initiated in more recent cohorts. The remainder is accounted for primarily by other firms, including a substantial contribution from non-top-1% diversified innovators, while universities and government entities represent small shares. Individual inventors are present among early assignees in older cohorts but decline sharply over time.

Panel (b) shows that similar patterns hold when expanding the window to include all pre-prominence patents. Large innovators account for an even larger share of pre-prominence patents than of pre-emergence patents, consistent with growing participation by large innovators as technologies diffuse from emergence to prominence. Non-top-1% firms continue to account for a meaningful share—particularly diversified innovators—while universities, government, and individuals remain comparatively small contributors. Taken together, the evidence indicates that the patent foundations of prominent technologies are typically laid within established firms, with large innovators playing a central and increasingly important role both at the earliest stage (pre-emergence) and through the run-up to prominence (pre-prominence).

These patterns motivate the firm-level analysis that follows. Because a substantial share of early patents in prominent technologies is held by firms—especially large innovators—rather than by universities or individual inventors, we focus on early assignees that can be linked to public firms in Compustat. We compare pre-emergence and pre-prominence public-firm innovators to other public firms active in the same technology-year, both in the decades leading up to prominence and during the post-prominence maturity phase.

pre-prominence windows, as well as the distribution of assignee types across technologies.

5 Early Innovators Across the Technology Life Cycle

5.1 Firm Outcomes Before Prominence

We next turn to firm-level outcomes and ask whether early innovators exhibit distinct patterns of investment, growth, operating performance, and valuation before a technology becomes prominent. Our sample comprises 9,047 global public firms matched to at least one patent in at least one prominent technology by 2023. For each firm–technology pair, we retain all calendar years on or after the technology's initiation for which firm financial data are available in Compustat.[23] Table 3 reports summary statistics for the firms in this sample.

To characterize how firm outcomes differ by entry cohort, we estimate the following linear model for firm i, technology j, and year t:

$$\begin{aligned} Y_{i,t} = {} & \beta_1 \, \text{Assignee}_{\text{PreEmergence},i,j} \times PreProminence_{j,t} \times PreEmergence_{j,t} \\ & + \beta_2 \, \text{Assignee}_{\text{PreEmergence},i,j} \times PreProminence_{j,t} + \beta_3 \, \text{Assignee}_{\text{PreEmergence},i,j} \\ & + \gamma_1 \, \text{Assignee}_{\text{PreProminence},i,j} \times PreProminence_{j,t} \times PreEmergence_{j,t} \\ & + \gamma_2 \, \text{Assignee}_{\text{PreProminence},i,j} \times PreProminence_{j,t} + \gamma_3 \, \text{Assignee}_{\text{PreProminence},i,j} \\ & + \delta \, X_{i,t} + \alpha_i + \mu_{j,t} + \epsilon_{i,j,t}, \end{aligned} \tag{2}$$

where the unit of observation is a firm–technology–year and $Y_{i,t}$ is a firm-level outcome. In different specifications, $Y_{i,t}$ measures investment intensity, growth, operating performance, or valuation, in either levels or changes. For consistency with the patent data, the explanatory variables retain the "Assignee" notation, but in this section they identify matched public firms. $\text{Assignee}_{\text{PreEmergence},i,j}$ equals one if firm i's first patent in technology j is granted on or before the technology's emergence year. $\text{Assignee}_{\text{PreProminence},i,j}$ equals one if firm i's first patent in technology j is granted after emergence but on or before the technology's prominence year, so the two early-entry cohorts are mutually exclusive. The omitted category therefore con-

[23] Figure A6 plots sample coverage.

sists of firms whose first patent in the technology occurs after prominence. $PreEmergence_{j,t}$ and $PreProminence_{j,t}$ indicate years on or before technology j's emergence year and prominence year, respectively. Thus, for pre-emergence innovators the implied effect is $\beta_1+\beta_2+\beta_3$ in the pre-emergence stage, $\beta_2+\beta_3$ between emergence and prominence, and β_3 after prominence; analogous expressions apply to the pre-prominence cohort. We control for firm age and lagged assets, along with their squared terms ($X_{i,t}$), include firm fixed effects (α_i) and technology-by-year fixed effects ($\mu_{j,t}$), and double-cluster standard errors at the firm and technology levels.

Table 6 reports estimates of equation (2).[24] The results show that early innovators invest more and grow faster well before a technology becomes prominent. In Panel A, the implied total effects for pre-emergence innovators in the pre-emergence stage are economically large: capital expenditure intensity is 3.3 percentage points higher, R&D intensity is 0.6 percentage points higher, and asset, sales, and employment growth are higher by 3.3, 2.9, and 1.9 percentage points, respectively. Pre-prominence innovators—firms that first patent between emergence and prominence—display the same qualitative pattern, although the magnitudes are generally smaller. The strongest differences therefore arise in the earliest stage of the technology life cycle.

Panel B shows that these differences reflect active adjustment rather than purely static level differences. During the run-up to prominence, early innovators increase capital expenditure intensity more rapidly than later entrants, and they also accelerate growth in assets, sales, and employment. These differential changes are especially clear in the broader pre-prominence window, indicating that early innovators are not merely firms with permanently high levels of investment and growth, but firms that actively ramp up as the technology develops toward prominence.

The same table also indicates that these advantages do not persist after prominence. In post-prominence years, both $PreEmergence_{j,t}$ and $PreProminence_{j,t}$ equal zero, so the

[24] Table A5 reports the corresponding estimates for the North American subsample.

coefficients on the uninteracted early-entry indicators capture how pre-emergence and pre-prominence innovators compare with post-prominence entrants once the technology has matured. Across the investment and growth outcomes, these coefficients are typically negative, indicating that after prominence early innovators invest less and grow more slowly than later entrants in the same technology-year. This reversal reinforces the life-cycle interpretation: early innovators concentrate capital deployment, R&D spending, and organizational expansion in the stages leading up to prominence rather than after diffusion is already well advanced. The investment and growth premia associated with early entry are therefore concentrated in the early stages of the technology life cycle rather than reflecting permanent firm characteristics.

The operating-performance results are more nuanced but clearly not flat. In the levels specifications (Panel A, Columns 6–8), return on assets (ROA) differences are generally small, but early innovators exhibit higher profit margins and, especially, higher return on invested capital (ROIC) before prominence. The ROIC effects are particularly consistent across the two early-entry cohorts. In the changes specifications (Panel B, Columns 6–8), early innovators also improve ROA and margins more quickly than later entrants during the run-up to prominence, whereas changes in ROIC are smaller. Taken together, these results indicate that the strongest pre-prominence operating-performance differences lie in margins, capital efficiency, and the pace of improvement rather than in large level differences in ROA.

These operating patterns are reflected in valuation. The clearest result is in asset market-to-book. In Panel A, the implied total asset market-to-book premium for pre-emergence innovators is about 0.15 log points in the pre-emergence stage, and valuation premia remain positive though smaller closer to prominence. Firms that eventually enter between emergence and prominence also command higher asset market-to-book ratios before prominence. The price-to-earnings results are weaker in these coarse stage partitions, but they also point to positive valuation dynamics during the pre-prominence window. Overall, the valuation evidence is consistent with markets capitalizing both the growth options

associated with early technological leadership and the investment acceleration and rapid expansion that accompany it.

Figures 7 and 8 provide a more granular view of these dynamics.[25] They show that investment, growth, and valuation premia are visible well before prominence and that valuation premia compress as prominence approaches, even as operating performance improves. This pattern is consistent with markets gradually shifting from pricing uncertain growth options to pricing realized fundamentals as diffusion proceeds. We further decompose these trajectories in Figures A8 through A11 by distinguishing between firms that remain active after prominence ("survivors") and firms that later exit the technology. This split delivers two additional insights. First, valuation premia are already present before prominence for both groups, indicating that markets initially price the general option value of exposure to an emerging technology rather than only the prospects of eventual winners. Second, the premium compresses more slowly for eventual survivors. This pattern suggests that investors gradually update from valuing broad technological opportunity to valuing the probability that a given firm will retain a durable position as the technology matures.

Taken together, these results show that early innovation is associated with a distinctive pre-prominence corporate profile: high investment intensity, rapid scaling, improving operating performance, and elevated valuation. These differences are strongest in the earliest stage of the technology life cycle, and the associated valuation premia begin to compress as prominence approaches.

5.2 Persistence and Decay of Early Assignees' Patent Shares

Having characterized firm outcomes in the run-up to prominence, we next ask whether early assignees retain technological leadership as technologies move into post-prominence maturity. A central challenge is that technologies diffuse at very different speeds. As Section 4.1 shows, some technologies rise and mature quickly, whereas others unfold over several decades.

[25] Figure A7 complements Figures 7 and 8 by showing how the pre-prominence growth differentials cumulate into levels of assets, sales, and employment.

Calendar years since prominence therefore do not map cleanly into comparable maturity stages across technologies.

We address this problem using a volume-based "technology clock." For each technology, $Batch_0$ contains all patents granted up to and including the prominence year. We then partition all subsequent patents into sequential batches of equal size, so that each post-prominence batch contains the same number of patents as $Batch_0$. This construction places fast- and slow-diffusing technologies on a common maturity scale and allows us to compare leadership dynamics across equivalent stages of technological maturity. Through 2023, the 289 prominent technologies span an average of 22.6 batches; the typical batch contains about 210 patents and spans roughly four years.

We validate this batch structure by showing that patent quality declines systematically across batches, as one would expect over the technology life cycle. Figure A12 plots patent characteristics over the life cycles of two representative technologies, *liquid crystal* and *semiconductor device*. In both cases, early batches are more likely to contain cited patents, breakthroughs, and highly original inventions. Table A3 shows the same pattern in the full sample, and Table 7 confirms that these declines are statistically significant in regression specifications. Later batches therefore correspond to progressively more mature phases of the technology life cycle.

We then measure technological leadership using an assignee's share of patents within a technology-batch, $PatentShare_{i,j,b}$, and estimate

$$PatentShare_{i,j,b} = \beta_1 \text{Assignee}_{\text{PreEmergence},i,j} + \beta_2 \left(\text{Assignee}_{\text{PreEmergence},i,j} \times Batch_{j,b}\right) + \mu_{j,b} + \epsilon_{i,j,b}, \tag{3}$$

where the unit of observation is an assignee–technology–batch. The dependent variable $PatentShare_{i,j,b}$ is the fraction of patents in batch b of technology j assigned to assignee i. $\text{Assignee}_{\text{PreEmergence},i,j}$ equals one if assignee i holds at least one pre-emergence patent in technology j. $Batch_{j,b}$ denotes the batch number, scaled by 100 so that β_2 measures the change

in patent share over 100 batches. The baseline specification includes technology-by-batch fixed effects, and later specifications additionally absorb assignee fixed effects or assignee-by-year fixed effects. Because $Batch_0$ contains all patents through the prominence year, β_1 captures the initial patent-share advantage of pre-emergence assignees at prominence, while β_2 captures the rate at which that advantage erodes as the technology matures.

Panel A of Table 8 reports the estimates. The coefficient on $\text{Assignee}_{\text{PreEmergence},i,j}$ is positive and statistically significant across specifications, indicating that pre-emergence assignees reach prominence with materially larger within-technology patent shares than later entrants. Economically, the initial premium is about 1.1 to 1.6 percentage points, roughly twice the unconditional mean patent share. The interaction with batch is negative but economically small: patent share declines by only about 0.3 percentage points per 100 batches. Given that the average technology in our sample spans 22.6 batches, this estimate implies that only about 0.07 percentage points of the initial advantage erode over the observed life cycle. Thus, early assignees retain a substantial lead well into the maturity phase.

Taken together, these results show that early technological leadership is persistent on average. Pre-emergence assignees begin with substantially larger patent shares at prominence, and this advantage decays only slowly as technologies mature. The next subsection shows, however, that this average persistence masks important heterogeneity across assignees and across technologies.

5.3 Heterogeneity in the Persistence of Early Assignees' Advantage: Scope and Early Competition

The slow average decay documented above masks important heterogeneity. The literature on first-mover advantage emphasizes that persistence should depend not only on entry timing, but also on firm capabilities and on the competitive environment at the technology's onset (Teece, 1986; Lieberman and Montgomery, 1988; Cohen and Klepper, 1996; Klepper, 1996). We therefore examine three sources of heterogeneity in early assignees' patent-share

advantage: assignee size, broader prior technological scope, and early competition among large pioneers.

We first ask whether persistence differs between large and non-large pre-emergence assignees. To do so, we partition pre-emergence assignees into large and non-large pioneers, denoted by $\text{Assignee}^{\text{Large}}_{\text{PreEmergence}}$ and $\text{Assignee}^{\text{NonLarge}}_{\text{PreEmergence}}$, depending on whether the assignee belongs to the top 1% of firm assignees by grant-year patent counts in at least one year during the pre-emergence window. We then re-estimate equation (3), allowing both the initial patent-share premium and its subsequent decay to vary by size.

Panel B of Table 8 shows that large pioneers begin the post-prominence phase with a much larger patent-share advantage than non-large pioneers. In the baseline specifications, the initial premium for large pioneers is about 1.9 to 2.1 percentage points, compared with only about 0.4 to 0.5 percentage points for non-large pioneers. Once assignee fixed effects or assignee-by-year fixed effects are added, the two premia become much closer—roughly 1.1 to 1.4 percentage points for large pioneers and 0.8 to 1.2 percentage points for non-large pioneers. This convergence indicates that much of the unconditional gap reflects pre-existing assignee scale. At the same time, the decay estimates show that large pioneers' advantage is more durable: in the richer specifications, their patent-share premium declines by about 0.2 to 0.3 percentage points per 100 batches, compared with about 0.4 to 0.5 percentage points for non-large pioneers. Thus, large pioneers not only begin with larger shares on average, but also retain their lead more persistently as technologies mature.

We next examine whether broader prior technological scope helps pioneers sustain leadership as the focal technology matures. To do so, we interact the size indicators with $\log(\#\,Tech_{\text{PreEmergence}})$, the log number of technologies in which the focal assignee is already a pre-emergence assignee when the focal technology reaches emergence. Panel A of Table 9 shows a consistent trade-off between initial lead and persistence. Among large pioneers, broader scope is associated with a smaller initial patent-share advantage but a slower subsequent decay. The same pattern is weaker, but still visible in the richer specifications, for

non-large pioneers. Put differently, specialists tend to start with larger leads, whereas more diversified pioneers retain their leads for longer. This pattern is consistent with theories of recombinant innovation (Weitzman, 1998; Fleming, 2001): firms with broader knowledge bases may be better able to adapt and recombine the focal technology with adjacent domains as it matures.

Persistence also depends on the competitive environment at the technology's onset. When the set of large pioneers is small, an early leader may be better positioned to shape standards, build complementary capabilities, and retain a durable edge. When many large assignees enter early, rivalry is more intense and leadership should be harder to sustain. To test this mechanism, we construct an indicator for intense early competition equal to one if a technology has more than five large pre-emergence assignees, $\mathbb{1}(\#\text{Assignee}^{\text{Large}}_{\text{PreEmergence}} > 5)$, and interact it with the large- and non-large-pioneer indicators. Panel B of Table 9 shows that the initial patent-share advantage of large pioneers is substantially smaller in these high-competition technologies. Relative to low-competition technologies, their initial premium is lower by roughly 0.7 to 1.1 percentage points, or about one-third to two-fifths, depending on the specification. The corresponding interaction with $Batch_{j,b}$ indicates markedly faster erosion—an additional decline of about 0.8 to 1.1 percentage points per 100 batches. We do not find a comparable pattern for non-large pioneers. These results suggest that early rivalry among large pioneers materially limits any one firm's ability to translate temporal priority into persistent leadership.

Finally, we ask whether persistence has weakened in more recent technologies. Table A4 interacts the size-specific pioneer indicators with $\mathbb{1}(Initiation \geq 1980)$. For large pioneers, the interaction with $Batch_{j,b}$ is consistently negative, implying an additional decline of about 2.0 to 2.7 percentage points per 100 batches in technologies initiated after 1980. By contrast, the corresponding recent-cohort interactions for non-large pioneers are imprecisely estimated. Thus, while large pioneers still begin with sizable patent-share advantages in recent technologies, those advantages erode much faster than in earlier cohorts.

Taken together, these results show that the persistence of early patent-share leadership depends on both assignee capabilities and market structure. Early entry alone does not guarantee durable dominance. Persistence is strongest for large pioneers, more durable when those pioneers have broader prior technological scope, and much weaker when early competition among large innovators is intense. In addition, the advantages of large pioneers erode more quickly in recent technology cohorts.

5.4 Firm Outcomes After Prominence

Finally, we ask whether the erosion of early patent-share leadership is mirrored in firm outcomes after prominence. Returning to the firm–technology–year sample used for the pre-prominence outcome analysis, we estimate

$$\begin{aligned} Y_{i,t} = & \sum_{group \in \{1\text{–}5, 6\text{–}20, 21+\}} \beta_{group} \, \text{Assignee}_{\text{PreEmergence},i,j} \times Batch_{group,j,t} \\ & + \sum_{group \in \{1\text{–}5, 6\text{–}20, 21+\}} \gamma_{group} \, \text{Assignee}_{\text{PreProminence},i,j} \times Batch_{group,j,t} \\ & + \beta \, \text{Assignee}_{\text{PreEmergence},i,j} + \gamma \, \text{Assignee}_{\text{PreProminence},i,j} \\ & + \delta X_{i,t} + \alpha_i + \mu_{j,t} + \epsilon_{i,j,t}, \end{aligned} \tag{4}$$

where the unit of observation is a firm–technology–year. As in the previous subsection, the explanatory variables retain the "Assignee" notation for consistency with the patent data, but here they identify matched public firms. $Batch_{1\text{–}5,j,t}$, $Batch_{6\text{–}20,j,t}$, and $Batch_{21+,j,t}$ indicate whether technology j in calendar year t lies in an early, middle, or late post-prominence maturity stage.[26] The omitted category again consists of firms whose first patent in the technology occurs after prominence. Thus, for a pre-emergence innovator the implied effect in batch group g is $\beta_g + \beta$, and for a pre-prominence innovator it is $\gamma_g + \gamma$. As in equation (2), we control for lagged firm size and age and their squared terms ($X_{i,t}$), include firm fixed

[26] If a calendar year contains multiple batches of a technology, we classify that year using the minimum batch number.

effects (α_i) and technology-by-year fixed effects ($\mu_{j,t}$), and double-cluster standard errors at the firm and technology levels.

Table 10 reports the estimates.[27] The results show a clear reversal in the corporate profile of early innovators after prominence. In Panel A, the implied total effects for both pre-emergence and pre-prominence innovators are negative for capital expenditure intensity, R&D intensity, and the growth rates of assets, sales, and employment in every post-prominence batch group. These gaps also widen as technologies mature. For pre-emergence innovators, capital expenditure intensity is about 0.6 percentage points lower than for post-prominence entrants in $Batch_{1\text{–}5}$ and about 2.1 percentage points lower in $Batch_{21+}$. The corresponding asset-growth differential widens from about 0.9 to 2.1 percentage points, and the sales-growth differential from about 0.8 to 2.3 percentage points. Pre-prominence innovators display the same qualitative pattern, although the shortfalls are generally smaller early in maturity. Thus, once a technology passes prominence, the rapid expansion that characterizes early innovators before prominence gives way to slower investment and growth.

Panel B shows that these lower post-prominence levels reflect active retrenchment rather than merely static differences. Relative to post-prominence entrants, early innovators reduce capital expenditure intensity more rapidly and experience progressively weaker growth in assets, sales, and employment as technologies move from $Batch_{1\text{–}5}$ to $Batch_{21+}$. The R&D-change results are weaker than those for capital expenditures, but the overall pattern still points to deceleration after prominence rather than continued ramp-up.

Operating performance also becomes less favorable after prominence. In Panel A, ROIC is lower for both early-entry cohorts throughout the post-prominence period, and profit margins also become less favorable, especially for pre-emergence innovators and in later batch groups. ROA differences are smaller, but they are generally negative as maturity deepens. In Panel B, changes in operating performance are more mixed than the investment and growth reversals, yet the drift is again downward: changes in margins, and to a lesser

[27] Table A6 reports the corresponding estimates for the North American subsample.

extent ROA, become less favorable in later maturity stages, particularly for pre-emergence innovators.

Financial markets mirror this reversal. The clearest result is again in asset market-to-book. For pre-emergence innovators, the implied asset market-to-book differential turns negative immediately after prominence and becomes substantially more negative as technologies mature, moving from about -0.02 log points in $Batch_{1\text{–}5}$ to about -0.08 in $Batch_{21+}$. Pre-prominence innovators show the same pattern, with the implied discount widening from about -0.02 to about -0.05 log points. The price-to-earnings results are weaker in levels, but the changes specifications also become more negative in later post-prominence stages, especially for pre-prominence innovators. The valuation premia documented before prominence therefore compress and often reverse as technologies move deeper into maturity.

Figures 9 and 10 visualize these post-prominence dynamics across investment, growth, operating performance, and valuation.[28] They make clear that the negative differentials widen as technologies move from early to late post-prominence maturity. Figures A14 through A17 split the sample by whether firms remain active after prominence. These figures show that the deterioration is more gradual for firms that continue innovating in the technology than for firms that later exit. Finally, Figures 11 and 12 report analogous life-cycle profiles for *active innovators*—firms that receive at least one patent in a given technology-year—and provide a benchmark for the characteristics of the firms currently driving innovation at each stage of the technology life cycle.

Taken together, the post-prominence results complete the life-cycle pattern documented in this section. Early innovators invest more, grow faster, and command higher valuations before prominence, but these advantages fade as technologies mature. After prominence, early innovators retrench relative to later entrants, their growth slows, their capital efficiency and margins weaken, and their valuation premia compress and often reverse. These results mirror the erosion of early patent-share leadership and show that the corporate-finance

[28] Figure A13 complements Figures 9 and 10 by showing how the post-prominence growth differentials cumulate into levels of assets, sales, and employment.

benefits of early participation are concentrated in the early stages of the technology life cycle.

6 Conclusion

We construct a new dataset of technology life cycles from the universe of USPTO patent titles spanning 1836–2023. By defining technologies as recurring multi-word phrases in patent titles, we obtain an interpretable, classification-free technology vocabulary with consistent long-run coverage. This framework recovers three familiar aggregate patterns in technological change: invention becomes increasingly recombinant, patenting becomes increasingly concentrated in a small set of dominant technologies, and large innovators account for a growing share of patenting within those dominant areas.

We then trace 289 prominent technologies—those initiated in 1920 or later that enter the annual top 100 by patent count at least once through 2023—and link their patent histories to 9,047 global public firms. The earliest patents in these technologies are disproportionately influential and are primarily assigned to firms, especially large innovators. Early assignees also reach prominence with materially larger within-technology patent shares than later entrants. This leadership advantage erodes only slowly on average, but it is more durable for large pioneers and for pioneers with broader prior technological scope, and much weaker when early competition among large innovators is intense.

Early participation in a technology is associated with a distinct pre-prominence corporate profile. Relative to later entrants in the same technology-year, firms that patent early invest more, grow faster, and command higher valuation ratios long before a technology becomes prominent. These differences are strongest in the earliest stage of the technology life cycle. As prominence approaches, valuation premia compress even as operating performance improves, consistent with markets shifting from pricing uncertain growth options to pricing realized fundamentals as diffusion proceeds.

After prominence, the pattern reverses. Innovation becomes less foundational, early patent-share leadership becomes harder to sustain, and the investment, growth, and valuation advantages of early participation fade and often reverse as technologies mature. Early innovators retrench relative to later entrants, their capital efficiency and margins become less favorable, and the valuation premia that characterize the run-up to prominence compress and often turn into discounts. Taken together, these findings provide new evidence on when—and for whom—technological leadership translates into sustained rents and offer a unified empirical framework for studying how corporate investment, growth, and valuation evolve during periods of major technological change.

References

Aghion, P. and Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, 60(2):323–351.

Akcigit, U. and Ates, S. T. (2023). What happened to us business dynamism? *Journal of Political Economy*, 131(8):2059–2124.

Atanassov, J. (2013). Do hostile takeovers stifle innovation? evidence from antitakeover legislation and corporate patenting. *The Journal of Finance*, 68(3):1097–1131.

Autor, D., Dorn, D., Hanson, G. H., Pisano, G., and 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., and 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.

Balsmeier, B., Fleming, L., and Manso, G. (2017). Independent boards and innovation. *Journal of Financial Economics*, 123(3):536–557.

Bena, J., Ferreira, M. A., Matos, P., and Pires, P. (2017). Are foreign investors locusts? the long-term effects of foreign institutional ownership. *Journal of Financial Economics*, 126(1):122–146.

Bena, J. and Li, K. (2014). Corporate innovations and mergers and acquisitions. *The Journal of Finance*, 69(5):1923–1960.

Berk, J. B., Green, R. C., and Naik, V. (1999). Optimal investment, growth options, and security returns. *The Journal of Finance*, 54(5):1553–1607.

Bloom, N., Hassan, T. A., Kalyani, A., Lerner, J., and Tahoun, A. (2021). The diffusion of new technologies. Technical report, National Bureau of Economic Research.

Bloom, N., Jones, C. I., Van Reenen, J., and Webb, M. (2020). Are ideas getting harder to find? *American Economic Review*, 110(4):1104–1144.

Bresnahan, T. F. and Trajtenberg, M. (1995). General purpose technologies "engines of growth"? *Journal of Econometrics*, 65(1):83–108.

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

Buffett, W. E. (2008). Letter to the shareholders of berkshire hathaway inc. (2007 annual report). Berkshire Hathaway Inc.

Chan, L. K. C., Lakonishok, J., and Sougiannis, T. (2001). The stock market valuation of research and development expenditures. *The Journal of Finance*, 56(6):2431–2456.

Chemmanur, T. J., Loutskina, E., and Tian, X. (2014). Corporate venture capital, value creation, and innovation. *The Review of Financial Studies*, 27(8):2434–2473.

Cohen, W. M. and Klepper, S. (1996). A reprise of size and r & d. *The Economic Journal*, 106(437):925–951.

Comin, D. and Hobijn, B. (2010). An exploration of technology diffusion. *American Economic Review*, 100(5):2031–2059.

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

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Science*, 47(1):117–132.

Gilbert, R. J. and Newbery, D. M. G. (1982). Preemptive patenting and the persistence of monopoly. *American Economic Review*, 72(3):514–526.

Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature*, 28(4):1661–1707.

Gross, D. P. and Sampat, B. N. (2023). America, jump-started: World war ii r&d and the takeoff of the us innovation system. *American Economic Review*, 113(12):3323–3356.

Gundlach, J. (2025). Jeffrey gundlach says almost all financial assets are now overvalued. Odd Lots (Bloomberg) podcast transcript.

Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2001). The NBER patent citation data file: Lessons, insights and methodological tools. Working Paper 8498, National Bureau of Economic Research.

Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2005). Market value and patent citations. *The RAND Journal of Economics*, 36(1):16–38.

He, J. and Tian, X. (2013). The dark side of analyst coverage: The case of innovation. *Journal of Financial Economics*, 109(3):856–878.

Hobijn, B. and Jovanovic, B. (2001). The information-technology revolution and the stock market: Evidence. *American Economic Review*, 91(5):1203–1220.

Hsu, P.-H., Tian, X., and Xu, Y. (2014). Financial development and innovation: Cross-country evidence. *Journal of Financial Economics*, 112(1):116–135.

Kalyani, A., Bloom, N., Carvalho, M., Hassan, T., Lerner, J., and Tahoun, A. (2025). The diffusion of new technologies. *The Quarterly Journal of Economics*, 140(2):1299–1365.

Kelly, B., Papanikolaou, D., Seru, A., and Taddy, M. (2021). Measuring technological innovation over the long run. *American Economic Review: Insights*, 3(3):303–320.

Klepper, S. (1996). Entry, exit, growth, and innovation over the product life cycle. *American Economic Review*, 86(3):562–583.

Kogan, L. and Papanikolaou, D. (2014). Growth opportunities, technology shocks, and asset prices. *The Journal of Finance*, 69(2):675–718.

Kogan, L., Papanikolaou, D., Seru, A., and Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics*, 132(2):665–712.

Lieberman, M. B. and Montgomery, D. B. (1988). First-mover advantages. *Strategic management journal*, 9(S1):41–58.

Meisenzahl, R. R. and Mokyr, J. (2011). The rate and direction of invention in the british industrial revolution: Incentives and institutions. In Lerner, J. and Stern, S., editors, *The Rate and Direction of Inventive Activity Revisited*, pages 443–479. University of Chicago Press, Chicago.

Mokyr, J. (1998). The second industrial revolution, 1870–1914. In Castronovo, V., editor, *Storia dell'economia Mondiale*, pages 219–245. Laterza, Rome.

Papanikolaou, D. (2011). Investment shocks and asset prices. *Journal of Political Economy*, 119(4):639–685.

Pástor, L. and Veronesi, P. (2006). Was there a Nasdaq bubble in the late 1990s? *Journal of Financial Economics*, 81(1):61–100.

Pástor, L. and Veronesi, P. (2009). Technological revolutions and stock prices. *American Economic Review*, 99(4):1451–1483.

Peters, R. H. and Taylor, L. A. (2017). Intangible capital and the investment-q relation. *Journal of Financial Economics*, 123(2):251–272.

Reinganum, J. F. (1981). On the diffusion of new technology: A game theoretic approach. *The Review of Economic Studies*, 48(3):395–405.

Seru, A. (2014). Firm boundaries matter: Evidence from conglomerates and R&D activity. *Journal of Financial Economics*, 111(2):381–405.

Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy*, 15(6):285–305.

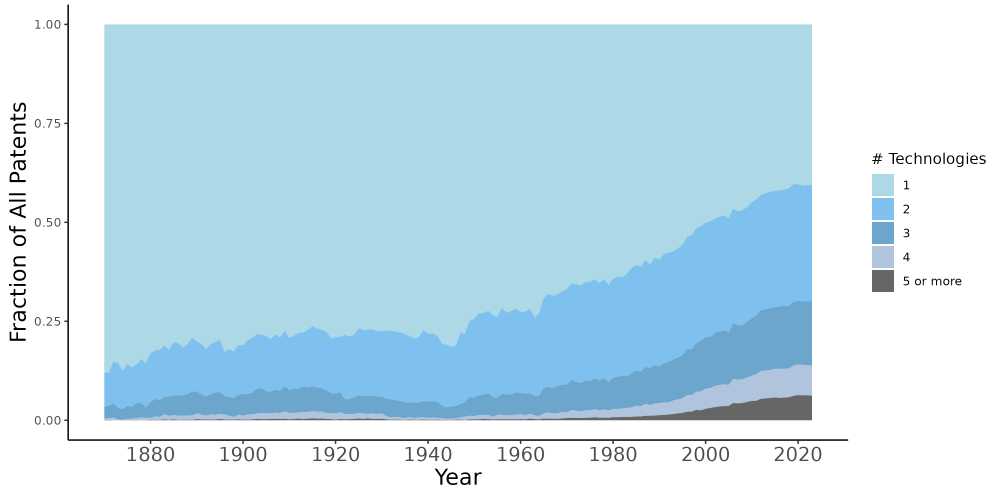
Trajtenberg, M. (1990). A penny for your quotes: Patent citations and the value of innovations. *The RAND Journal of Economics*, 21(1):172–187.

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

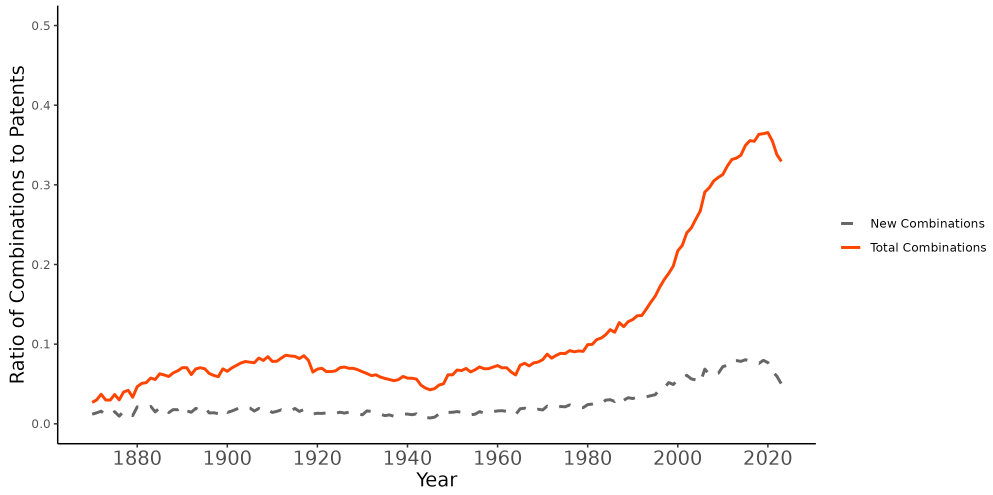
Figures and Tables

Figure 1: Technological Recombination

(a) Distribution of Patents by the Number of Technologies



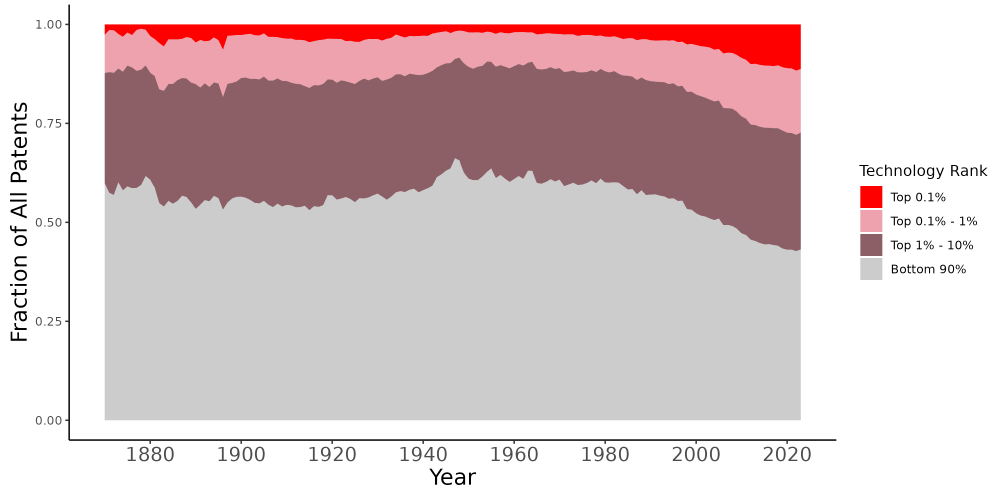
(b) Distinct Pairwise Technology Combinations per Patent



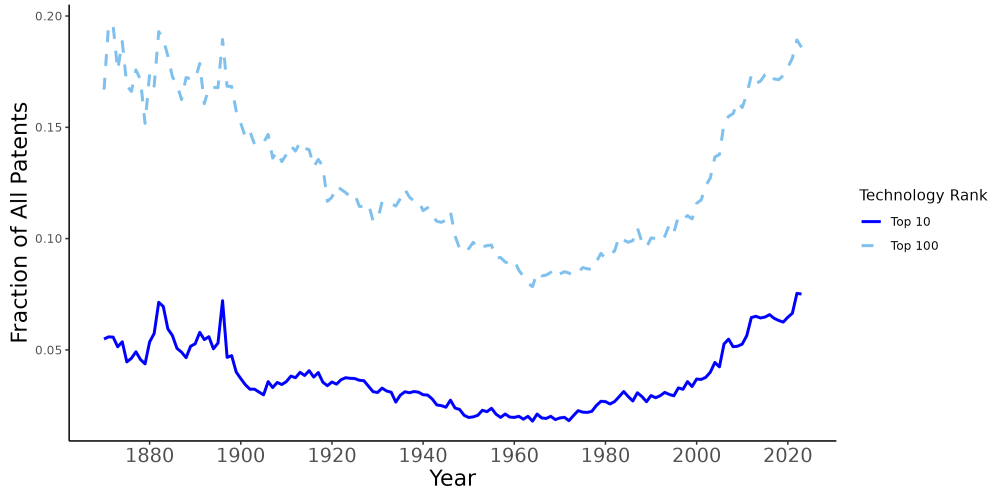
Note: This figure documents the growth of technological recombination in patent titles, 1870–2023. Panel (a) reports the distribution of patents by the number of distinct title-based technologies they mention, expressed as a share of total patents in each year. Panel (b) tracks pairwise co-occurrence of technologies within titles. The solid orange line plots the ratio of unique technology pairs observed in a given year to the number of patents granted that year. The dashed gray line plots the analogous ratio for new technology pairs—pairs that appear for the first time in that year in at least two patents.

Figure 2: Concentration of Patenting in Dominant Technologies

(a) Distribution of Patents by Technology Rank



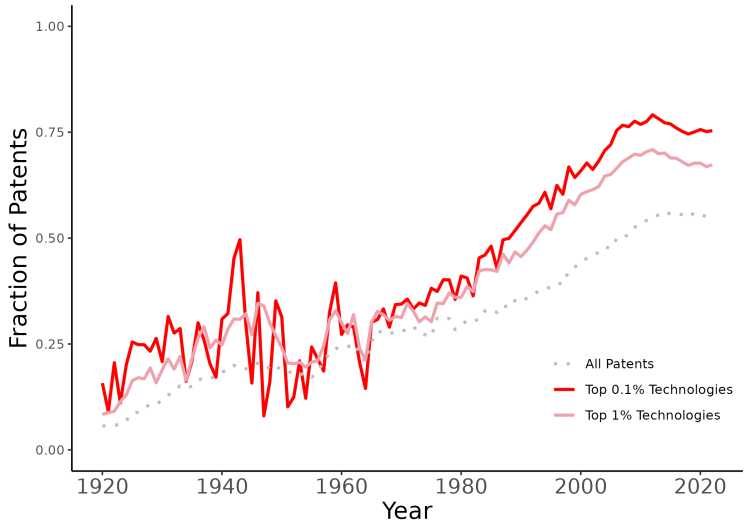
(b) Fraction of Patents in Top-10 and Top-100 Technologies



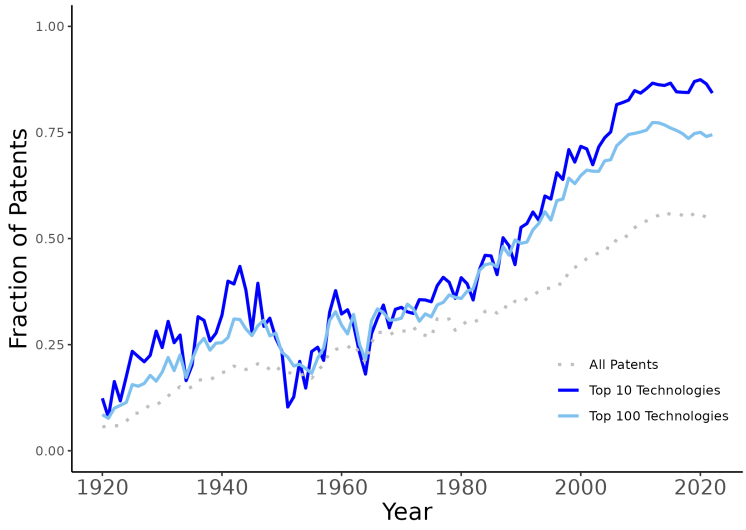
Note: This figure tracks the concentration of patenting across title-based technologies, 1870–2023. Panel (a) ranks technologies each year by patent count (using fractional counting when titles contain multiple technologies) and reports the share of fractionalized patent mass accounted for by technologies in the top 0.1%, top 0.1–1%, top 1–10%, and bottom 90% of the annual rank distribution. Panel (b) reports the fraction of patents that contain at least one of the top 10 or top 100 technologies in the corresponding year.

Figure 3: Large Innovators' Share of Patents in Dominant Technologies

(a) Top-0.1% and Top-1% Technologies

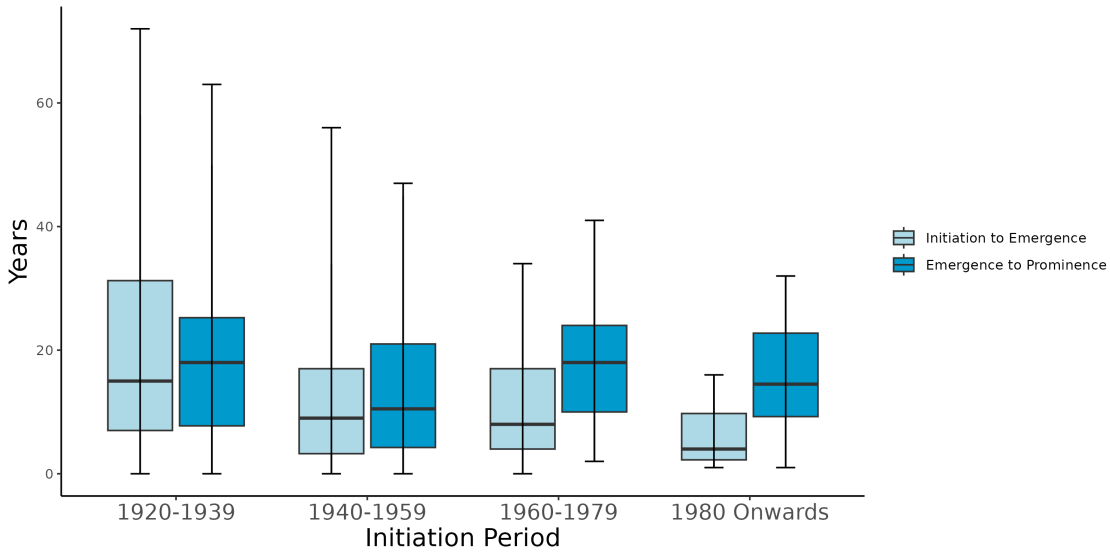


(b) Top-10 and Top-100 Technologies



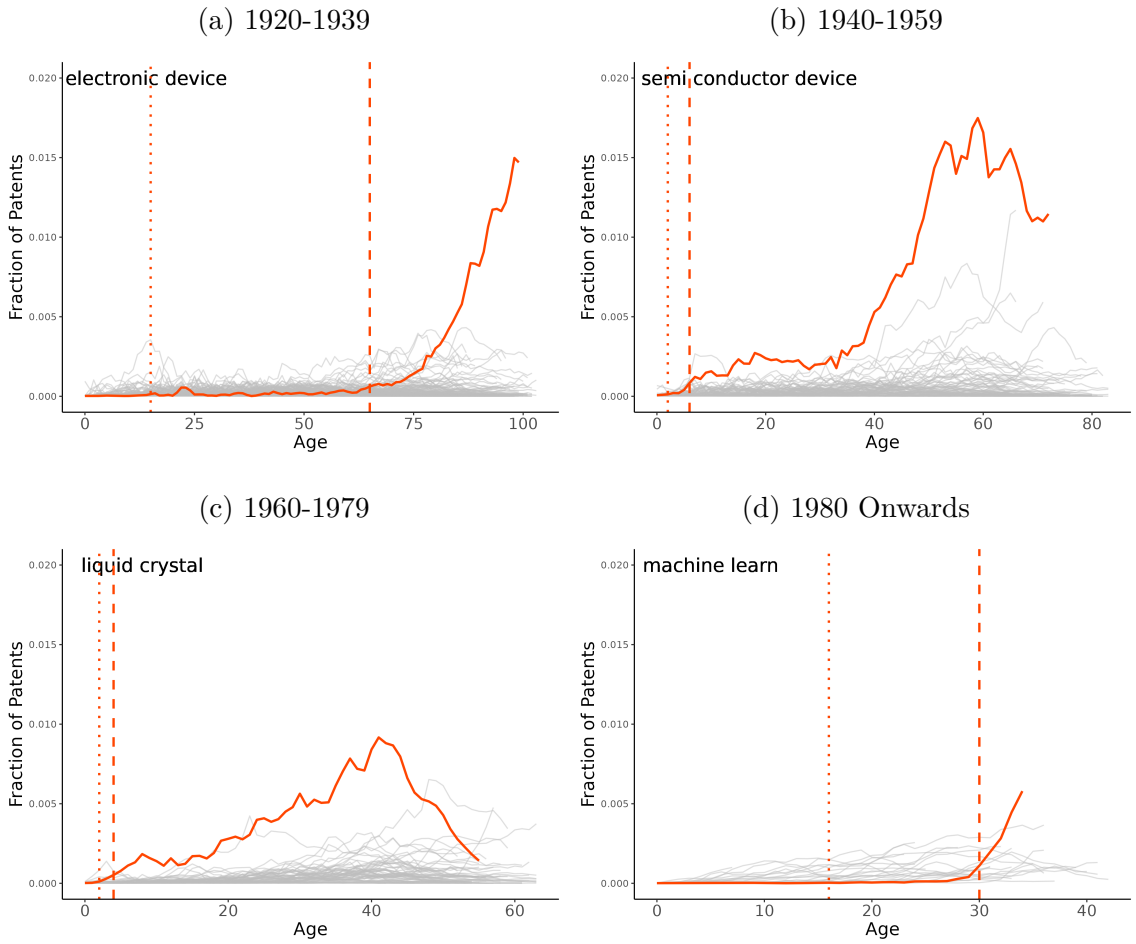
Note: This figure reports the share of patents assigned to large innovators—defined each year as the top 1% of firm assignees by grant-year patent counts—and how this share differs for patents associated with dominant technologies. Panel (a) plots the large-innovator share for patents that contain technologies in the top 0.1% (red) or top 1% (pink) of the annual technology rank distribution. Panel (b) plots the corresponding shares for patents that contain at least one of the top 10 (blue) or top 100 (light blue) technologies in that year. In both panels, the gray dotted line is the benchmark large-innovator share across all patents granted in the same year.

Figure 4: Time from Initiation to Emergence and Prominence: Prominent Technologies



Note: This figure reports, for each of the 289 prominent technologies (defined in Section 4.1), the time from initiation to emergence (light blue) and from emergence to prominence (dark blue). Technologies are grouped into four initiation-year cohorts (x-axis). Within each cohort, the box shows the interquartile range (25th–75th percentiles), the horizontal line marks the median, and whiskers indicate the minimum and maximum values.

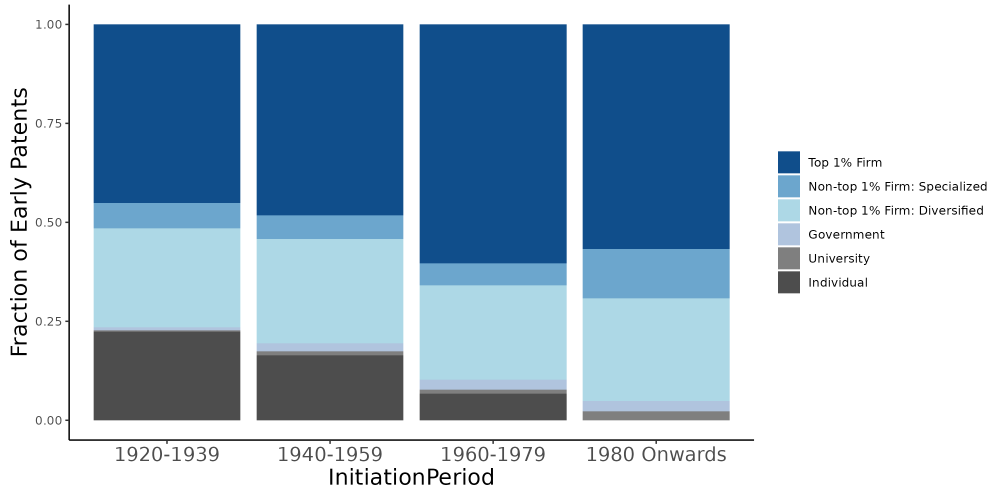
Figure 5: Life-Cycle Patent-Share Trajectories: Prominent Technologies



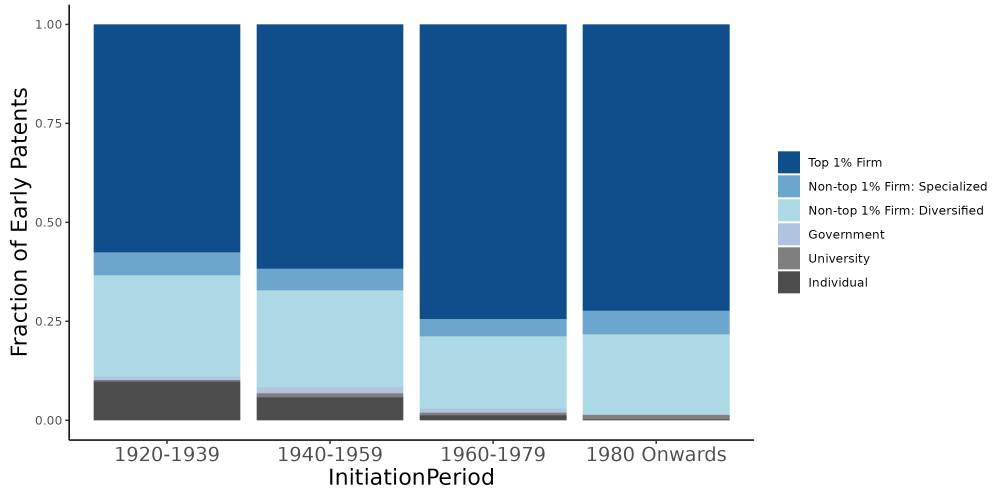
Note: This figure plots life-cycle trajectories for the 289 prominent technologies defined in Section 4.1. Technologies are grouped by initiation cohort, with panels (a)–(d) showing all technologies initiated in each cohort; each line corresponds to one technology. The x-axis is technology age (years since initiation). The y-axis is the technology's annual patent share, defined as the number of patents associated with the technology in a given year divided by the total number of USPTO patents granted that year. In each panel, one representative technology is highlighted in orange. Vertical dotted and dashed lines mark the ages at the technology's emergence and prominence years, respectively.

Figure 6: Composition of Early Assignees in Prominent Technologies: Pre-emergence vs. Pre-prominence

(a) Pre-emergence Patents



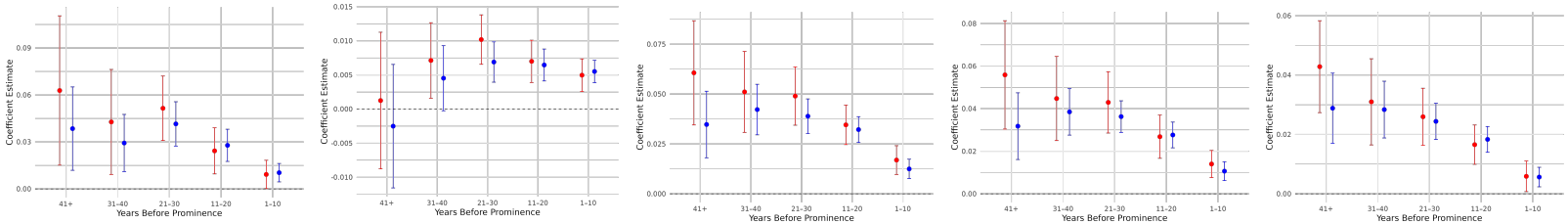
(b) Pre-prominence Patents



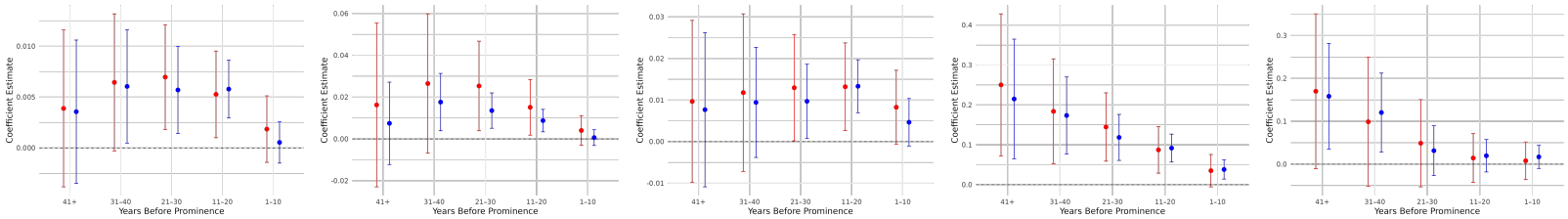
Note: This figure reports the assignee composition of early patents in the 289 prominent technologies defined in Section 4.1, separately for pre-emergence patents (Panel (a)) and pre-prominence patents (Panel (b)). Technologies are grouped by initiation cohort, and within each cohort the stacked bars show the fraction of early patents assigned to six assignee types: large innovators, non-top-1% specialized firms, non-top-1% diversified firms, government, universities, and individual inventors. *Large innovators* are firm assignees that rank in the top 1% of firm assignees by grant-year patent counts (defined at the patent's grant year). Among the remaining firm assignees, we classify an assignee as *specialized* if it has no patents in title-based technologies outside the focal prominent technology, and as *diversified* otherwise.

Figure 7: Firm Outcomes Before Prominence: Pre-emergence and Pre-prominence Innovators (Levels)

(a) $\frac{CapEx_t}{PPENT_{t-1}}$ (b) $\frac{R\&D_t}{Asset_{t-1}}$ (c) $\log\left(\frac{Asset_{t+1}}{Asset_t}\right)$ (d) $\log\left(\frac{Sales_{t+1}}{Sales_t}\right)$ (e) $\log\left(\frac{Empl_{t+1}}{Empl_t}\right)$



(f) ROA_t (g) $Margin_t$ (h) $ROIC_t$ (i) $\log(\text{Asset M/B}_t)$ (j) $\log(\text{P/E}_t)$

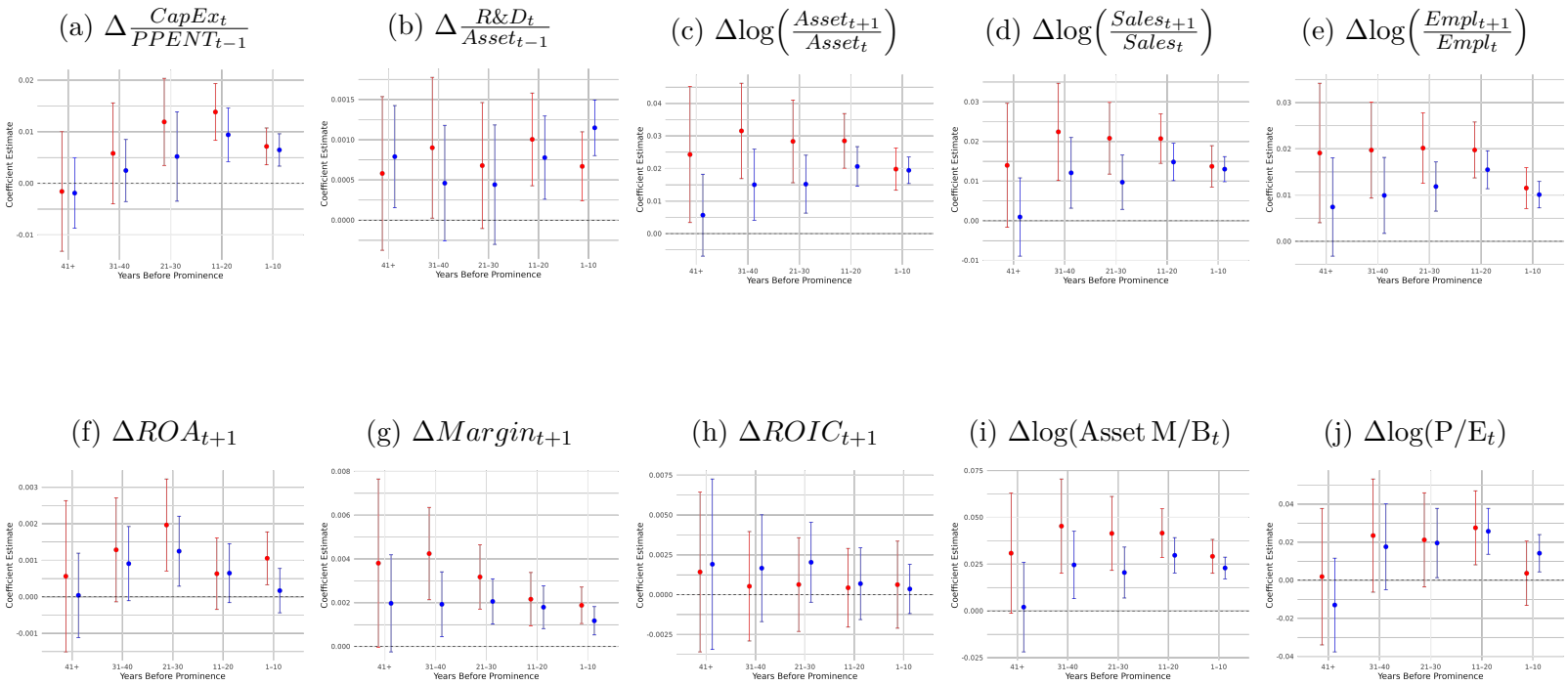


Note: This figure reports coefficient estimates from linear regression

$$
\begin{aligned}
Y_{i,t} = & \sum_{group \in \{1\text{–}10, 11\text{–}20, 21\text{–}30, 31\text{–}40, 41+\}} \beta_{group} \, \text{Assignee}_{\text{PreEmergence},i,j} \times Decade_{group,j,t} \\
& + \sum_{group \in \{1\text{–}10, 11\text{–}20, 21\text{–}30, 31\text{–}40, 41+\}} \gamma_{group} \, \text{Assignee}_{\text{PreProminence},i,j} \times Decade_{group,j,t} \\
& + \beta \, \text{Assignee}_{\text{PreEmergence},i,j} + \gamma \, \text{Assignee}_{\text{PreProminence},i,j} \\
& + \delta X_{i,t} + \alpha_i + \mu_{j,t} + \epsilon_{i,j,t},
\end{aligned}
\tag{5}
$$

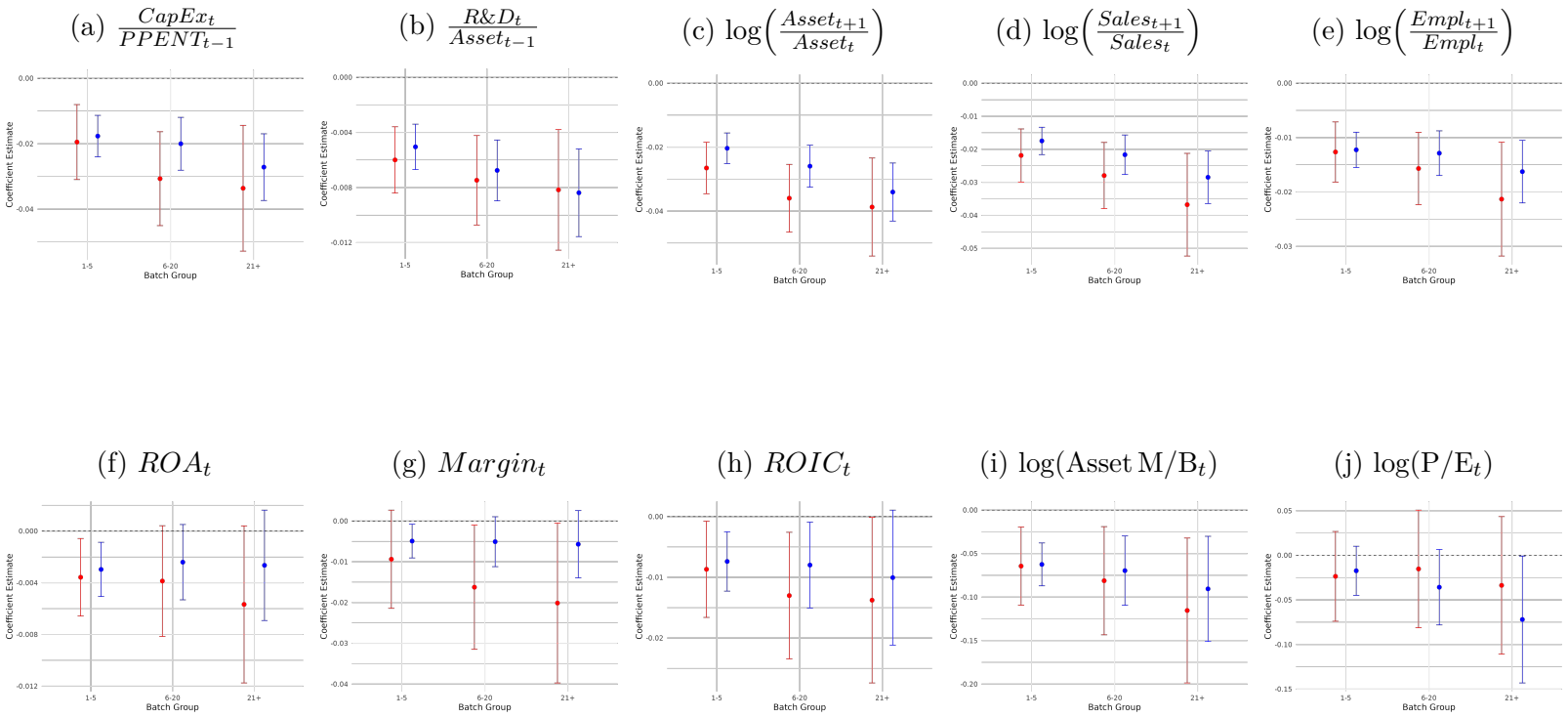
where i, j, and t represent firm, technology and year, respectively. The dependent variable $Y_{i,t}$ corresponds to the outcome indicated in the title of each panel. The dummy variable $\text{Assignee}_{\text{PreEmergence},i,j}$ ($\text{Assignee}_{\text{PreProminence},i,j}$) indicates whether firm i is a pre-emergence (pre-prominence) assignee of technology j. $Decade_{group,j,t}$ is dummy variable indicating decades before prominence, as shown by the x-axis of each figure. We control for firm fixed effects (α_i), technology-by-year fixed effects ($\mu_{j,t}$) and firm-year level variables ($X_{i,t}$) including firm age and assets, along with their squared terms. We double-cluster standard errors at the firm and technology levels. The red (blue) points present the estimated coefficient for β_{group} (γ_{group}) for the specific decade, with vertical lines indicating the 95% confidence interval.

Figure 8: Firm Outcomes Before Prominence: Pre-emergence and Pre-prominence Innovators (Changes)



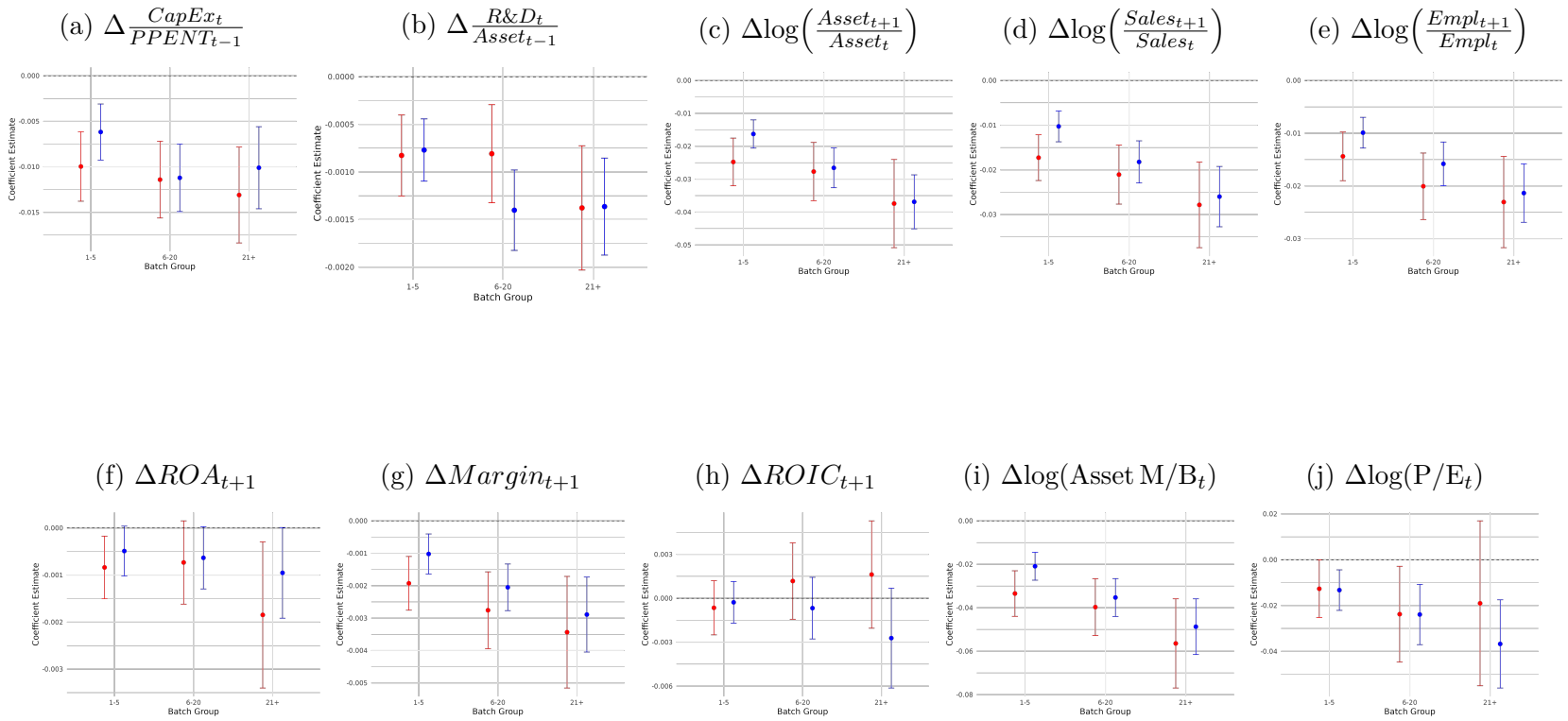
Note: This figure reports coefficient estimates from equation (5). The dependent variable $Y_{i,t}$ corresponds to the outcome indicated in the title of each panel. The red (blue) points present the estimated coefficient for β_{group} (γ_{group}) for the specific decade, with vertical lines indicating the 95% confidence interval.

Figure 9: Firm Outcomes After Prominence: Pre-emergence and Pre-prominence Innovators (Levels)



Note: The figure displays the coefficient estimates from equation (4). The dependent variable $Y_{i,t}$ corresponds to the outcome indicated in the title of each panel. The red (blue) points present the estimated coefficient for β_{group} (γ_{group}) for the specific batch group shown by the x-axis, with vertical lines indicating the 95% confidence interval.

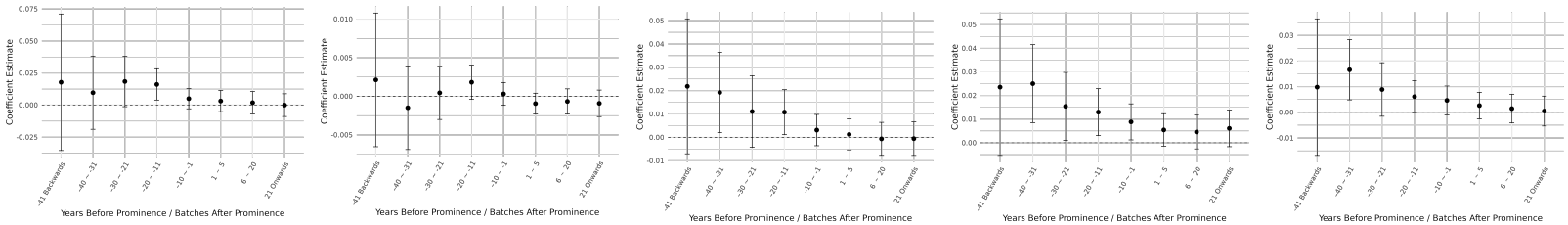
Figure 10: Firm Outcomes After Prominence: Pre-emergence and Pre-prominence Innovators (Changes)



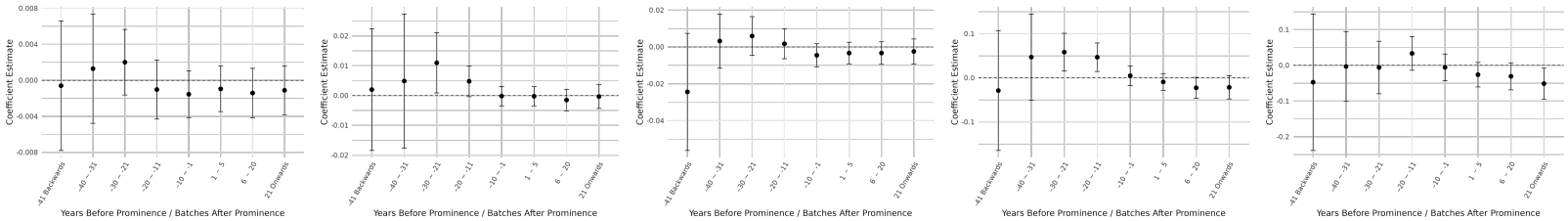
Note: The figure displays the coefficient estimates from equation (4). The dependent variable $Y_{i,t}$ corresponds to the outcome indicated in the title of each panel. The red (blue) points present the estimated coefficient for β_{group} (γ_{group}) for the specific batch group shown by the x-axis, with vertical lines indicating the 95% confidence interval.

Figure 11: Firm Outcomes of Active Innovators Across the Technology Life Cycle (Levels)

(a) $\frac{CapEx_t}{PPENT_{t-1}}$ (b) $\frac{R\&D_t}{Asset_{t-1}}$ (c) $\log\left(\frac{Asset_{t+1}}{Asset_t}\right)$ (d) $\log\left(\frac{Sales_{t+1}}{Sales_t}\right)$ (e) $\log\left(\frac{Empl_{t+1}}{Empl_t}\right)$



(f) ROA_t (g) $Margin_t$ (h) $ROIC_t$ (i) $\log(\text{Asset M/B}_t)$ (j) $\log(\text{P/E}_t)$

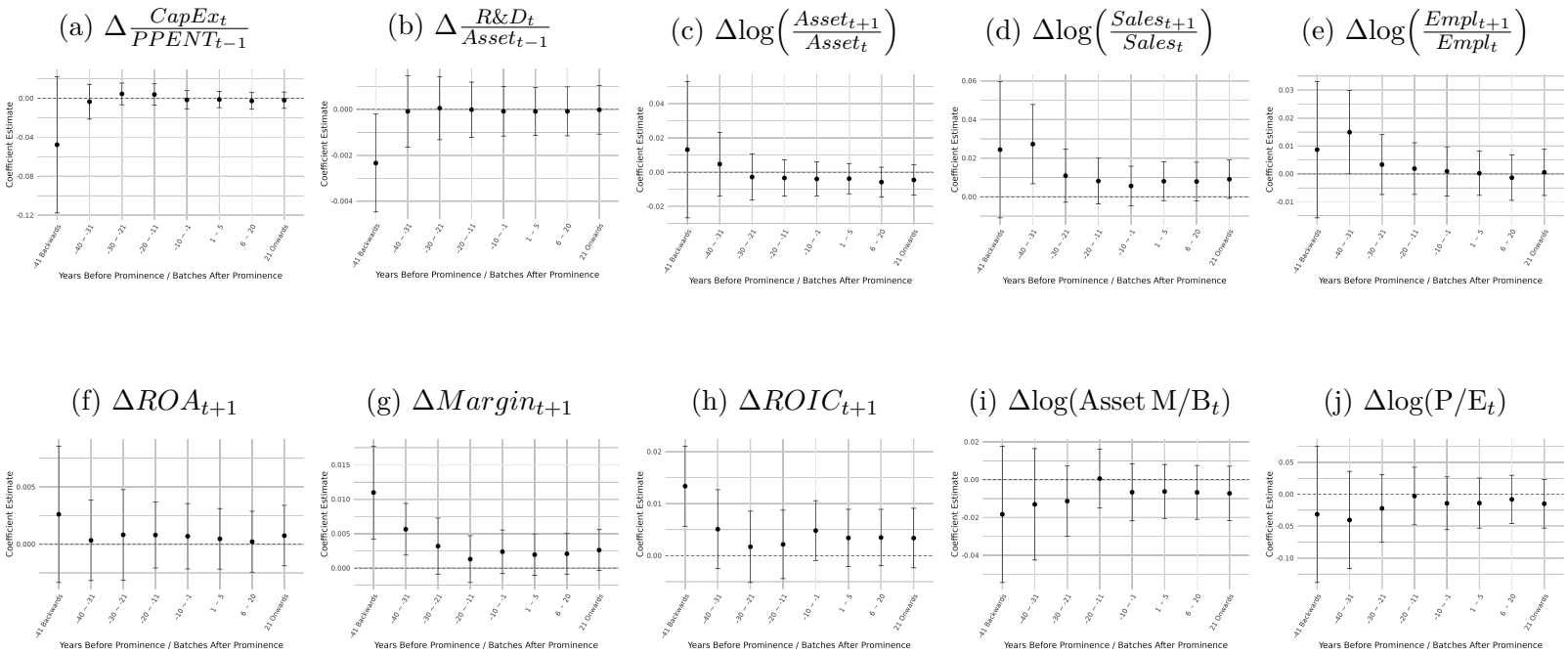


Note: This figure reports coefficient estimates from linear regression

$$
\begin{aligned}
Y_{i,t} = & \sum_{group \in \{1\text{–}10, 11\text{–}20, 21\text{–}30, 31\text{–}40, 41+\}} \beta_{group} \, \text{Assignee}_{\text{Active},i,j} \times Decade_{group,j,t} \\
& + \sum_{group \in \{1\text{–}5, 6\text{–}20, 21+\}} \beta_{group} \, \text{Assignee}_{\text{Active},i,j} \times Batch_{group,j,t} \\
& + \delta X_{i,t} + \alpha_i + \mu_{j,t} + \epsilon_{i,j,t},
\end{aligned}
\tag{6}
$$

where i, j, and t represent firm, technology and year, respectively. The dependent variable $Y_{i,t}$ corresponds to the outcome indicated in the title of each panel. The dummy variable $\text{Assignee}_{\text{Active},i,j}$ indicates whether firm i has received at least one patent in technology j and year t. $Decade_{group,j,t}$ is dummy variable indicating decades *before* prominence. $Batch_{group,j,t}$ is dummy variable indicating batch groups after prominence. We control for firm fixed effects (α_i), technology-by-year fixed effects ($\mu_{j,t}$) and firm-year level variables ($X_{i,t}$) including firm age and assets, along with their squared terms. We double-cluster standard errors at the firm and technology levels. The points present the estimated coefficient for β_{group} for the specific decade or batch group, with vertical lines indicating the 95% confidence interval.

Figure 12: Firm Outcomes of Active Innovators Across the Technology Life Cycle (Changes)



Note: This figure reports coefficient estimates from equation (6). The dependent variable $Y_{i,t}$ corresponds to the outcome indicated in the title of each panel. The points present the estimated coefficient for β_{group} for the specific decade or batch group, with vertical lines indicating the 95% confidence interval.

Table 1: Summary Statistics: Prominent Technologies

Statistic	Obs.	Mean	St. Dev.	Min	.25	.50	.75	Max
Panel A. Initiation, Emergence and Prominence								
Initiation (Year)	289	1,948.83	21.10	1,920	1,930	1,946	1,965	2,016
Emergence (Year)	289	1,964.01	21.09	1,920	1,949	1,963	1,978	2,017
Prominence (Year)	289	1,980.57	24.01	1,920	1,963	1,982	2,000	2,023
Initiation to Emergence	289	15.18	13.96	0	4	10	22	72
Initiation to Prominence	289	31.74	19.27	0	17	29	42	99
Emergence to Prominence	289	16.56	11.75	0	7	15	24	63
Panel B. Pre-emergence Patents and Assignees								
Patents (#)	289	20.10	10.44	6	12	17	26	61
Assignees (#)	289	12.64	8.28	1	7	11	17	50
Large innovators (#)	289	4.92	3.56	0	2	4	6	21
Governments (#)	289	0.18	0.48	0	0	0	0	3
Universities (#)	289	0.13	0.38	0	0	0	0	2
Panel C. Pre-prominence Patents and Assignees								
Patents (#)	289	482.82	550.44	19	137	292	577	3,731
Assignees (#)	289	178.49	162.29	1	65	145	251	1,146
Large innovators (#)	289	70.49	69.72	0	17	47	111	438
Governments (#)	289	2.37	3.69	0	0	1	4	27
Universities (#)	289	3.68	8.62	0	0	1	3	78
Panel D. Number of Batches								
Batches (#)	289	22.57	77.76	1	4	8	18	1,160

Note: This table reports summary statistics for the 289 prominent technologies in our sample. Panel A summarizes the initiation year (grant year of the first patent), the emergence year (first year in which the technology has at least five granted patents), and the prominence year (first year in which the technology ranks among the annual top 100 technologies by patent count). Panels B and C report counts of patents and assignees in the pre-emergence window (initiation through the emergence year, inclusive) and the pre-prominence window (initiation through the prominence year, inclusive), respectively, together with the number of government and university assignees. *Large innovators (window top 1%)* are assignees that rank in the top 1% of firm assignees by grant-year patent counts in *any* year during the corresponding window. Panel D reports the total number of patent batches observed for each technology through 2023 (including Batch_0). Detailed variable definitions are provided in Appendix A.

Table 2: Summary Statistics: Patent-level, Patent Batch-level and Assignee-batch Level

Statistic	Obs.	Mean	St. Dev.	Min	.25	.50	.75	Max
Panel A. Patent-level								
Citations	7,065,491	3.21	11.35	0	0	1	3	1,432
$\mathbb{1}(Citations > 0)$	7,065,491	0.65	0.48	0	0	1	1	1
$\mathbb{1}(Breakthrough)$	4,914,943	0.05	0.22	0	0	0	0	1
Originality	7,682,712	0.76	0.24	0.00	0.69	0.83	0.91	1.00
Panel B. Patent Batch-level								
Patents (#)	6,524	209.96	236.26	1	80	126	245	3,731
Grant-year span	6,524	4.15	8.42	1	1	2	3	100
Panel C. Assignee-batch Level								
PatentShare	1,169,166	0.006	0.020	0.000	0.000	0.000	0.004	1.000
$\text{Assignee}_{\text{PreEmergence}}$	1,169,166	0.014	0.119	0	0	0	0	1
$\text{Assignee}^{\text{Large}}_{\text{PreEmergence}}$	1,169,166	0.010	0.099	0	0	0	0	1
$\text{Assignee}^{\text{NonLarge}}_{\text{PreEmergence}}$	1,169,166	0.004	0.066	0	0	0	0	1

Note: This table reports summary statistics at three levels of observation. Panel A reports patent-level characteristics for patents granted from 1947 onward (the period for which forward-citation data are available). *Citations* is the count of forward citations a patent receives from its grant year through the end of the fifth subsequent calendar year; $\mathbb{1}(Citations > 0)$ indicates whether the patent receives at least one such citation. $\mathbb{1}(Breakthrough)$ is an indicator for breakthrough patents following Kelly et al. (2021) and is available for a subset of patents. *Originality* measures the diversity of technologies cited by the patent, following Hall et al. (2001) (constructed using title-based technologies rather than patent classifications), and is available for a subset of patents. Panel B reports technology–batch statistics, including the number of patents per batch and the inclusive *Grant-year span* of the batch, defined as $max(\text{grant year}) - min(\text{grant year}) + 1$. Panel C reports assignee–technology–batch statistics. *PatentShare* is an assignee's share of patents within a technology–batch. $\text{Assignee}_{\text{PreEmergence}}$ indicates assignees with at least one pre-emergence patent in the focal technology. $\text{Assignee}^{\text{Large}}_{\text{PreEmergence}}$ identifies large pre-emergence assignees (window top 1%), defined as pre-emergence innovators that rank in the top 1% of firm assignees by grant-year patent counts in at least one year between the technology's initiation and emergence; $\text{Assignee}^{\text{NonLarge}}_{\text{PreEmergence}}$ are pre-emergence assignees that never rank in the top 1% during that window. Detailed variable definitions are provided in Appendix A.

Table 3: Summary Statistics: Firm-Level Sample

Statistic	Obs.	Mean	St. Dev.	Min	.25	.50	.75	Max
Panel A. Basic Characteristics								
Age (Year)	1,843,679	23.34	16.70	1	11	20	32	73
Assets (mil. USD)	1,843,635	15,313.90	49,746.57	0.00	594.74	2,863.66	11,956.77	2,111,534.00
Sales (mil. USD)	1,843,489	10,277.22	20,554.30	−805.84	549.89	2,715.84	9,859.52	286,525.00
Employment (thous.)	1,648,408	60.24	108.49	0.00	5.10	22.68	72.34	2,300.00
Panel B. Investment								
$CapEx_t/PPENT_{t-1}$	1,843,679	0.29	0.29	0.01	0.14	0.21	0.32	5.01
$R\&D_t/Asset_{t-1}$	1,991,171	0.05	0.07	0.00	0.01	0.03	0.07	0.86
$\Delta CapEx_t/PPENT_{t-1}$	1,775,433	−0.01	0.24	−4.41	−0.05	−0.0004	0.05	3.37
$\Delta R\&D_t/Asset_{t-1}$	1,926,656	−0.001	0.03	−0.47	−0.003	0.00	0.003	0.23
Panel C. Growth								
$\log\left(\frac{Asset_{t+1}}{Asset_t}\right)$	1,967,522	0.05	0.21	−0.91	−0.05	0.03	0.12	2.17
$\log\left(\frac{Sales_{t+1}}{Sales_t}\right)$	1,963,743	0.04	0.21	−1.09	−0.05	0.03	0.12	1.92
$\log\left(\frac{Empl_{t+1}}{Empl_t}\right)$	1,651,468	0.03	0.16	−1.10	−0.03	0.02	0.08	1.29
$\Delta\log\left(\frac{Asset_{t+1}}{Asset_t}\right)$	1,904,166	−0.01	0.27	−2.33	−0.12	−0.005	0.11	1.33
$\Delta\log\left(\frac{Sales_{t+1}}{Sales_t}\right)$	1,899,704	−0.02	0.28	−2.18	−0.13	−0.01	0.11	1.42
$\Delta\log\left(\frac{Empl_{t+1}}{Empl_t}\right)$	1,548,164	−0.01	0.21	−1.44	−0.08	−0.003	0.06	1.22
Panel D. Performance								
ROA_t	1,840,609	0.05	0.08	−0.55	0.02	0.05	0.09	0.34
$Margin_t$	1,983,173	0.16	0.13	−0.42	0.08	0.14	0.22	0.63
$ROIC_t$	1,849,252	0.14	0.20	−0.78	0.05	0.10	0.19	1.23
ΔROA_{t+1}	1,810,832	−0.001	0.07	−0.44	−0.02	0.0005	0.02	0.31
$\Delta Margin_{t+1}$	1,957,129	−0.001	0.07	−0.57	−0.03	0.0001	0.02	0.39
$\Delta ROIC_{t+1}$	1,816,585	−0.005	0.15	−0.97	−0.04	0.0005	0.03	0.67
Panel E. Valuation								
$\log(\text{Asset } M/B_t)$	1,845,788	0.29	0.88	−1.90	−0.33	0.15	0.80	4.64
$\log(P/E_t)$	1,575,553	3.09	1.06	−0.76	2.44	3.08	3.68	8.38
$\Delta\log(\text{Asset } M/B_t)$	1,778,364	−0.01	0.41	−3.09	−0.20	0.01	0.21	2.31
$\Delta\log(P/E_t)$	1,402,052	−0.02	0.80	−4.41	−0.38	−0.003	0.34	5.69

Note: This table reports summary statistics for the firm-level sample used in the assignee–technology–year analyses. Each observation is an firm–technology–year with Compustat financial data for firms that are assigned at least one patent in a prominent technology (defined in Section 4.1). Panel A reports basic firm characteristics (age, assets, sales, and employment). Panel B reports investment measures (capital expenditures scaled by lagged tangible fixed assets and R&D scaled by lagged assets). Panel C reports one-year growth rates of assets, sales, and employment. Panel D reports profitability measures (ROA, operating margin and return on invested capital). Panel E reports valuation measures (log asset market-to-book and log price-to-earnings). For variables in Panels B–E, we report both levels and the corresponding one-year changes. Variables in Panels B, C, and E are winsorized at the 1st and 99th percentiles, whereas those in Panel D are winsorized at the 3rd and 97th percentiles. The statistics in Panel A are computed on the sample used in column (1) of Table 6, Panel A; the statistics in Panels B–E are computed on the estimation sample for the corresponding dependent variable in Table 6, Panel A (and therefore the reported observation counts can differ across panels). Detailed variable definitions are provided in Appendix A.

Table 4: Patent Quality of Prominent Technologies: Pre-emergence and Pre-prominence

	(1) log(1+Citations)	(2) $\mathbb{1}$ (Citations > 0)	(3) $\mathbb{1}$ (Breakthrough)	(4) Originality
Panel A.				
ProminentTech	0.015***	0.046***	0.008***	0.007***
	(0.002)	(0.005)	(0.002)	(0.001)
Class × Year	X	X	X	X
Assignee × Year	X	X	X	X
Obs	7,065,491	7,065,491	4,914,943	7,682,712
Adj. R^2	0.140	0.204	0.143	0.185
Panel B.				
ProminentTech × PreProminence × PreEmergence	0.040***	0.164***	0.045***	0.009**
	(0.008)	(0.023)	(0.008)	(0.004)
ProminentTech × PreProminence	0.045***	0.147***	0.030***	0.001
	(0.004)	(0.011)	(0.003)	(0.001)
ProminentTech	0.008***	0.025***	0.002*	0.007***
	(0.002)	(0.005)	(0.001)	(0.001)
Class × Year	X	X	X	X
Assignee × Year	X	X	X	X
Obs	7,065,491	7,065,491	4,914,943	7,682,712
Adj. R^2	0.140	0.204	0.143	0.185

Note: This table reports estimates of equation (1) relating patent outcomes to whether a patent is associated with a prominent technology. The unit of observation is the patent. *ProminentTech* equals one if the patent title contains at least one of the 289 prominent technologies in our sample. *PreEmergence* (*PreProminence*) equals one if the patent is granted on or before the emergence (prominence) year of the earliest prominent technology appearing in its title. *Interpretation:* in specifications that include *PreEmergence* and/or *PreProminence*, the coefficient on *ProminentTech* captures the baseline difference for prominent-technology patents *outside* the corresponding early window(s), while the interaction terms capture the incremental difference within the pre-emergence and pre-prominence windows. *Citations* is the count of forward citations received within five years of grant; $\mathbb{1}(Citations > 0)$ indicates whether the patent receives at least one such citation. $\mathbb{1}(Breakthrough)$ indicates breakthrough patents following Kelly et al. (2021). *Originality* follows Hall et al. (2001) and measures the diversity of cited technologies, constructed using title-based technologies rather than patent classifications. Regressions include 3-digit CPC-by-year fixed effects and assignee-by-year fixed effects. Standard errors are clustered by grant year. *** denotes p-value < 0.01, ** denotes p-value< 0.05, and * denotes p-value< 0.1.

Table 5: Patent Quality of Prominent Technologies: Large Innovators

	(1) log(1+Citations)	(2) $\mathbb{1}$(Citations > 0)	(3) $\mathbb{1}$(Breakthrough)	(4) Originality
Panel A.				
ProminentTech × Assignee$^{\text{Large}}$	-0.028***	-0.093***	-0.005***	-0.010***
	(0.002)	(0.005)	(0.002)	(0.001)
Assignee$^{\text{Large}}$	0.016***	0.024***	0.006***	-0.001
	(0.003)	(0.009)	(0.001)	(0.001)
ProminentTech	0.018***	0.061***	0.008***	0.001
	(0.005)	(0.010)	(0.001)	(0.002)
Class × Year	X	X	X	X
Obs	7,065,491	7,065,491	4,914,943	7,682,712
Adj. R^2	0.077	0.129	0.172	0.095
Panel B.				
ProminentTech × Assignee$^{\text{Large}}$ × PreProminence × PreEmergence	0.053***	0.203***	0.031**	0.007
	(0.019)	(0.039)	(0.012)	(0.009)
ProminentTech × Assignee$^{\text{Large}}$ × PreProminence	-0.005	-0.004	-0.007*	-0.001
	(0.004)	(0.014)	(0.004)	(0.002)
ProminentTech × Assignee$^{\text{Large}}$	-0.026***	-0.088***	-0.004	-0.010***
	(0.002)	(0.005)	(0.002)	(0.001)
Assignee$^{\text{Large}}$	0.016***	0.023***	0.006***	-0.001
	(0.003)	(0.009)	(0.001)	(0.001)
ProminentTech × PreProminence × PreEmergence	0.030**	0.110***	0.033***	0.011
	(0.012)	(0.029)	(0.010)	(0.007)
ProminentTech × PreProminence	0.067***	0.200***	0.043***	0.008***
	(0.007)	(0.019)	(0.005)	(0.002)
ProminentTech	0.008	0.029***	-0.001	0.000
	(0.005)	(0.010)	(0.001)	(0.002)
Class × Year	X	X	X	X
Obs	7,065,491	7,065,491	4,914,943	7,682,712
Adj. R^2	0.078	0.130	0.172	0.095

Note: This table reports estimates of equation (1) augmented with innovator-size indicators to examine whether the quality premium of prominent technologies varies with large-innovator status. The unit of observation is the patent. $ProminentTech$ equals one if the patent title contains at least one of the 289 prominent technologies in our sample. $PreEmergence$ ($PreProminence$) equals one if the patent is granted on or before the emergence (prominence) year of the earliest prominent technology appearing in its title. $Assignee^{\text{Large}}$ is an indicator equal to one if the patent's assignee ranks in the top 1% of firm assignees by grant-year patent counts (our definition of *large innovators*). The coefficient on $Assignee^{\text{Large}}$ captures the unconditional large-innovator premium; $ProminentTech \times Assignee^{\text{Large}}$ captures how this premium differs within prominent technologies; and interactions with $PreEmergence$ and $PreProminence$ isolate how these differences vary across life-cycle stages. *Citations* is the count of forward citations received within five years of grant; $\mathbb{1}(Citations > 0)$ indicates whether the patent receives at least one such citation. $\mathbb{1}(Breakthrough)$ indicates breakthrough patents following Kelly et al. (2021). *Originality* follows Hall et al. (2001) and measures the diversity of cited technologies, constructed using title-based technologies rather than patent classifications. Regressions include 3-digit CPC-by-year fixed effects. Standard errors are clustered by grant year. *** denotes p-value < 0.01, ** denotes p-value< 0.05, and * denotes p-value< 0.1.

Table 6: Outcomes of Pre-emergence and Pre-prominence Innovators Before Prominence

Panel A.	(1) $\frac{CapEx_t}{PPENT_{t-1}}$	(2) $\frac{R\&DExp_t}{Asset_{t-1}}$	(3) $log(\frac{Asset_{t+1}}{Asset_t})$	(4) $log(\frac{Sales_{t+1}}{Sales_t})$	(5) $log(\frac{Empl_{t+1}}{Empl_t})$	(6) ROA_t	(7) $Margin_t$	(8) $ROIC_t$	(9) $log(\text{Asset M/B}_t)$	(10) $log(P/E_t)$
$\text{Assignee}_{\text{PreEmergence}} \times \text{PreProminence} \times \text{PreEmergence}$	0.029***	0.002	0.023***	0.022***	0.016***	0.002	0.013	0.002	0.100***	0.077*
	(0.010)	(0.002)	(0.006)	(0.005)	(0.004)	(0.002)	(0.009)	(0.005)	(0.037)	(0.041)
$\text{Assignee}_{\text{PreEmergence}} \times \text{PreProminence}$	0.015***	0.006***	0.023***	0.018***	0.009***	0.003*	0.008*	0.010**	0.049**	0.002
	(0.005)	(0.001)	(0.004)	(0.004)	(0.003)	(0.002)	(0.005)	(0.004)	(0.024)	(0.022)
$\text{Assignee}_{\text{PreEmergence}}$	−0.011***	−0.002***	−0.013***	−0.011***	−0.006***	−0.001**	−0.006*	−0.004**	−0.035***	−0.017
	(0.003)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.003)	(0.002)	(0.010)	(0.011)
$\text{Assignee}_{\text{PreProminence}} \times \text{PreProminence} \times \text{PreEmergence}$	0.017***	−0.001	0.014***	0.015***	0.012***	0.002	0.007**	−0.001	0.070***	0.050**
	(0.005)	(0.001)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.022)	(0.020)
$\text{Assignee}_{\text{PreProminence}} \times \text{PreProminence}$	0.015***	0.006***	0.020***	0.016***	0.010***	0.002**	0.004*	0.008***	0.054***	0.016
	(0.003)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)	(0.013)	(0.014)
$\text{Assignee}_{\text{PreProminence}}$	−0.008***	−0.002***	−0.009***	−0.008***	−0.005***	−0.001**	−0.002**	−0.002***	−0.024***	−0.011*
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.006)	(0.006)
Log Asset	−0.165***	−0.032***	−0.097***	−0.083***	−0.064***	−0.003	0.006	−0.005	−0.108***	0.046
	(0.013)	(0.003)	(0.006)	(0.006)	(0.006)	(0.003)	(0.005)	(0.009)	(0.035)	(0.039)
Log Asset^2	0.007***	0.001***	0.002***	0.002***	0.002***	0.000**	0.000	−0.002***	0.001	−0.002
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.003)	(0.003)
Log Age	−0.382***	−0.050***	−0.051**	−0.061***	−0.089***	−0.028***	−0.033**	−0.077***	−0.255**	0.327***
	(0.038)	(0.006)	(0.022)	(0.019)	(0.019)	(0.010)	(0.014)	(0.024)	(0.115)	(0.115)
Log Age^2	0.084***	0.020***	0.016**	0.016***	0.020***	0.007**	0.004	0.011	−0.036	−0.218***
	(0.010)	(0.002)	(0.007)	(0.006)	(0.006)	(0.003)	(0.005)	(0.008)	(0.041)	(0.041)
Assignee	X	X	X	X	X	X	X	X	X	X
Technology × Year	X	X	X	X	X	X	X	X	X	X
Obs	1,843,679	1,991,171	1,967,522	1,963,743	1,651,468	1,840,609	1,983,173	1,849,252	1,845,788	1,575,553
Adj. R^2	0.404	0.708	0.283	0.266	0.225	0.426	0.602	0.409	0.704	0.542

Panel B.	(1) $\Delta\frac{CapEx_t}{PPENT_{t-1}}$	(2) $\Delta\frac{R\&DExp_t}{Asset_{t-1}}$	(3) $\Delta log(\frac{Asset_{t+1}}{Asset_t})$	(4) $\Delta log(\frac{Sales_{t+1}}{Sales_t})$	(5) $\Delta log(\frac{Empl_{t+1}}{Empl_t})$	(6) ΔROA_{t+1}	(7) $\Delta Margin_{t+1}$	(8) $\Delta ROIC_{t+1}$	(9) $\Delta log(\text{Asset M/B}_t)$	(10) $\Delta log(P/E_t)$
$\text{Assignee}_{\text{PreEmergence}} \times \text{PreProminence} \times \text{PreEmergence}$	−0.001	0.000	0.005	0.003	0.003	0.000	0.001	0.000	0.007	0.006
	(0.003)	(0.000)	(0.004)	(0.004)	(0.003)	(0.000)	(0.001)	(0.001)	(0.006)	(0.010)
$\text{Assignee}_{\text{PreEmergence}} \times \text{PreProminence}$	0.011***	0.001***	0.024***	0.018***	0.016***	0.001***	0.002***	0.000	0.035***	0.015*
	(0.002)	(0.000)	(0.004)	(0.003)	(0.002)	(0.000)	(0.000)	(0.001)	(0.005)	(0.008)
$\text{Assignee}_{\text{PreEmergence}}$	−0.004***	0.000***	−0.009***	−0.007***	−0.006***	0.000**	−0.001***	0.000	−0.013***	−0.006*
	(0.001)	(0.000)	(0.002)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.002)	(0.003)
$\text{Assignee}_{\text{PreProminence}} \times \text{PreProminence} \times \text{PreEmergence}$	−0.005**	−0.001***	−0.009***	−0.006**	−0.004*	0.000	0.001	0.001	−0.009*	−0.004
	(0.002)	(0.000)	(0.003)	(0.002)	(0.002)	(0.000)	(0.000)	(0.001)	(0.005)	(0.006)
$\text{Assignee}_{\text{PreProminence}} \times \text{PreProminence}$	0.009***	0.001***	0.022***	0.014***	0.013***	0.001*	0.001***	0.000	0.028***	0.018***
	(0.002)	(0.000)	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)	(0.001)	(0.003)	(0.004)
$\text{Assignee}_{\text{PreProminence}}$	−0.002***	0.000***	−0.007***	−0.004***	−0.004***	0.000*	0.000***	0.000	−0.008***	−0.006***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.002)
Log Asset	−0.033***	−0.003**	−0.078***	−0.041***	−0.044***	−0.002	−0.006***	−0.002	−0.139***	−0.031*
	(0.011)	(0.001)	(0.008)	(0.006)	(0.006)	(0.001)	(0.001)	(0.003)	(0.012)	(0.018)
Log Asset^2	0.001**	0.000	0.001**	0.000	0.000	0.000	0.000	0.000	0.003***	0.000
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Log Age	0.247***	0.038***	0.268***	0.379***	0.144***	0.007	0.003	−0.002	0.540***	0.543***
	(0.051)	(0.004)	(0.032)	(0.029)	(0.028)	(0.005)	(0.004)	(0.010)	(0.049)	(0.098)
Log Age^2	−0.047***	−0.009***	−0.057***	−0.086***	−0.022***	−0.001	0.000	0.004	−0.109***	−0.131***
	(0.013)	(0.001)	(0.008)	(0.008)	(0.008)	(0.001)	(0.001)	(0.003)	(0.013)	(0.026)
Assignee	X	X	X	X	X	X	X	X	X	X
Technology × Year	X	X	X	X	X	X	X	X	X	X
Obs	1,775,433	1,926,656	1,904,166	1,899,704	1,548,164	1,810,832	1,957,129	1,816,585	1,778,364	1,402,052
Adj. R^2	0.059	0.071	0.148	0.141	0.075	0.078	0.104	0.060	0.182	0.159

Note: This table reports estimates from equation (2). The unit of observation is the firm–technology–year. All dependent variables are defined in Appendix A. The key explanatory variables identify early innovators: $Assignee_{\text{PreEmergence}}$ is an indicator equal to one if the firm's first patent in the technology is granted on or before the technology's emergence year; $Assignee_{\text{PreProminence}}$ is an indicator equal to one if the firm's first patent in the technology is granted after emergence but on or before the technology's prominence year. These variables are interacted with indicators *PreEmergence* and *PreProminence*, which equal one if the observation year is on or before the technology's emergence and prominence, respectively. We control for firm size and age, and include firm fixed effects and technology-by-year fixed effects. Standard errors are double-clustered at the technology and firm levels. *** denotes p-value < 0.01, ** denotes p-value< 0.05, and * denotes p-value< 0.1.

Table 7: Patent Quality Across Post-prominence Patent Batch Groups

	(1) log(1+Citations)	(2) $\mathbb{1}$ (Citations >0)	(3) $\mathbb{1}$ (Breakthrough)	(4) Originality
Batch_{1-5}	-0.024***	-0.120***	-0.038***	-0.004***
	(0.003)	(0.008)	(0.005)	(0.001)
Batch_{6-20}	-0.035***	-0.164***	-0.053***	-0.005***
	(0.005)	(0.010)	(0.007)	(0.002)
Batch_{21+}	-0.052***	-0.209***	-0.080***	-0.010***
	(0.007)	(0.015)	(0.009)	(0.003)
Technology	X	X	X	X
Class × Year	X	X	X	X
Assignee × Year	X	X	X	X
Obs	923,524	923,524	512,260	1,034,792
Adj. R^2	0.168	0.276	0.287	0.137

Note: This table presents the results from estimating equation

$$Y_k = \sum_{group \in \{1\text{–}5, 6\text{–}20, 21+\}} \beta \, Batch_{group,j,k} + \mu_j + \delta_{c,t} + \alpha_{i,t} + \epsilon_k,$$

where j and k represent technology and patent, respectively. $Batch_{group,j,k}$ indicates the batch group of technology j to which the patent k is classified. *Citations* is the count of forward citations received within five years of grant; $\mathbb{1}(Citations > 0)$ indicates whether the patent receives at least one such citation. $\mathbb{1}(Breakthrough)$ indicates breakthrough patents following Kelly et al. (2021). *Originality* follows Hall et al. (2001) and measures the diversity of cited technologies, constructed using title-based technologies rather than patent classifications. The omitted category is $Batch_0$, which contains all patents granted up to and including the technology's prominence year. We include technology fixed effects (μ_j), patent classification (3-digit CPC)-year fixed effects ($\delta_{c,t}$) and assignee-year fixed effects ($\alpha_{i,t}$). Standard errors are clustered by the patent grant year. *** denotes p-value < 0.01, ** denotes p-value< 0.05, and * denotes p-value< 0.1.

Table 8: Pre-emergence Assignees' Patent Shares Across the Technology Life Cycle

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A.						
$Assignee_{PreEmergence}$	0.014***	0.016***	0.011***	0.013***	0.011***	0.013***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$Assignee_{PreEmergence} \times Batch$		-0.003***		-0.003***		-0.003***
		(0.000)		(0.000)		(0.001)
Technology × Batch	X	X	X	X	X	X
Assignee			X	X		
Assignee × Year					X	X
Obs	1,169,156	1,169,156	1,123,417	1,123,417	1,032,865	1,032,865
Adj. R^2	0.113	0.114	0.251	0.251	0.277	0.278
Panel B.						
$Assignee^{Large}_{PreEmergence}$	0.019***	0.021***	0.012***	0.014***	0.011***	0.013***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
$Assignee^{Large}_{PreEmergence} \times Batch$		-0.003***		-0.003***		-0.002***
		(0.000)		(0.000)		(0.000)
$Assignee^{Non\text{-}Large}_{PreEmergence}$	0.004**	0.005**	0.008***	0.010***	0.009**	0.012***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)
$Assignee^{Non\text{-}Large}_{PreEmergence} \times Batch$		-0.003***		-0.004***		-0.005***
		(0.001)		(0.001)		(0.002)
Technology × Batch	X	X	X	X	X	X
Assignee			X	X		
Assignee × Year					X	X
Obs	1,169,156	1,169,156	1,123,417	1,123,417	1,032,865	1,032,865
Adj. R^2	0.115	0.116	0.251	0.251	0.277	0.278

Note: The table presents the result from estimating (3). Each observation is at assignee-technology-batch level. The dependent variable *PatentShare*, is the fraction of patents in the batch that are assigned to a given assignee. $Assignee_{PreEmergence}$ indicates pre-emergence assignees, defined as assignees with at least one pre-emergence patent in the technology. $Assignee^{Large}_{PreEmergence}$ indicates assignees in the top 1% of the patent-count distribution in any year of the pre-emergence period; $Assignee^{NonLarge}_{PreEmergence}$ indicates pre-emergence assignees outside this top 1%. *Batch* is the technology batch to which the patent is classified, scaled by 100. We include technology-batch fixed effects and assignee–batch start-year fixed effects. Standard errors are double-clustered at the technology and firm levels. *** denotes p-value < 0.01, ** denotes p-value< 0.05, and * denotes p-value< 0.1.

Table 9: Pre-emergence Assignees' Patent Shares Across the Technology Life Cycle: Scope and Early Competition

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A.						
$\text{Assignee}^{\text{Large}}_{\text{PreEmergence}} \times \text{Log}(\#\text{Tech}_{\text{PreEmergence}})$	-0.004*	-0.005**	-0.005***	-0.006***	-0.004**	-0.005***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
$\text{Assignee}^{\text{Large}}_{\text{PreEmergence}} \times \text{Batch} \times \text{Log}(\#\text{Tech}_{\text{PreEmergence}})$		0.014***		0.012***		0.009**
		(0.005)		(0.003)		(0.003)
$\text{Assignee}^{\text{Large}}_{\text{PreEmergence}}$	0.027***	0.033***	0.027***	0.032***	0.021***	0.026***
	(0.007)	(0.007)	(0.005)	(0.005)	(0.005)	(0.006)
$\text{Assignee}^{\text{Large}}_{\text{PreEmergence}} \times \text{Batch}$		-0.050***		-0.041***		-0.031**
		(0.018)		(0.011)		(0.012)
$\text{Assignee}^{\text{NonLarge}}_{\text{PreEmergence}} \times \text{Log}(\#\text{Tech}_{\text{PreEmergence}})$	-0.011***	-0.011***	-0.014***	-0.016***	-0.019***	-0.028***
	(0.003)	(0.004)	(0.005)	(0.006)	(0.007)	(0.009)
$\text{Assignee}^{\text{NonLarge}}_{\text{PreEmergence}} \times \text{Batch} \times \text{Log}(\#\text{Tech}_{\text{PreEmergence}})$		0.006		0.021***		0.039***
		(0.004)		(0.007)		(0.014)
$\text{Assignee}^{\text{NonLarge}}_{\text{PreEmergence}}$	0.014***	0.015***	0.023***	0.028***	0.033***	0.048***
	(0.005)	(0.005)	(0.008)	(0.009)	(0.011)	(0.015)
$\text{Assignee}^{\text{NonLarge}}_{\text{PreEmergence}} \times \text{Batch}$		-0.010		-0.036***		-0.070***
		(0.007)		(0.012)		(0.025)
Technology × Batch	X	X	X	X	X	X
Assignee			X	X		
Assignee × Year					X	X
Obs	1,169,156	1,169,156	1,123,417	1,123,417	1,032,865	1,032,865
Adj. R^2	0.116	0.118	0.251	0.253	0.278	0.279
Panel B.						
$\text{Assignee}^{\text{Large}}_{\text{PreEmergence}} \times \mathbb{1}\left(\#\text{Assignee}^{\text{Large}}_{\text{PreEmergence}} > 5\right)$	-0.009*	-0.011***	-0.007*	-0.008**	-0.006	-0.007
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$\text{Assignee}^{\text{Large}}_{\text{PreEmergence}} \times \text{Batch} \times \mathbb{1}\left(\#\text{Assignee}^{\text{Large}}_{\text{PreEmergence}} > 5\right)$		-0.011***		-0.011***		-0.008***
		(0.004)		(0.002)		(0.002)
$\text{Assignee}^{\text{Large}}_{\text{PreEmergence}}$	0.023***	0.028***	0.016***	0.020***	0.015***	0.018***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
$\text{Assignee}^{\text{Large}}_{\text{PreEmergence}} \times \text{Batch}$		-0.004***		-0.003***		-0.003***
		(0.000)		(0.000)		(0.000)
$\text{Assignee}^{\text{NonLarge}}_{\text{PreEmergence}} \times \mathbb{1}\left(\#\text{Assignee}^{\text{Large}}_{\text{PreEmergence}} > 5\right)$	-0.005	-0.007*	-0.006	-0.008	-0.006	-0.009
	(0.004)	(0.004)	(0.005)	(0.005)	(0.007)	(0.008)
$\text{Assignee}^{\text{NonLarge}}_{\text{PreEmergence}} \times \text{Batch} \times \mathbb{1}\left(\#\text{Assignee}^{\text{Large}}_{\text{PreEmergence}} > 5\right)$		0.004		-0.002		-0.006
		(0.003)		(0.004)		(0.005)
$\text{Assignee}^{\text{NonLarge}}_{\text{PreEmergence}}$	0.006*	0.008**	0.010**	0.013***	0.011*	0.016**
	(0.003)	(0.003)	(0.005)	(0.005)	(0.006)	(0.007)
$\text{Assignee}^{\text{NonLarge}}_{\text{PreEmergence}} \times \text{Batch}$		-0.004***		-0.005**		-0.006**
		(0.001)		(0.002)		(0.002)
Technology × Batch	X	X	X	X	X	X
Assignee			X	X		
Assignee × Year					X	X
Obs	1,169,156	1,169,156	1,123,417	1,123,417	1,032,865	1,032,865
Adj. R^2	0.115	0.117	0.251	0.252	0.277	0.278

Note: This table reports estimates from equation (3). The unit of observation is the assignee-technology-batch level. The dependent variable, $PatentShare$, is the fraction of patents in a specific technology batch assigned to a given assignee. The key explanatory variables classify assignees based on their activity during the technology's pre-emergence period. $Assignee^{\text{Large}}_{\text{PreEmergence}}$ indicates assignees in the top 1% of the patent-count distribution in any year of the pre-emergence period; $Assignee^{\text{NonLarge}}_{\text{PreEmergence}}$ indicates pre-emergence assignees outside this top 1%. Panel A interacts these variables with $Log(\#Tech_{\text{PreEmergence}})$, the natural logarithm of the number of technologies where the firm was already a pre-emergence assignee at the time of the focal technology's emergence. Panel B interacts them with $\mathbb{1}(\#Assignee^{\text{Large}}_{\text{PreEmergence}} > 5)$, an indicator equal to one for technologies with more than five large pre-emergence assignees. $Batch$ denotes the technology batch, scaled by 100. We include technology-batch fixed effects and assignee-by-batch-start-year fixed effects. Standard errors are double-clustered at the technology and firm levels. *** denotes p-value < 0.01, ** denotes p-value< 0.05, and * denotes p-value< 0.1.

Table 10: Outcomes of Pre-emergence and Pre-prominence Innovators After Prominence

Panel A.	(1) $\frac{CapEx_t}{PPENT_{t-1}}$	(2) $\frac{R\&DExp_t}{Asset_{t-1}}$	(3) $log(\frac{Asset_{t+1}}{Asset_t})$	(4) $log(\frac{Sales_{t+1}}{Sales_t})$	(5) $log(\frac{Empl_{t+1}}{Empl_t})$	(6) ROA_t	(7) $Margin_t$	(8) $ROIC_t$	(9) log(Asset M/B$_t$)	(10) $log(P/E_t)$
$\text{Assignee}_{\text{PreEmergence}} \times \text{Batch}_{1-5}$	−0.019***	−0.006***	−0.027***	−0.022***	−0.013***	−0.004**	−0.009	−0.009**	−0.064***	−0.024
	(0.006)	(0.001)	(0.004)	(0.004)	(0.003)	(0.002)	(0.006)	(0.004)	(0.023)	(0.026)
$\text{Assignee}_{\text{PreEmergence}} \times \text{Batch}_{6-20}$	−0.031***	−0.007***	−0.036***	−0.028***	−0.016***	−0.004*	−0.016**	−0.013**	−0.081**	−0.015
	(0.007)	(0.002)	(0.005)	(0.005)	(0.003)	(0.002)	(0.008)	(0.005)	(0.032)	(0.034)
$\text{Assignee}_{\text{PreEmergence}} \times \text{Batch}_{21+}$	−0.034***	−0.008***	−0.039***	−0.037***	−0.021***	−0.006*	−0.020**	−0.014**	−0.115***	−0.034
	(0.010)	(0.002)	(0.008)	(0.008)	(0.005)	(0.003)	(0.010)	(0.007)	(0.042)	(0.039)
$\text{Assignee}_{\text{PreEmergence}}$	0.013***	0.004***	0.018***	0.014***	0.008***	0.002**	0.006	0.007**	0.040**	0.005
	(0.004)	(0.001)	(0.003)	(0.003)	(0.002)	(0.001)	(0.004)	(0.003)	(0.016)	(0.017)
$\text{Assignee}_{\text{PreProminence}} \times \text{Batch}_{1-5}$	−0.018***	−0.005***	−0.020***	−0.017***	−0.012***	−0.003***	−0.005**	−0.007***	−0.062***	−0.017
	(0.003)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.013)	(0.014)
$\text{Assignee}_{\text{PreProminence}} \times \text{Batch}_{6-20}$	−0.020***	−0.007***	−0.026***	−0.022***	−0.013***	−0.002	−0.005	−0.008**	−0.070***	−0.036*
	(0.004)	(0.001)	(0.003)	(0.003)	(0.002)	(0.001)	(0.003)	(0.004)	(0.020)	(0.022)
$\text{Assignee}_{\text{PreProminence}} \times \text{Batch}_{21+}$	−0.027***	−0.008***	−0.034***	−0.028***	−0.016***	−0.003	−0.006	−0.010*	−0.091***	−0.072**
	(0.005)	(0.002)	(0.005)	(0.004)	(0.003)	(0.002)	(0.004)	(0.006)	(0.031)	(0.036)
$\text{Assignee}_{\text{PreProminence}}$	0.012***	0.004***	0.014***	0.012***	0.008***	0.002**	0.003**	0.005***	0.042***	0.015
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.010)	(0.010)
Log Asset	−0.165***	−0.032***	−0.097***	−0.083***	−0.064***	−0.003	0.006	−0.005	−0.108***	0.046
	(0.013)	(0.003)	(0.006)	(0.006)	(0.006)	(0.003)	(0.005)	(0.009)	(0.035)	(0.039)
Log Asset^2	0.007***	0.001***	0.002***	0.002***	0.002***	0.000**	0.000	−0.002***	0.001	−0.002
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.003)	(0.003)
Log Age	−0.381***	−0.050***	−0.050**	−0.061***	−0.088***	−0.028***	−0.033**	−0.077***	−0.253**	0.329***
	(0.038)	(0.006)	(0.022)	(0.019)	(0.019)	(0.010)	(0.014)	(0.024)	(0.115)	(0.115)
Log Age^2	0.084***	0.020***	0.016**	0.016***	0.020***	0.007**	0.003	0.011	−0.037	−0.219***
	(0.010)	(0.002)	(0.007)	(0.006)	(0.006)	(0.003)	(0.005)	(0.008)	(0.041)	(0.041)
Assignee	X	X	X	X	X	X	X	X	X	X
Technology × Year	X	X	X	X	X	X	X	X	X	X
Obs	1,843,679	1,991,171	1,967,522	1,963,743	1,651,468	1,840,609	1,983,173	1,849,252	1,845,788	1,575,553
Adj. R^2	0.404	0.708	0.283	0.266	0.225	0.426	0.602	0.409	0.704	0.542

Panel B.	(1) $\Delta\frac{CapEx_t}{PPENT_{t-1}}$	(2) $\Delta\frac{R\&DExp_t}{Asset_{t-1}}$	(3) $\Delta log(\frac{Asset_{t+1}}{Asset_t})$	(4) $\Delta log(\frac{Sales_{t+1}}{Sales_t})$	(5) $\Delta log(\frac{Empl_{t+1}}{Empl_t})$	(6) ΔROA_{t+1}	(7) $\Delta Margin_{t+1}$	(8) $\Delta ROIC_{t+1}$	(9) $\Delta log(\text{Asset M/B}_t)$	(10) $\Delta log(P/E_t)$
$\text{Assignee}_{\text{PreEmergence}} \times \text{Batch}_{1-5}$	−0.010***	−0.001***	−0.025***	−0.017***	−0.014***	−0.001**	−0.002***	−0.001	−0.033***	−0.013**
	(0.002)	(0.000)	(0.004)	(0.003)	(0.002)	(0.000)	(0.000)	(0.001)	(0.005)	(0.006)
$\text{Assignee}_{\text{PreEmergence}} \times \text{Batch}_{6-20}$	−0.011***	−0.001***	−0.028***	−0.021***	−0.020***	−0.001	−0.003***	0.001	−0.040***	−0.024**
	(0.002)	(0.000)	(0.004)	(0.003)	(0.003)	(0.000)	(0.001)	(0.001)	(0.007)	(0.011)
$\text{Assignee}_{\text{PreEmergence}} \times \text{Batch}_{21+}$	−0.013***	−0.001***	−0.037***	−0.028***	−0.023***	−0.002**	−0.003***	0.002	−0.056***	−0.019
	(0.003)	(0.000)	(0.007)	(0.005)	(0.004)	(0.001)	(0.001)	(0.002)	(0.010)	(0.018)
$\text{Assignee}_{\text{PreEmergence}}$	0.007***	0.001***	0.018***	0.013***	0.011***	0.001**	0.001***	0.000	0.025***	0.011**
	(0.001)	(0.000)	(0.003)	(0.002)	(0.002)	(0.000)	(0.000)	(0.001)	(0.004)	(0.005)
$\text{Assignee}_{\text{PreProminence}} \times \text{Batch}_{1-5}$	−0.006***	−0.001***	−0.016***	−0.010***	−0.010***	0.000*	−0.001***	0.000	−0.021***	−0.013***
	(0.002)	(0.000)	(0.002)	(0.002)	(0.001)	(0.000)	(0.000)	(0.001)	(0.003)	(0.005)
$\text{Assignee}_{\text{PreProminence}} \times \text{Batch}_{6-20}$	−0.011***	−0.001***	−0.027***	−0.018***	−0.016***	−0.001*	−0.002***	−0.001	−0.035***	−0.024***
	(0.002)	(0.000)	(0.003)	(0.002)	(0.002)	(0.000)	(0.000)	(0.001)	(0.004)	(0.007)
$\text{Assignee}_{\text{PreProminence}} \times \text{Batch}_{21+}$	−0.010***	−0.001***	−0.037***	−0.026***	−0.021***	−0.001*	−0.003***	−0.003	−0.049***	−0.037***
	(0.002)	(0.000)	(0.004)	(0.003)	(0.003)	(0.000)	(0.001)	(0.002)	(0.007)	(0.010)
$\text{Assignee}_{\text{PreProminence}}$	0.005***	0.001***	0.013***	0.009***	0.008***	0.000**	0.001***	0.000	0.018***	0.011***
	(0.001)	(0.000)	(0.002)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.002)	(0.003)
Log Asset	−0.033***	−0.003**	−0.078***	−0.041***	−0.044***	−0.002	−0.006***	−0.002	−0.139***	−0.031*
	(0.011)	(0.001)	(0.008)	(0.006)	(0.006)	(0.001)	(0.001)	(0.003)	(0.012)	(0.018)
Log Asset^2	0.001**	0.000	0.001**	0.000	0.000	0.000	0.000	0.000	0.003***	0.000
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Log Age	0.247***	0.038***	0.269***	0.380***	0.145***	0.007	0.003	−0.002	0.542***	0.544***
	(0.051)	(0.004)	(0.032)	(0.029)	(0.028)	(0.005)	(0.004)	(0.010)	(0.049)	(0.098)
Log Age^2	−0.047***	−0.009***	−0.057***	−0.086***	−0.022***	−0.001	0.000	0.004	−0.110***	−0.131***
	(0.013)	(0.001)	(0.008)	(0.008)	(0.007)	(0.001)	(0.001)	(0.003)	(0.013)	(0.026)
Assignee	X	X	X	X	X	X	X	X	X	X
Technology × Year	X	X	X	X	X	X	X	X	X	X
Obs	1,775,433	1,926,656	1,904,166	1,899,704	1,548,164	1,810,832	1,957,129	1,816,585	1,778,364	1,402,052
Adj. R^2	0.059	0.071	0.148	0.141	0.075	0.078	0.104	0.060	0.182	0.159

Note: This table reports estimates from equation (4). The unit of observation is the firm–technology–year. All dependent variables are defined in Appendix A. The key explanatory variables identify early entrants: $Assignee_{\text{PreEmergence}}$ is an indicator equal to one if the firm's first patent in the technology is granted on or before the technology's emergence year; $Assignee_{\text{PreProminence}}$ is an indicator equal to one if the firm's first patent in the technology is granted after emergence but on or before the technology's prominence year. These variables are interacted with lifecycle indicators $Batch_{1-5}$, $Batch_{6-20}$, and $Batch_{21+}$, which equal one if the observation year corresponds to the respective batch groups of the technology. We control for firm size and age, and include firm fixed effects and technology-by-year fixed effects. Standard errors are double-clustered at the technology and firm levels. *** denotes p-value < 0.01, ** denotes p-value< 0.05, and * denotes p-value< 0.1.

Internet Appendix to

Technology Life Cycles: Investment, Growth, and Firm Value

March 4, 2026

A Variable Definitions

Technology-level Variables	
Initiation (Year)	The grant year of the technology's first patent.
Emergence (Year)	The first year that the technology has at least five patents granted.
Prominence (Year)	The first year that the technology ranks among top 100 technologies by patent count in that year.

Patent-level Variables	
Citations	Count of forward citations received by a patent from the grant year through the end of the fifth subsequent calendar year.
$\mathbb{1}(Citations > 0)$	A dummy variable that equals one if $Citations > 0$.
$\mathbb{1}$(Breakthrough)	A dummy variable that equals one if the patent is defined as a breakthrough patent by Kelly et al. (2021).
Originality	A patent originality measure capturing the diversity of technologies cited by the patent (Hall et al., 2001). We construct the measure using n-gram technologies extracted from patent titles (described in Section 2.2 and Appendix C) rather than patent classifications as in Hall et al. (2001).
ProminentTech	A dummy variable indicating whether the patent contains any of the 289 prominent technologies (defined in Section 4.1) in our sample.
PreEmergence	A dummy variable that equals one if the patent is granted on/before the earliest emergence year among prominent technologies in its title, where emergence year is the first year that the technology has at least five patents granted.

PreProminence	A dummy variable that equals one if a patent is granted on/before the earliest prominence year of any prominent technology in its title, where prominence year is the first year that the technology ranks among top 100 technologies by patent count in that year.
$Assignee^{\text{Large}}$	A dummy variable that equals one if the assignee of the patent ranks in the top 1% by patent count in the patent grant year.

Assignee-technology-level Variables

PatentShare	The assignee's share of patents within a technology–batch (number of patents assigned to the assignee divided by the total number of patents in the technology–batch).
$Assignee_{\text{PreEmergence}}$	A dummy variable indicating assignees with at least one patent granted on/before the year of emergence of the focal technology, where year of emergence is the first year that the technology has at least five patents granted.
$Assignee^{\text{Large}}_{\text{PreEmergence}}$	A dummy variable indicating assignees that (1) has at least one patent granted on/before the emergence of the focal technology, and (2) ranks in the top 1% of firms by patent count in at least one year between the technology's initiation and emergence.
$Assignee^{\text{NonLarge}}_{\text{PreEmergence}}$	A dummy variable indicating assignees that (1) has at least one patent granted on/before the emergence of the focal technology, and (2) never ranks in the top 1% of firms by patent count in any year between the technology's initiation and emergence.
$Assignee_{\text{PreProminence}}$	A dummy variable indicating assignees whose first patent in the focal technology is granted after the technology's year of emergence but on/before its year of prominence.

$\log(\#Tech_{\text{PreEmergence}})$	The natural logarithm of the number of technologies for which the focal assignee was already a pre-emergence assignee ($Assignee_{\text{PreEmergence}}$) at the emergence year of the focal technology.
$\mathbb{1}\left(\#\text{Assignee}^{\text{Large}}_{\text{PreEmergence}} > 5\right)$	A dummy variable indicating technologies with more than five assignees classified as $Assignee^{\text{Large}}_{\text{PreEmergence}}$.

Assignee-year-level Variables

$Assets_t$	Total assets (Compustat `at`) in year t.
$Sales_t$	Total sales (Compustat `sale`) in year t.
$Employment_t$	Total employment (Compustat `emp`) in year t.
$\log\left(\frac{Asset_{t+1}}{Asset_t}\right)$	Natural logarithm of total assets (Compustat `at`) in year $t+1$ divided by total assets in year t.
$\log\left(\frac{Sales_{t+1}}{Sales_t}\right)$	Natural logarithm of total sales (Compustat `sale`) in year $t+1$ divided by total sales in year t.
$\log\left(\frac{Empl_{t+1}}{Empl_t}\right)$	Natural logarithm of total employment (Compustat `emp`) in year $t+1$ divided by total employment in year t.
$\Delta\log\left(\frac{Asset_{t+1}}{Asset_t}\right)$	$\log\left(\frac{Asset_{t+1}}{Asset_t}\right)$ minus $\log\left(\frac{Asset_{t-1}}{Asset_{t-2}}\right)$.
$\Delta\log\left(\frac{Sales_{t+1}}{Sales_t}\right)$	$\log\left(\frac{Sales_{t+1}}{Sales_t}\right)$ minus $\log\left(\frac{Sales_{t-1}}{Sales_{t-2}}\right)$.
$\Delta\log\left(\frac{Empl_{t+1}}{Empl_t}\right)$	$\log\left(\frac{Empl_{t+1}}{Empl_t}\right)$ minus $\log\left(\frac{Empl_{t-1}}{Empl_{t-2}}\right)$.
$CapEx_t/PPENT_{t-1}$	Capital expenditure (Compustat `capx`) in year t divided by tangible assets (Compustat `ppent`) in year $t-1$.
$R\&D_t/Asset_{t-1}$	Research & development expenditure (Compustat `xrd`) in year t divided by total assets (Compustat `at`) in year $t-1$.
$\Delta CapEx_t/PPENT_{t-1}$	$CapEx_t/PPENT_{t-1}$ minus $CapEx_{t-1}/PPENT_{t-2}$.
$\Delta R\&D_t/Asset_{t-1}$	$R\&D_t/Asset_{t-1}$ minus $R\&D_{t-1}/Asset_{t-2}$.

ROA_t	Net income (Compustat `ni` for North American firms, Compustat `nicon` for other firms) in year t divided by total assets (Compustat `at`) in year $t-1$.
$Margin_t$	Operating income (Compustat `oibdp`) in year t divided by total sales (Compustat `sale`) in year $t-1$.
$ROIC_t$	Net operating profit after tax (Compustat `oiadp` $\times$ (1 - `txt/pi`)) in year t divided by invested capital (Compustat `dltt + dlc + seq - che`) in year $t-1$. We compute this measure only when the invested capital in year $t-1$ is greater than 0 (and set it to missing otherwise).
ΔROA_{t+1}	ROA_{t+1} minus ROA_t.
$\Delta Margin_{t+1}$	$Margin_{t+1}$ minus $Margin_t$.
$\Delta ROIC_{t+1}$	$ROIC_{t+1}$ minus $ROIC_t$.
$\log(Asset\, M/B_t)$	Sum of market value of equity (Compustat `csho` $\times$ `prcc_f` for North American firms, Compustat `cshoc` $\times$ `prccd` / `qunit` for other firms), debt in current liabilities (Compustat `dlc`), total long-term debt (Compustat `dltt`), and preferred stock (Compustat `pstkl`, or `pstk` if `pstkl` is missing) in year t, all divided by total assets (Compustat `at`) in year $t-1$. We compute this measure only when $\texttt{at}_{t-1} > 0$ (and set it to missing otherwise). We then take natural logarithm of this ratio.
$\log(P/E_t)$	Market value of equity (Compustat `csho` $\times$ `prcc_f` for North American firms, Compustat `cshoc` $\times$ `prccd` / `qunit` for other firms) in year t divided by income before extraordinary items (Compustat `ib`) in the year $t-1$. We compute this measure only when $\texttt{ib}_{t-1} > 0$ (and set it to missing otherwise). We then take natural logarithm of this ratio.

$\Delta\log(Asset\,M/B_t)$	$\log(Asset\,M/B_t)$ minus $\log(Asset\,M/B_{t-1})$.
$\Delta\log(P/E_t)$	$\log(P/E_t)$ minus $\log(P/E_{t-1})$.

B Pre-1976 Patent Titles

This appendix describes how we construct patent titles for patents granted before 1976 (grant year < 1976). We treat this period separately because, starting with patents granted in 1976, the United States Patent and Trademark Office (USPTO) systematically collects and disseminates structured patent metadata, whereas earlier records more often require extracting bibliographic fields from scanned documents.[29]

B.1 Source and Coverage

Our starting point is Google Patents,[30] which provides patent-level bibliographic fields on each patent's HTML page. For many early patents, the title displayed on Google Patents is derived from Optical Character Recognition (OCR) applied to scanned PDF images of the original document. We collect these titles via HTML scraping of all pre-1976 patents issued by USPTO starting with US1, yielding 3,923,712 OCR-based title records.

B.2 Error Patterns in OCR-derived Titles

OCR-derived titles contain systematic noise due to heterogeneity in front-page layouts and scan quality. In our data, two error patterns are particularly common.

Non-title text captured as the title. Front pages contain multiple prominent text blocks (e.g., inventor names, assignees, locations, and firms). OCR and downstream field extraction can misidentify these blocks and return strings that are not invention titles. A frequent variant of this error is that an inventor name (or fragments of it) appears in the title field, which is typically straightforward to detect. For example, Google Patents lists US958 with the title "hunt and j," which corresponds to inventor-name text rather than the invention title.

[29] Structured patent metadata are available at https://data.uspto.gov/bulkdata/datasets.
[30] Available at https://patents.google.com/.

Character-level OCR errors in the title region. Even when the title region is correctly identified, low-quality scans can produce insertions, deletions, and character confusions. For instance, Google Patents lists US1572 as "Apparatus forx t the t treatment of cilub-foot," whereas the intended title is "Apparatus for the treatment of club-foot." These errors complicate matching, parsing, and downstream technology classification.

B.3 Cleaning and Screening

We apply an iterative procedure that (a) corrects recurring OCR artifacts that are mechanically fixable and (b) flags titles that are likely incorrect, incomplete, or non-titles.

Step 1: Dictionary-based normalization for recurring OCR artifacts. We first target frequent artifacts that are reliably corrected with deterministic rules, especially those that manifest as short or fragmented tokens. We focus on short tokens for two reasons. First, common title function words that affect segmentation (e.g., "for," "of") are short; correcting OCR variants of these tokens improves consistent parsing. Second, OCR often introduces spurious whitespace that splits a single word into short fragments (e.g., "wheels" misread as "wh eels").

Rather than relying on a generic dictionary, we use custom dictionaries tailored to patent-title language and to how titles evolved over time (including period-specific vocabulary and orthographic conventions). These dictionaries are compiled and expanded iteratively as we observe recurring OCR artifacts in different historical periods.

Step 2: NLP-based detection of implausible or incomplete titles. Next, we use NLP screening rules to flag titles exhibiting one or more of the following patterns:

- **Human-name strings** consistent with inventor text appearing in the title field (e.g., "George t" US25438).
- **Other proper nouns** unlikely to be invention titles (e.g., organizations or locations

such as "london" in US749465), including cases arising from OCR or misclassification.

- **Incomplete or truncated titles**, such as strings that begin with conjunctions or prepositions or terminate prematurely (e.g., US95974 appears as "Improved apparatus for" instead of "Improved apparatus for applying roofing-composition to felt").

Titles flagged by these rules are treated as missing or unreliable and routed to supplementation. Because these patterns are not mutually exclusive, a single title may trigger multiple flags.

B.4 Supplementing Missing or Unreliable Titles

When OCR titles are missing or unreliable, we supplement them using two additional sources.

Amazon Mechanical Turk. We deploy Human Intelligence Tasks (HITs) on Amazon Mechanical Turk (MTurk)[31] in which workers review the original patent document and transcribe the invention title following standardized instructions. We assign one worker per title, and one of the co-authors subsequently checks the output quality and corrects issues when needed. This process yields 138,512 titles.

ChatGPT-4 extraction from the patent PDF header. We also extract titles using ChatGPT-4[32] by first extracting the top portion of the first page of the patent PDF (i.e., the header region where the title is typically printed) and then prompting the model to return the invention title as plain text. This targeted input reduces irrelevant page content and improves reliability relative to feeding the full document. This process yields 184,090 titles.

[31] See https://www.mturk.com/.
[32] For this task, we use https://openai.com/api/.

B.5 Iteration and Validation

The pipeline is iterative. After incorporating MTurk- and ChatGPT-derived titles, we re-apply the same screening rules to identify residual anomalies and address them in subsequent passes.

We validate accuracy via manual review using a stratified sampling design: for each grant-decade in the pre-1976 period, we sample 100 patents and verify the extracted title against the underlying document. We proceed with the pipeline until our audit indicates that, in each grant-decade, the resulting title dataset achieves over 98% accuracy.

C Extracting Technologies from Patent Titles

This appendix describes how we extract technology terms from patent titles. We combine (i) the pre-1976 titles constructed in Appendix B with (ii) post-1975 titles from PatentsView,[33] yielding titles for 11,804,242 USPTO patents granted between 1836 and the end of 2023.[34]

Our goal is to construct a vocabulary of multi-word expressions that proxy for technologies and to assign each patent the set of such technologies that appear in its title. To this end, we apply the following text-processing pipeline.

Before describing the text-processing pipeline, we briefly motivate why patent titles provide a defensible basis for technology identification. Patent titles are produced under legal and administrative constraints that make them informative summaries of the underlying invention. Under U.S. patent law, the application must disclose the invention in "full, clear, concise, and exact terms" sufficient to enable a person skilled in the art to make and use it, and it must conclude with claims "particularly pointing out and distinctly claiming" what the inventor regards as the invention (35 U.S.C. § 112(a)–(b)).[35] In addition, USPTO implementing regulations require a "title of the invention" that is "as short and specific as possible" and "reasonably indicative of the subject matter of the invention" (37 C.F.R. §1.72).[36] Together, these requirements create strong incentives for applicants to choose titles that foreground the key technological elements embodied in the claimed subject matter, while minimizing extraneous narrative detail. This legal design supports using cleaned n-grams from patent titles as a defensible proxy vocabulary for technologies in the patent record.

Step 1. Normalize titles. We standardize titles to ensure consistent tokenization across time and data sources: (i) convert all text to lowercase; (ii) remove parenthetical content

[33] Available at `https://patentsview.org/`.

[34] The last patent granted in 2023 is `US11856880`, implying coverage of 99.56% of granted patents in this time period. We verified that missing patents are not due to data gaps or processing errors: patents not present in our title data are those that are *secret or retracted*. For example, certain nuclear-energy-related inventions are subject to secrecy orders and are not publicly available (see, e.g., 35 U.S.C. §181 (Cornell LII)).

[35] See, e.g., 35 U.S.C. §112 (Cornell LII).

[36] See, e.g., 37 C.F.R. §1.72 (Cornell LII).

(e.g., abbreviations such as "(LED)"); (iii) remove punctuation except hyphens and commas; (iv) collapse repeated whitespace to single spaces; and (v) lemmatize nouns (singularization) using a standard lemmatization package augmented with custom dictionaries. The custom dictionaries are constructed to reflect the specific language of patent titles and how that language has evolved historically.

Step 2. Segment titles into phrases. We prevent n-grams from spanning unrelated components by first splitting each title into shorter *phrases*. We compile a list of 55 separators, including (a) prepositions (e.g., "to", "for", "with", "including"), (b) conjunctions (e.g., "and", "or"), (c) relative/adverbial markers (e.g., "where", "which"), and (d) multi-word connectors (e.g., "based on", "as well as"). We then split titles at these separators.

For example, US4236901 has the title "air cleaner with secondary air filter element." Using "with" as a separator yields two phrases: (1) "air cleaner" and (2) "secondary air filter element."

The separator list is constructed iteratively. We start from long titles, identify common connector terms that link distinct technological components, add them to the separator list, and re-segment titles. We repeat until 90% of resulting phrases contain fewer than four words. The four-word target balances granularity (isolating components) with interpretability (avoiding overly long phrases that bundle components).

Step 3. Remove stopwords within phrases. Within each segmented phrase, we remove a curated list of 566 stopwords. This list includes standard English stopwords (e.g., "a", "an", "any", "the") and patent-specific additions that we assemble manually based on recurring non-substantive title language. Stopword removal is applied *within* phrases (so that phrase boundaries from Step 2 are preserved).

Step 4. Generate contiguous n-grams from phrases. For each phrase $p = (w_1, \ldots, w_P)$ consisting of P words (after stopword removal), we generate all *contiguous* n-grams of lengths

$n \in \{1, \ldots, P\}$:

$$(w_i, \ldots, w_{i+n-1}) \quad \text{for } i = 1, \ldots, P - n + 1.$$

Thus, each phrase yields $P(P+1)/2$ n-grams. For instance, the phrase "secondary air filter element" generates: "secondary air filter element", "secondary air filter", "air filter element", "secondary air", "air filter", "filter element", "secondary", "air", "filter", and "element".

Step 5. Exclude uni-grams. We discard uni-grams because single words often cannot reliably distinguish technologies from generic vocabulary (Bloom et al., 2021). This approach excludes a small number of early patents with one-word titles (e.g., US1768757 "Lamp") or technologies described solely by uni-grams (e.g., in US14938 "Improvement in machinery for making rope and cordage," both "rope" and "cordage" are discarded). This limitation is less consequential for later patents that we use for our analyses, where titles typically contain multi-word technological descriptions.

Step 6. Standardize hyphenation. Hyphenation in n-grams is highly inconsistent and can generate redundant variants of the same technological concept (e.g., "typewriting machine", "type-writing machine", and "type writing machine"). We therefore standardize hyphenated forms *before* applying frequency filters and the phrase-completeness filter below.

We focus this step on high-frequency terms. For each grant year, we select the 300 n-grams with the largest patent counts and identify those containing hyphens. This yields 3,683 distinct hyphenated n-grams requiring standardization (e.g., "steam-engine", "paddle-wheel", "car-coupling"). For each, we replace hyphens with spaces (e.g., "steam engine") and treat this spacing-only form as the canonical representation throughout the corpus.

Step 7. Apply a minimum-frequency filter. We retain only n-grams that appear in at least three distinct patents over the full sample period. This reduces noise from spelling errors, OCR artifacts, and idiosyncratic fragments, and yields a more stable vocabulary for longitudinal analysis.

Step 8. Filter by phrase completeness. Many n-grams occur mainly as modifiers or fragments rather than stand-alone technological concepts (e.g., "high speed"), or as incomplete parts of longer phrases (e.g., "internal combustion"). To remove such cases, we compute a *full phrase ratio* for each n-gram g and year t:

$$\text{Full Phrase Ratio}_{g,t} = \frac{\#\{\text{patents in year } t \text{ where } g \text{ equals an entire phrase}\}}{\#\{\text{patents in year } t \text{ containing } g \text{ within any phrase}\}}.$$

Here, "phrase" refers to the segmented phrases from Step 2 *after* stopword removal in Step 3. We exclude n-grams with Full Phrase Ratio$_{g,t} < 0.10$. This criterion keeps terms that frequently function as complete technological units in titles while removing partial descriptors and fragments that rarely stand alone.

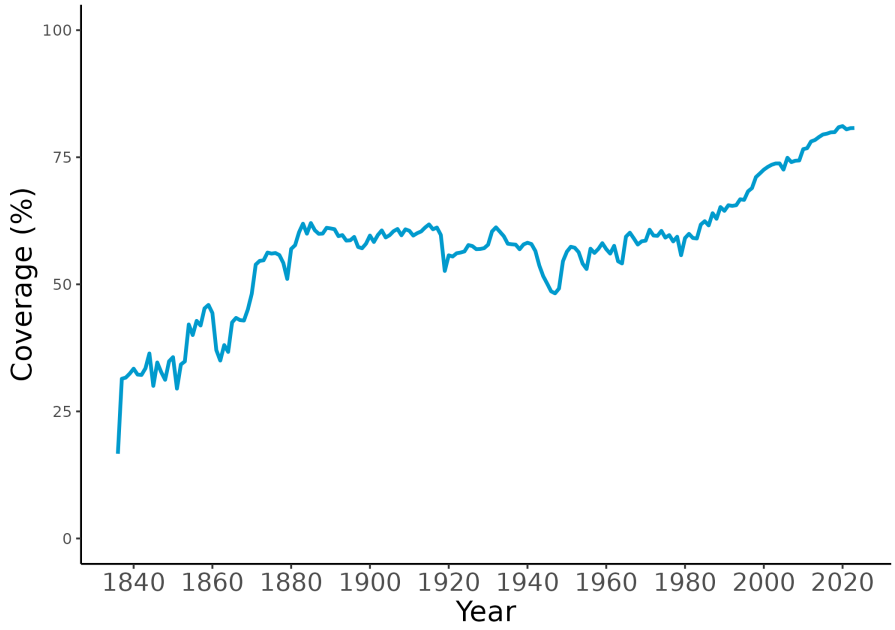
Step 9. Remove invention-type descriptors. Finally, to focus on substantive technologies rather than generic invention categories, we compile a list of 22 non-technological descriptor terms (e.g., "method", "process") and remove any n-gram composed exclusively of these terms (e.g., "processing method").

Output. After applying Steps 1–9, we obtain 715,788 unique multi-word n-grams. We interpret these n-grams as technology proxies and assign each patent the set of retained n-grams that appear in its processed title.

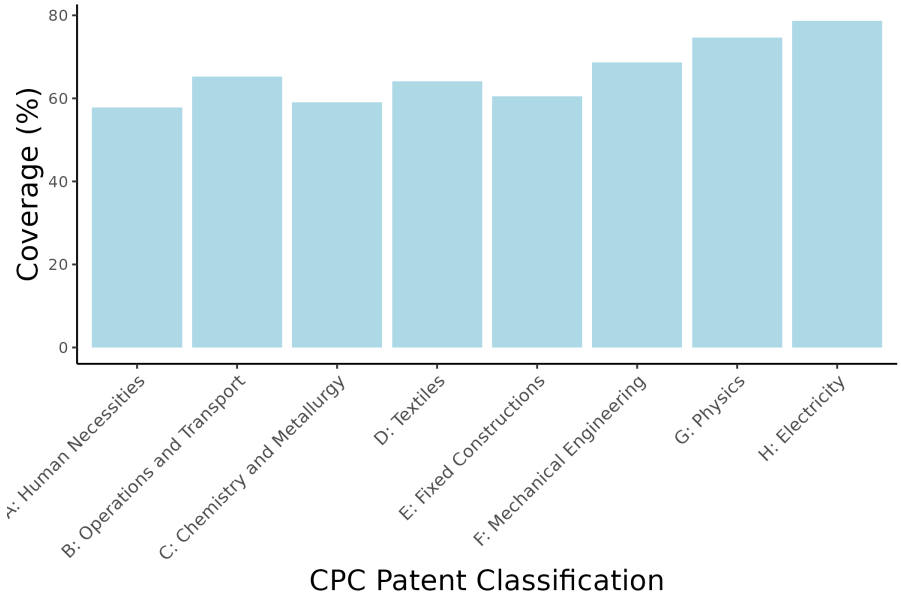
Descriptive statistics. Figure C1a and Figure C1b report the share of patents covered by these technologies over time and across CPC classes, respectively. Figure C2 plots the mean number of extracted technologies per title by grant year. Figure C3 plots the log number of distinct co-occurring technology pairs (where a pair is defined as two technologies appearing in the same title). Table C1 lists the largest technologies by decade, measured by the total number of patents associated with each technology.

Figure C1: Patent Coverage of Title-Based Technologies

(a) Coverage by Grant Year

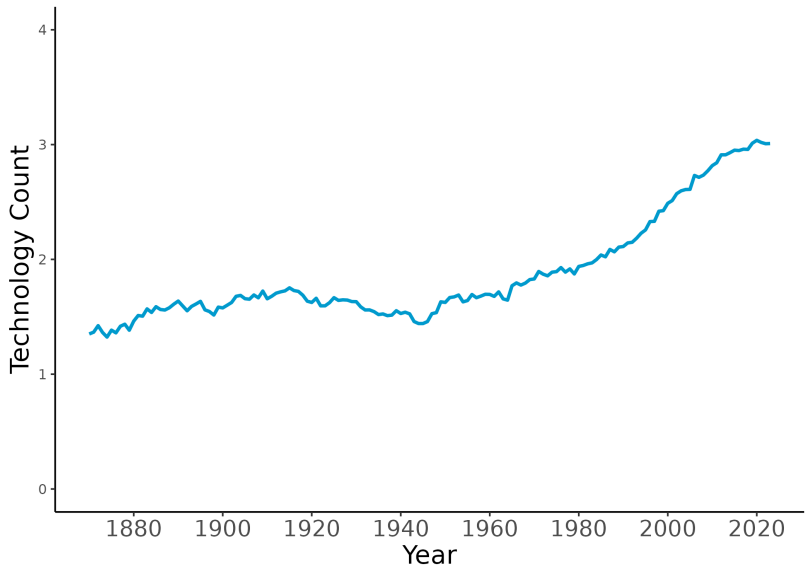


(b) Coverage by CPC Section



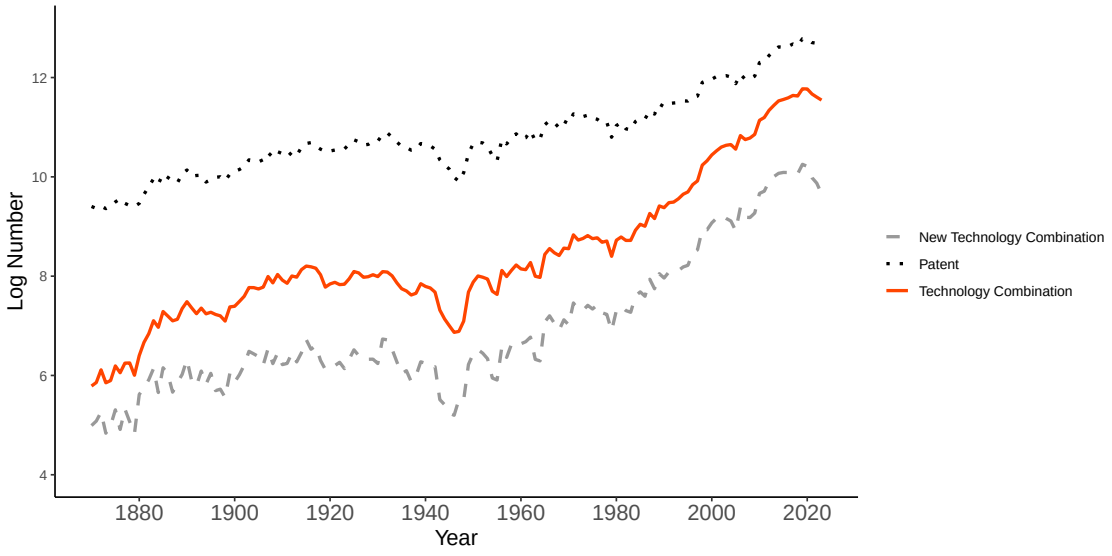
Note: This figure reports the share of USPTO patents covered by our technology vocabulary, where technologies are proxied by n-grams extracted from patent titles. The sample period is 1836–2023. Panel (a) shows annual coverage by grant year. Panel (b) shows coverage by one-digit CPC section.

Figure C2: Technologies per Patent Title Over Time



Note: This figure plots the average number of technologies identified from each patent title by grant year, 1870–2023. Technologies are proxied by n-grams extracted from titles.

Figure C3: Patents and Technology Combinations Over Time



Note: This figure plots the natural logarithm of (i) the number of granted patents, (ii) the number of distinct technology combinations, and (iii) the number of new technology combinations by grant year, 1870–2023. A technology combination is defined as the co-occurrence of two technologies within a patent title. Technologies are proxied by n-grams extracted from titles.

Table C1: Top Five Technologies by Decade

Period	1	2	3	4	5
Pre-1870	steam engine	fire arm	corn planter	washing machine	sewing machine
1870-1879	car coupling	steam engine	corn planter	sewing machine	bed bottoms
1880-1889	car coupling	sewing machine	steam engine	fire escape	washing machine
1890-1899	car coupling	sewing machine	electric railway	type writing machine	steam engine
1900-1909	type writing machine	rail joint	rotary engine	sewing machine	nut lock
1910-1919	internal combustion engine	sewing machine	type writing machine	rail joint	nut lock
1920-1929	internal combustion engine	motor vehicle	type writing machine	sewing machine	telephone system
1930-1939	internal combustion engine	motor vehicle	sewing machine	refrigerate apparatus	electric switch
1940-1949	internal combustion engine	sewing machine	circuit breaker	signal system	motor vehicle
1950-1959	internal combustion engine	sewing machine	motor vehicle	circuit breaker	gas turbine
1960-1969	internal combustion engine	control circuit	sewing machine	nuclear reactor	control valve
1970-1979	internal combustion engine	control circuit	motor vehicle	control valve	coat composition
1980-1989	internal combustion engine	motor vehicle	heat exchanger	optical fiber	pharmaceutical composition
1990-1999	internal combustion engine	motor vehicle	communication system	record medium	image form apparatus
2000-2009	communication system	display device	record medium	liquid crystal display	image form apparatus
2010-2019	display device	communication system	electronic device	image form apparatus	liquid crystal display
Post-2019	display device	electronic device	storage medium	communication system	display panel

Note: This table lists the five most prevalent technologies in each decade, ranked by the number of patents whose titles include the corresponding technology n-gram. Technologies are proxied by n-grams extracted from patent titles.

D Patent–Firm Matching: Constructing 2024 version of the Global Corporate Patent Dataset (GCPD)

This appendix describes matching of patents to firms. We match USPTO patents from PatentsView[37] to all public firms in the S&P Compustat North America and Compustat Global universe (henceforth, "Compustat"). Our approach builds on the Global Corporate Patent Dataset (GCPD),[38] first employed by Bena et al. (2017), and yields a new release of the GCPD. Conceptually, we map patent assignees to firm identifiers. Below we describe the construction in detail and provide summary statistics and figures for the resulting dataset. The workflow is: (i) define the sets of names used for matching; (ii) generate candidate links using automated methods; (iii) aggregate candidates into a set of matches using decision rules; and (iv) manually refine and validate the matches.

D.1 Sets of Firm Names Used for Matching

PatentsView. We collect assignee entities from PatentsView that are classified as organizations. PatentsView provides a statistically based assignee-name disambiguation that often corrects misspellings, but it can also over-aggregate by mapping multiple distinct assignee names to a single disambiguated entity. We improve this disambiguation by clustering the raw assignee names that share the same PatentsView disambiguated identifier and then splitting overly broad clusters into distinct assignee entities. This procedure increases the number of disambiguated assignee names from 475,277 to 542,305.

Compustat. From Compustat we collect the current and historical company names in `conm` and `conml`. To expand historical coverage, we additionally use the WRDS-maintained `wciklink` table, which links Compustat `gvkeys` to former company names.

[37] Available at https://patentsview.org.
[38] Available at https://patents.darden.virginia.edu/.

Normalization. We normalize all company and assignee names using a standardized set of rules tailored to corporate-name formatting. For example, we apply a library of more than 1,000 regular expressions that capture common variants and map them to a standardized form. We also standardize capitalization and whitespace and replace malformed characters arising from historical inconsistencies in USPTO data entry.

Output. We combine all Compustat company names and PatentsView assignee organization names into a single name file indexed by `data_src` (Compustat or PatentsView), `name_src` (e.g., `conm` vs. `conml`), and `id` (`gvkey` for Compustat; assignee identifier for PatentsView). The combined file contains 695,406 distinct names in total: 542,305 from PatentsView and 156,744 from Compustat.

D.2 Methods of Matching

Fuzzy string matching. Using the normalized names, we compute the OSA string similarity for each pair of PatentsView assignee names and Compustat company names. For each PatentsView assignee, we retain **all** Compustat `gvkey`s for which at least one associated Compustat name attains a similarity score of 0.8 or higher.

Google searches. For each name in our master list, we query Google Search and collect the first 100 results. From these results we extract and normalize URLs from Bloomberg, Facebook, LinkedIn, Twitter/X, and Wikipedia. We then parse platform-specific identifiers from URLs (e.g., from `https://www.bloomberg.com/profile/company/AAPL:CB` we extract "AAPL:CB"). For each platform, we add a candidate match between a PatentsView assignee name and a Compustat company name when the evidence is sufficiently specific for that platform (e.g., Bloomberg and LinkedIn results typically correspond to a particular company, whereas generic topic pages on Wikipedia are not always informative). We also compare URLs from Google Search to the firm homepage URL recorded in Compustat. Finally, we check for overlap in the top five non-platform results via common aggregator

websites, which can provide additional corroboration.

D.3 Aggregation of Potential Matches

The previous step produces a set of candidate links. Each candidate link connects a PatentsView assignee name to a Compustat company name and is enriched with a vector of indicators/weights capturing support from the different matching methods. We construct the final set of matches using a rule-based aggregation that specifies which combinations of evidence constitute an accepted match. For instance, candidates with fuzzy string similarity equal to 1 are accepted mechanically. We also incorporate geographic information (city, state, country) collected from the underlying sources to resolve ambiguous cases. PatentsView assignees often correspond to subsidiaries within multinational business groups that operate in multiple countries; in such cases, we search within the set of Compustat firms belonging to the relevant business group and select the best-supported match.

D.4 Refinement of Matches

Last, we refine the match set by benchmarking against matches produced by other researchers, most notably the 2017 release of the Global Corporate Patent Dataset (GCPD) and the patent-to-securities links for North American firms in Kogan et al. (2017). We manually reconcile all discrepancies relative to the 2017 GCPD and the matches provided by Kogan et al. (2017). In addition, we apply the following refinements.

Highest never matched. We identify disambiguated assignee names that remain unmatched and rank them by the number of patents assigned to the assignee. We then manually determine whether each assignee corresponds to a Compustat firm; if not (e.g., research institutes or government entities), we add the assignee to a blacklist of entities that should not be matched.

Highest potentially matched. We identify assignees that receive candidate links to Compustat firms but fail to satisfy any aggregation rule, and we again rank them by patent count. We manually verify whether a candidate link is correct, identify the correct match when one exists, or blacklist the assignee when it should not map to a public firm.

Short, unique company identifiers. We take the 1,000 assignees with the largest patent counts and extract short, unique company identifiers from the beginning of their names (e.g., from "Samsung Electronics Co Ltd" we extract "Samsung"). We manually validate these identifiers. We then use them to screen the remaining unmatched assignees with few patents by flagging names that begin with a validated identifier. To avoid overly aggressive matching, we do not assign a match when other firm identifiers appear in the assignee name (e.g., joint ventures).

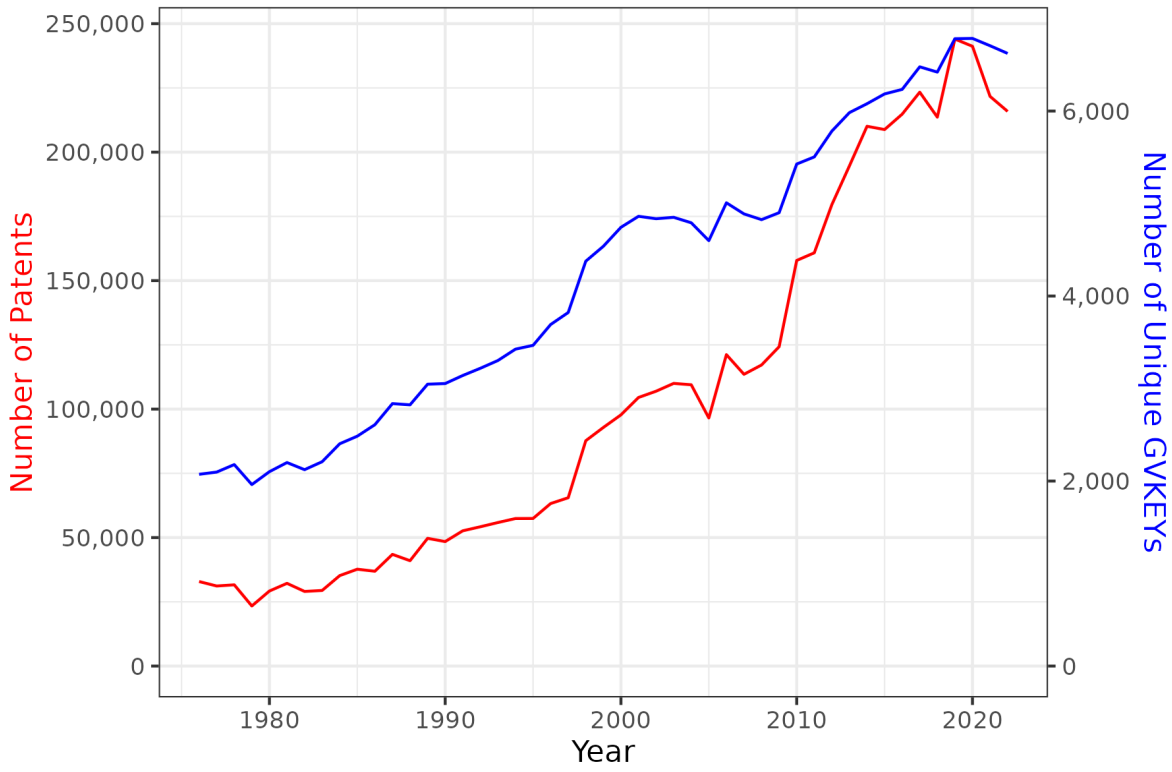
Timing correction. Up to this point, each assignee name is linked to a single `gvkey` representing the best match under the rules above. However, the relevant corporate entity can change over time (e.g., due to mergers and acquisitions), so the correct mapping may be time-varying. To enforce timing consistency, we use S&P Capital IQ's link between Compustat `gvkey`s and Capital IQ company identifiers. We only link a patent to a `gvkey` if that `gvkey` is valid at the patent's filing date; when it is not, we use the `gvkey`–CIQ linkage to identify the appropriate `gvkey` that is valid at filing.

D.5 Descriptive Statistics

Our initial patent sample contains 7,596,563 USPTO patents granted between January 1976 and September 2023. We match 5,328,861 of these patents to public firms in Compustat North America or Compustat Global. Because a patent can have multiple matched assignees, the final matched file contains 5,441,198 assignee–patent records. Figure D1 plots, by grant year, the number of matched patents and the number of distinct matched Compustat firms (`gvkey`s). Figure D2 compares matched patent counts in the 2017 and 2024 vintages of

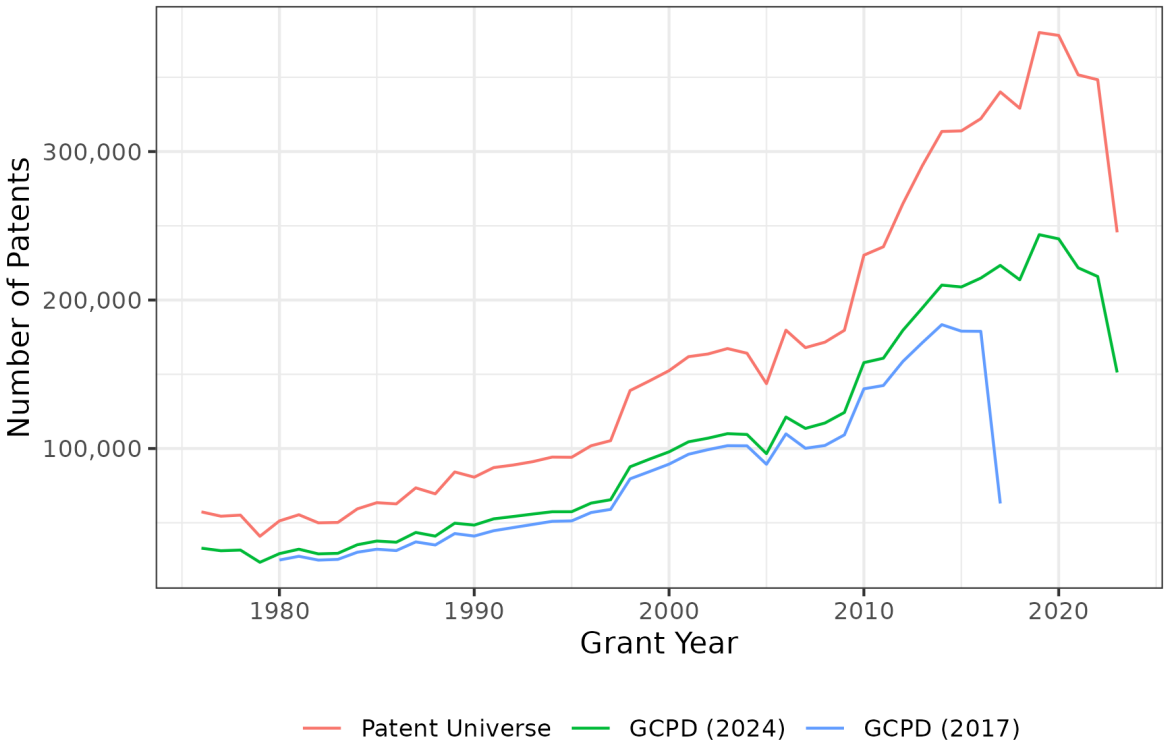
GCPD to the total number of granted patents by year. Figure D3 and Figure D4 summarize the composition of matched patents by headquarters country and industry, respectively. Figure D5 shows the distribution of patent grants across assignee types.

Figure D1: Matched Patents and Distinct Compustat Firms by Grant Year



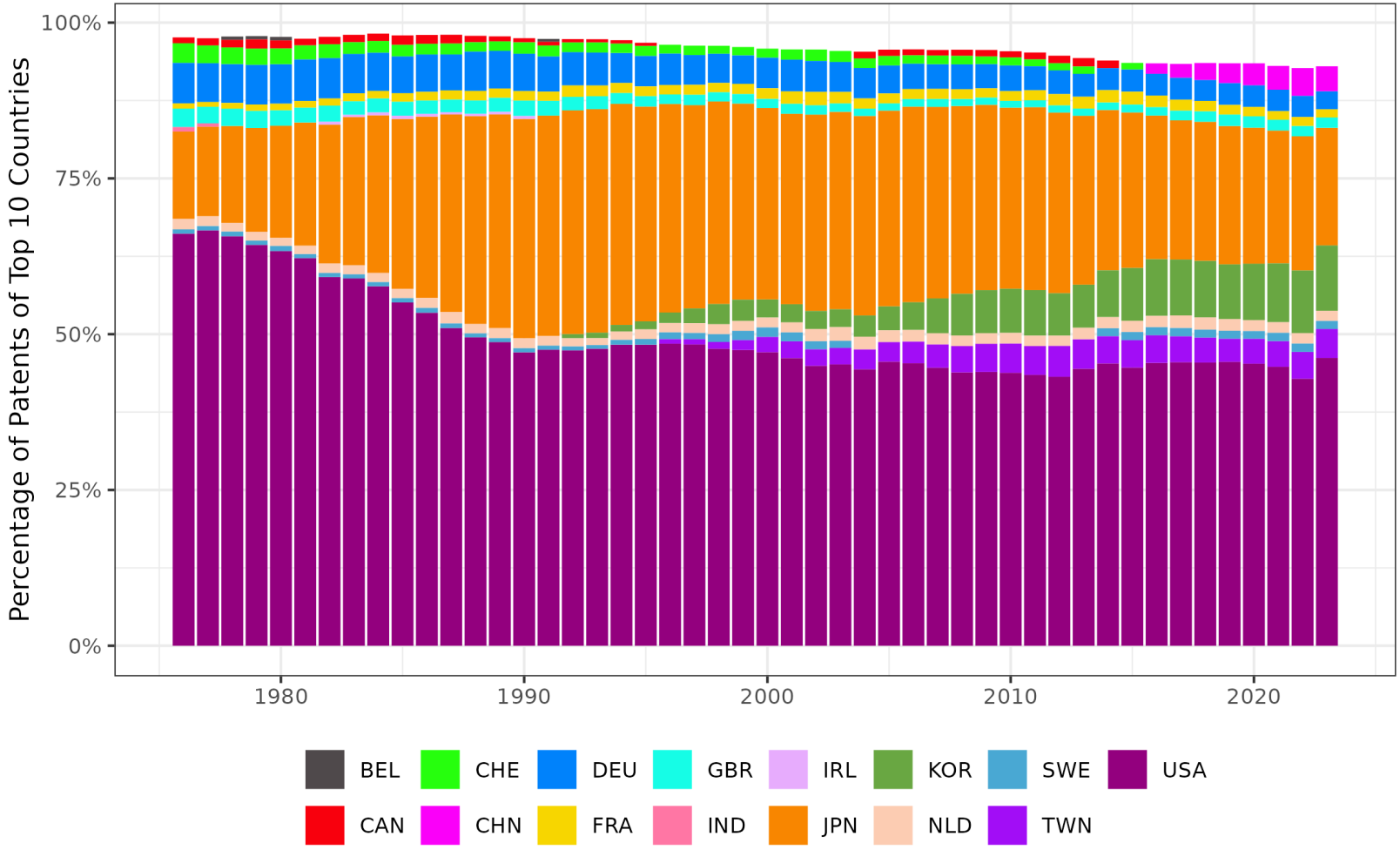
Note: The red line shows the number of USPTO patents matched to Compustat firms in each grant year. The blue line shows the number of distinct Compustat firm identifiers (`gvkeys`) receiving at least one matched patent in that year.

Figure D2: Patent-Matching Coverage Across GCPD Vintages



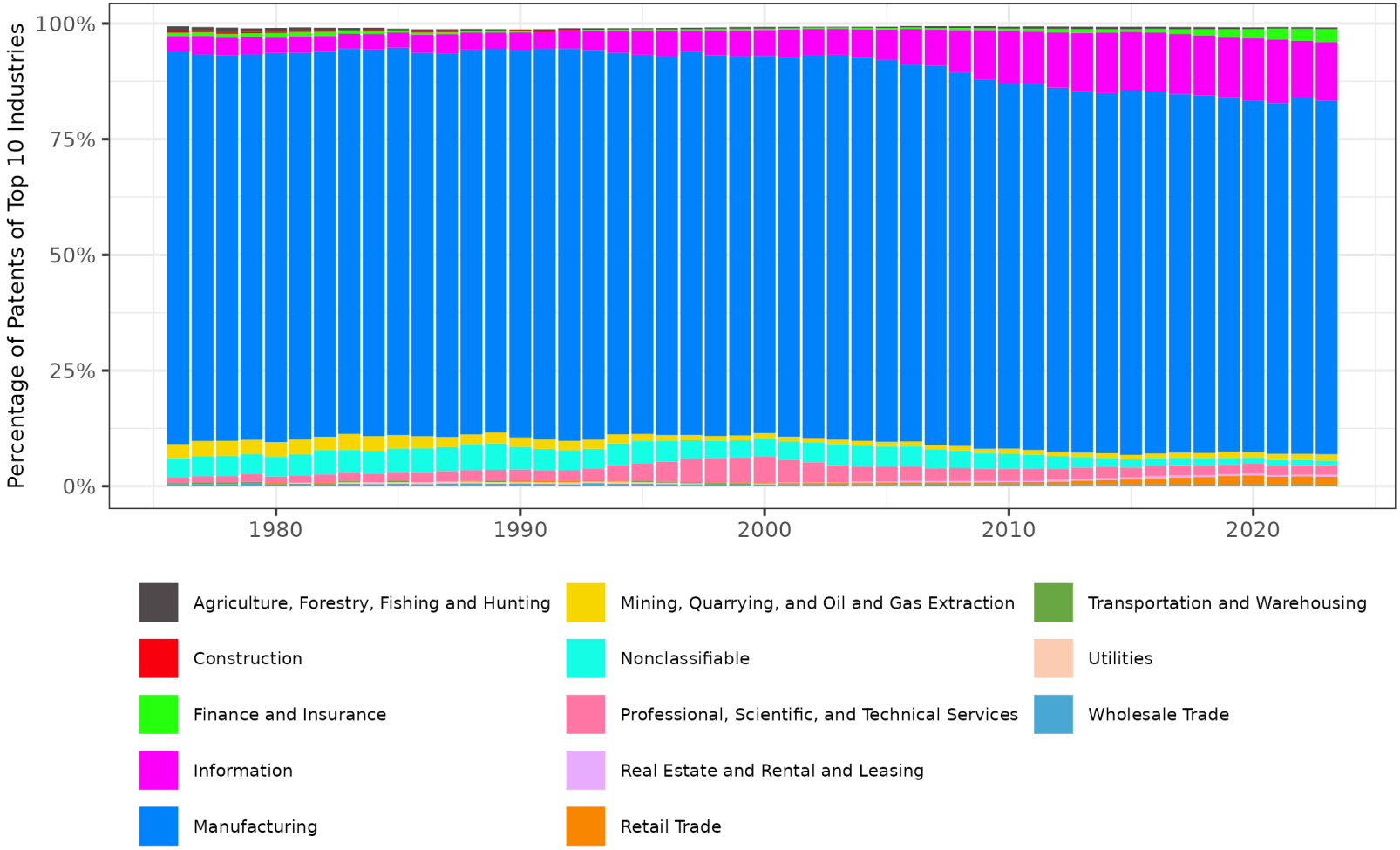
Note: The red line shows the total number of USPTO patents granted in each year. The blue line shows the number of patents matched in the 2017 GCPD release. The green line shows the number of patents matched in the current (2024) GCPD release.

Figure D3: Distribution of Matched Patents by Headquarters Country



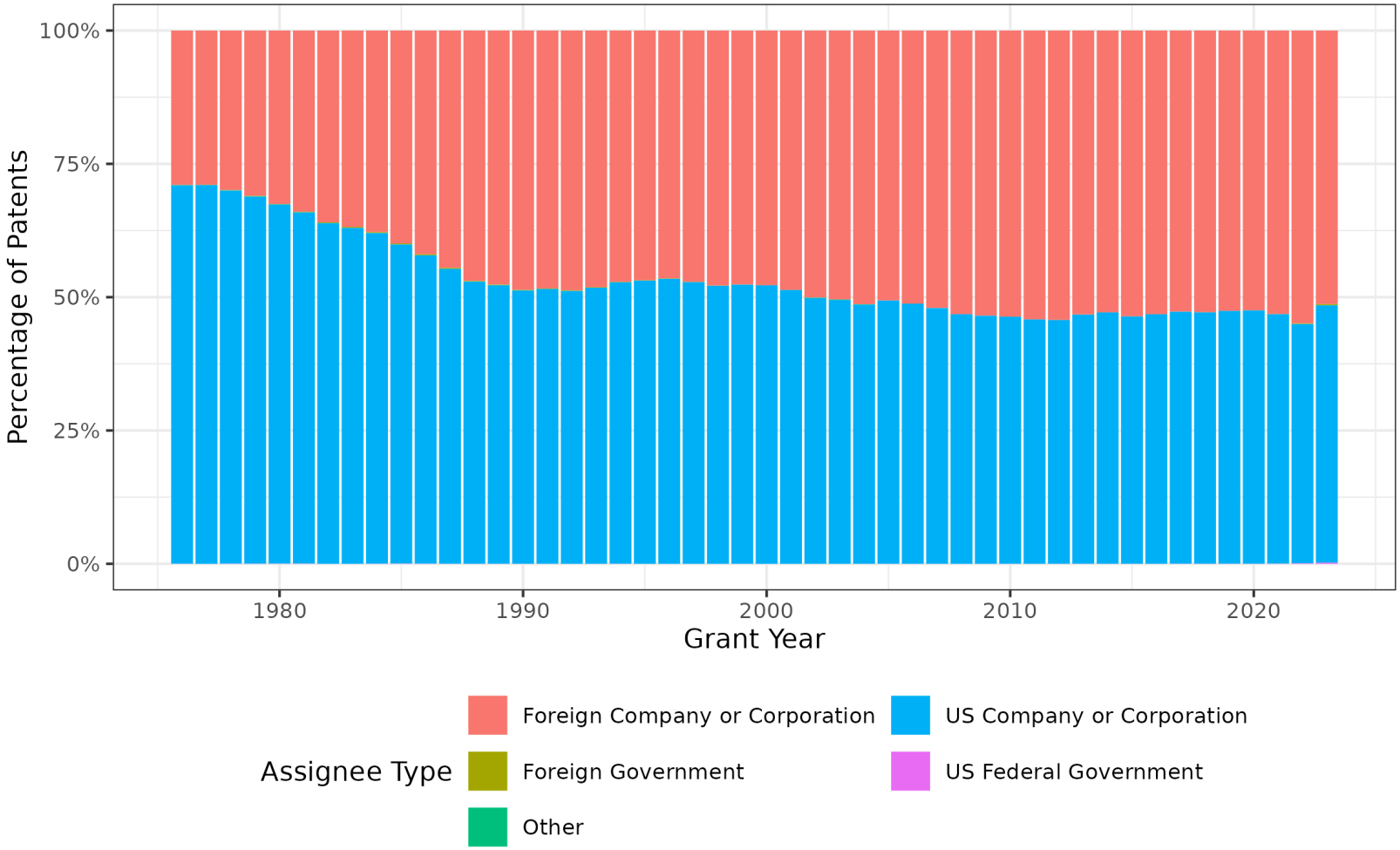
Note: The figure reports, by grant year, the share of matched patents assigned to firms headquartered in each country. Countries are defined using the Compustat headquarters location of the matched firm. The plot displays the 10 countries with the largest number of matched patent grants in each year.

Figure D4: Distribution of Matched Patents by Industry



Note: The figure reports, by grant year, the share of matched patents by industry. Each patent is assigned the industry of its matched firm using the first two digits of the firm's 2-digit NAICS code in Compustat. The plot displays the 10 industries with the largest number of matched patent grants in each year.

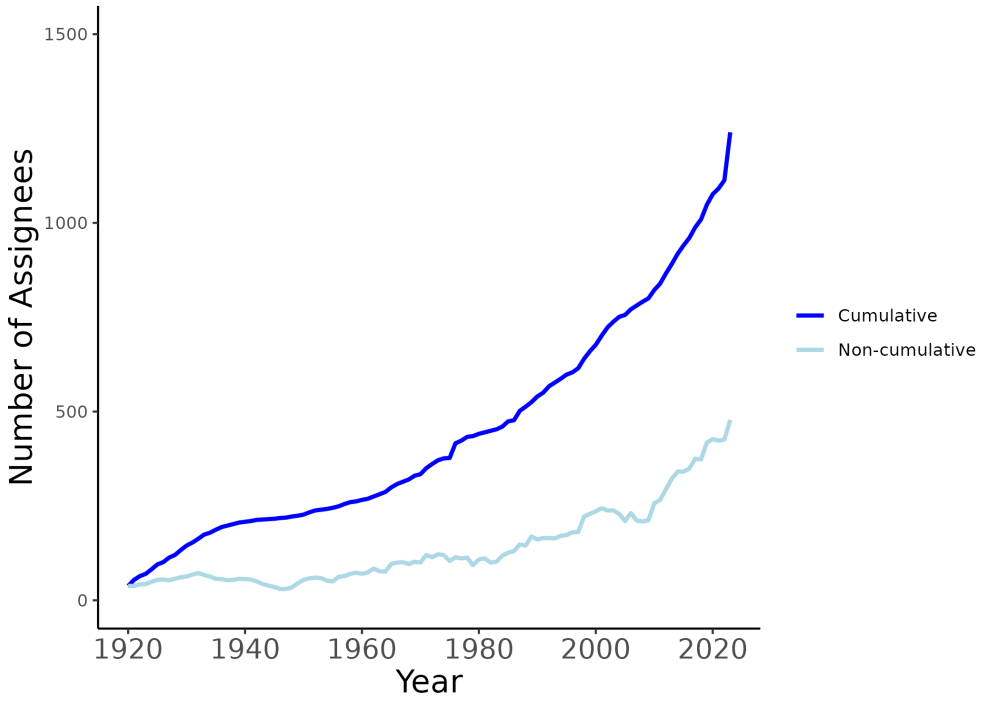
Figure D5: Distribution of Matched Patents by Assignee Type



Note: The figure reports, by grant year, the share of matched patents by assignee type. Assignee types follow USPTO designation and combine organization type (e.g., company, government, other) and location (U.S. vs. foreign).

E Supplementary Figures

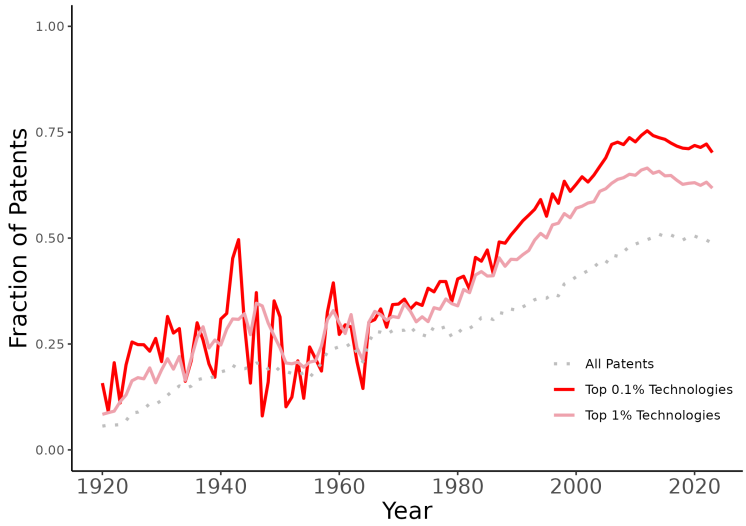
Figure A1: Top-1% Firm Innovators Over Time



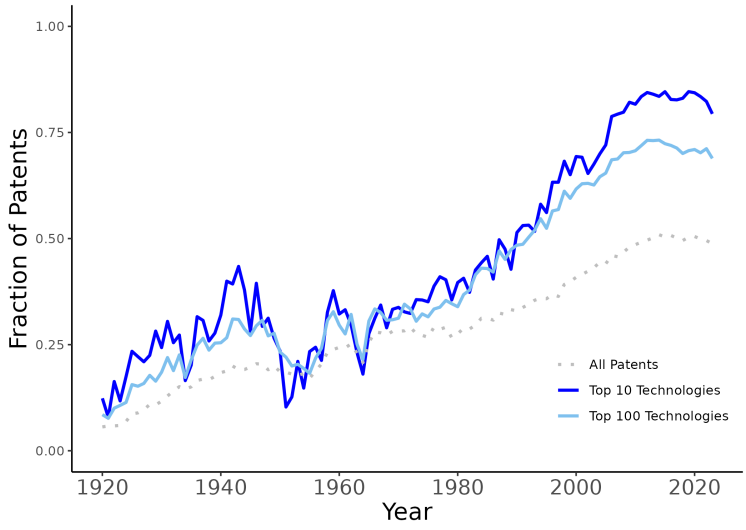
Note: This figure plots the number of firm assignees that rank in the top 1% of grant-year patent counts in each year (light blue line) and the cumulative number of distinct firms that have ever ranked in the top 1% since 1920 (dark blue line), 1920–2023.

Figure A2: Presence of Large Innovators in Dominant Technologies: PatentsView Data

(a) Top-0.1% and Top-1% Technologies

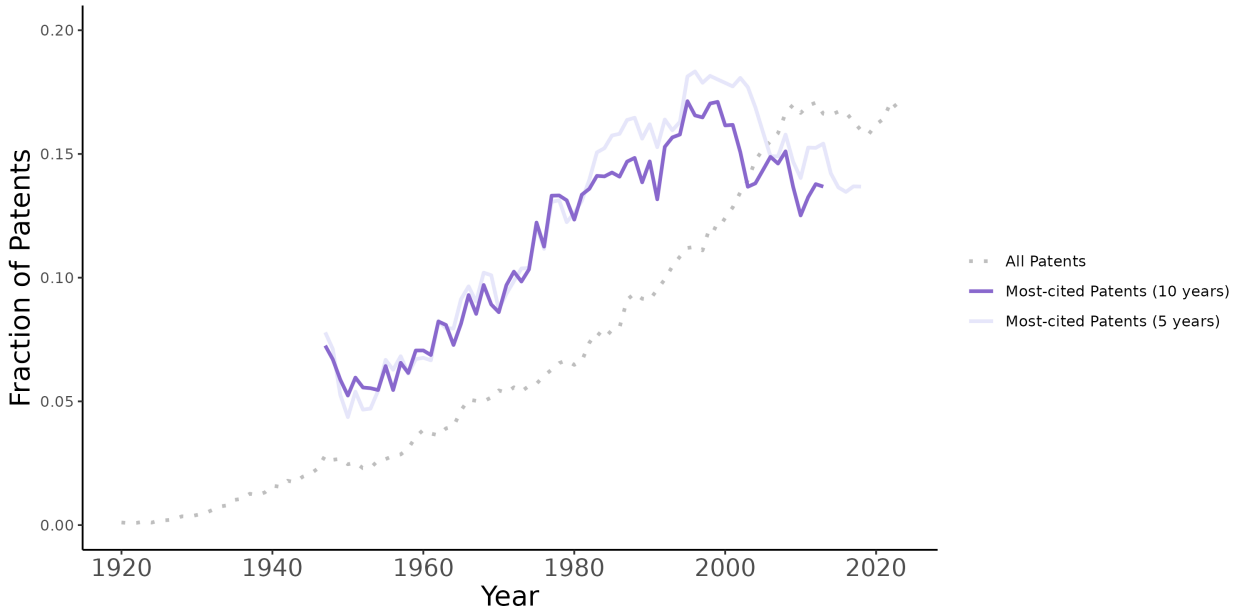


(b) Top-10 and Top-100 Technologies



Note: This figure tracks the fraction of patents granted to the top 1% of firm assignees over time, where Post-1976 patent data comes from PatentsView instead of GCPD. Panel (a) displays this fraction specifically for patents containing technologies ranked in the top 0.1% (red line) or top 1% (pink line). Panel (b) shows the same metric for patents that include at least one of the top 10 (blue line) or top 100 (light blue line) technologies. In both panels, the gray dotted line serves as a benchmark, representing the overall fraction of all patents granted to the top 1% of firms.

Figure A3: Patent Coverage by Prominent Technologies



Note: This figure plots the share of patents covered by 289 prominent technologies in our sample. The gray dotted line represents the coverage of all patents granted between 1920 and 2023. The purple (light purple) solid line indicates the fraction of highly cited patents within ten (five) years that are covered by prominent technologies.

Figure A4: Pre-emergence Innovators in Prominent Technologies: Examples

(a) Electronic Device



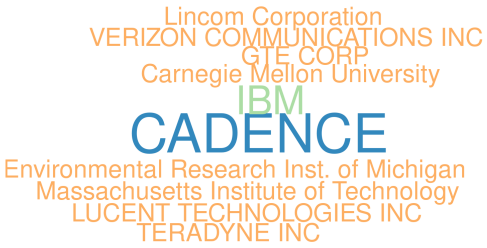
(b) Liquid Crystal



(c) Circuit Board



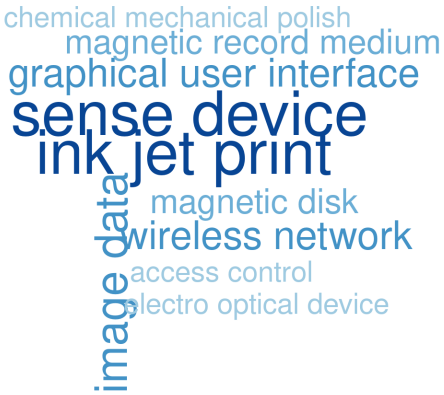
(d) Machine Learning



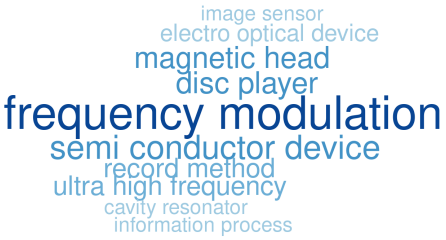
Note: This figure illustrates pre-emergence innovators (assignees with at least one patent granted on or before the technology's emergence year) for four example prominent technologies: *electronic device* (Panel (a)), *liquid crystal* (Panel (b)), *circuit board* (Panel (c)), and *machine learn* (Panel (d)). In each panel, the size of an assignee label is proportional to the number of pre-emergence patents assigned to that assignee in the focal technology.

Figure A5: Prominent Technologies of Selected Pre-emergence Innovators

(a) IBM



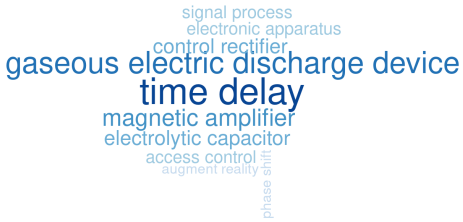
(b) RCA



(c) AT&T

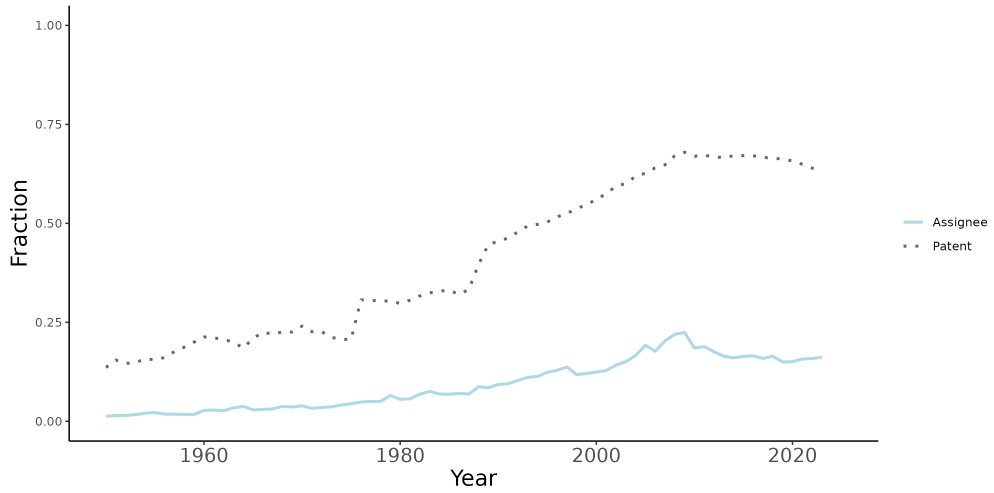


(d) GE



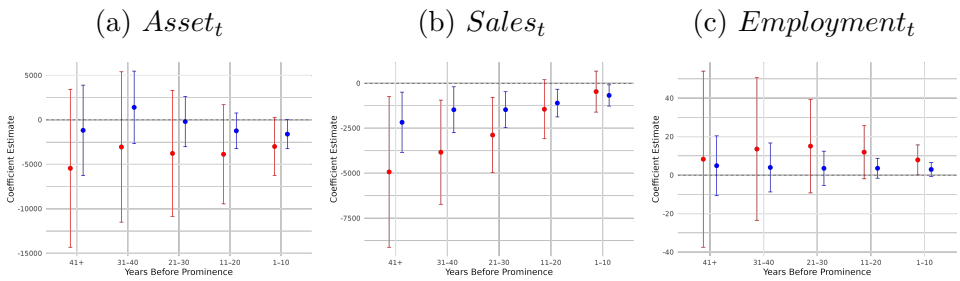
Note: This figure illustrates the set of prominent technologies in which four selected assignees—IBM (Panel (a)), RCA (Panel (b)), AT&T (Panel (c)), and GE (Panel (d))—are pre-emergence innovators (i.e., hold at least one patent granted on or before the technology's emergence year). Within each panel, the size of a technology term is proportional to the number of the technology's pre-emergence patents assigned to the focal assignee.

Figure A6: Coverage of Sample Assignees and Patents



Note: This figure illustrates the data coverage for the sample used in the firm investment and performance analyses. The solid blue line depicts the annual share of assignees with at least one patent grant that are included in the sample. The gray dotted line represents the proportion of total patents granted that are attributed to these sampled assignees.

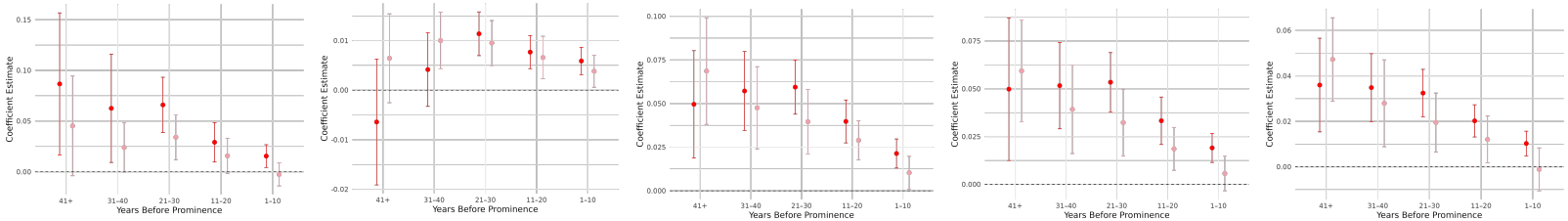
Figure A7: Firm Size Before Prominence: Pre-emergence and Pre-prominence Innovators



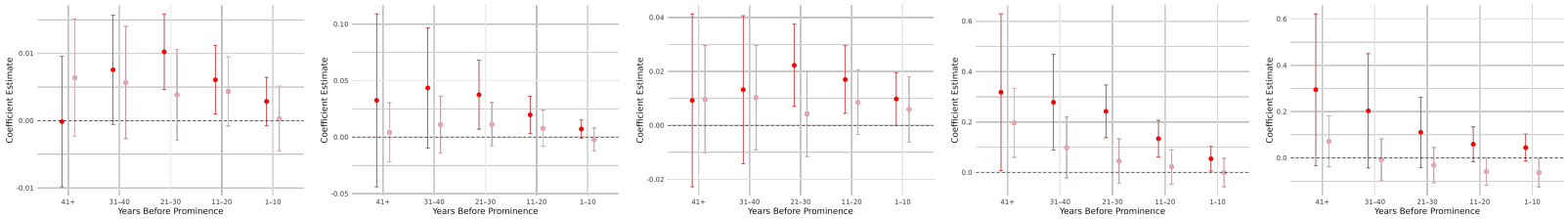
Note: This figure reports coefficient estimates from equation (5). The dependent variable $Y_{i,t}$ corresponds to the outcome indicated in the title of each panel. The red (blue) points present the estimated coefficient for β_{group} (γ_{group}) for the specific decade, with vertical lines indicating the 95% confidence interval.

Figure A8: Firm Outcomes Before Prominence: Pre-emergence Innovators (Levels)

(a) $\frac{CapEx_t}{PPENT_{t-1}}$ (b) $\frac{R\&D_t}{Asset_{t-1}}$ (c) $\log\left(\frac{Asset_{t+1}}{Asset_t}\right)$ (d) $\log\left(\frac{Sales_{t+1}}{Sales_t}\right)$ (e) $\log\left(\frac{Empl_{t+1}}{Empl_t}\right)$



(f) ROA_t (g) $Margin_t$ (h) $ROIC_t$ (i) $\log(\text{Asset M/B}_t)$ (j) $\log(\text{P/E}_t)$

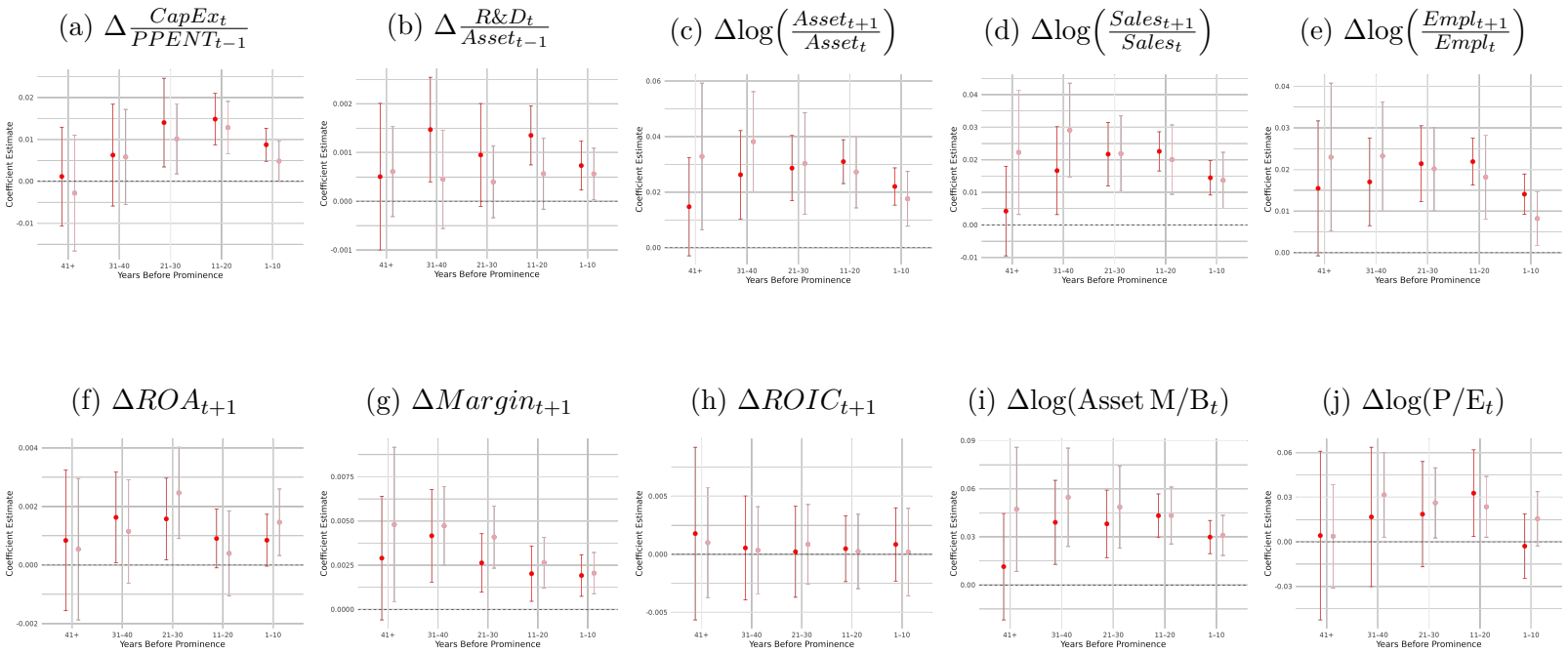


Note: This figure reports coefficient estimates from linear regression

$$
\begin{aligned}
Y_{i,t} = & \sum_{group \in \{1\text{–}10,11\text{–}20,21\text{–}30,31\text{–}40,41+\}} \beta_{group,1}\, \text{Assignee}_{\text{PreEmergence},i,j} \times \text{Assignee}^{\text{Active}}_{\text{PostProminence},i,j} \times Decade_{group,j,t} \\
& + \sum_{group \in \{1\text{–}10,11\text{–}20,21\text{–}30,31\text{–}40,41+\}} \beta_{group,2}\, \text{Assignee}_{\text{PreEmergence},i,j} \times \text{Assignee}^{\text{Inactive}}_{\text{PostProminence},i,j} \times Decade_{group,j,t} \\
& + \sum_{group \in \{1\text{–}10,11\text{–}20,21\text{–}30,31\text{–}40,41+\}} \gamma_{group,1}\, \text{Assignee}_{\text{PreProminence},i,j} \times \text{Assignee}^{\text{Active}}_{\text{PostProminence},i,j} \times Decade_{group,j,t} \\
& + \sum_{group \in \{1\text{–}10,11\text{–}20,21\text{–}30,31\text{–}40,41+\}} \gamma_{group,2}\, \text{Assignee}_{\text{PreProminence},i,j} \times \text{Assignee}^{\text{Inactive}}_{\text{PostProminence},i,j} \times Decade_{group,j,t} \\
& + \delta X_{i,t} + \alpha_i + \mu_{j,t} + \epsilon_{i,j,t},
\end{aligned}
\tag{7}
$$

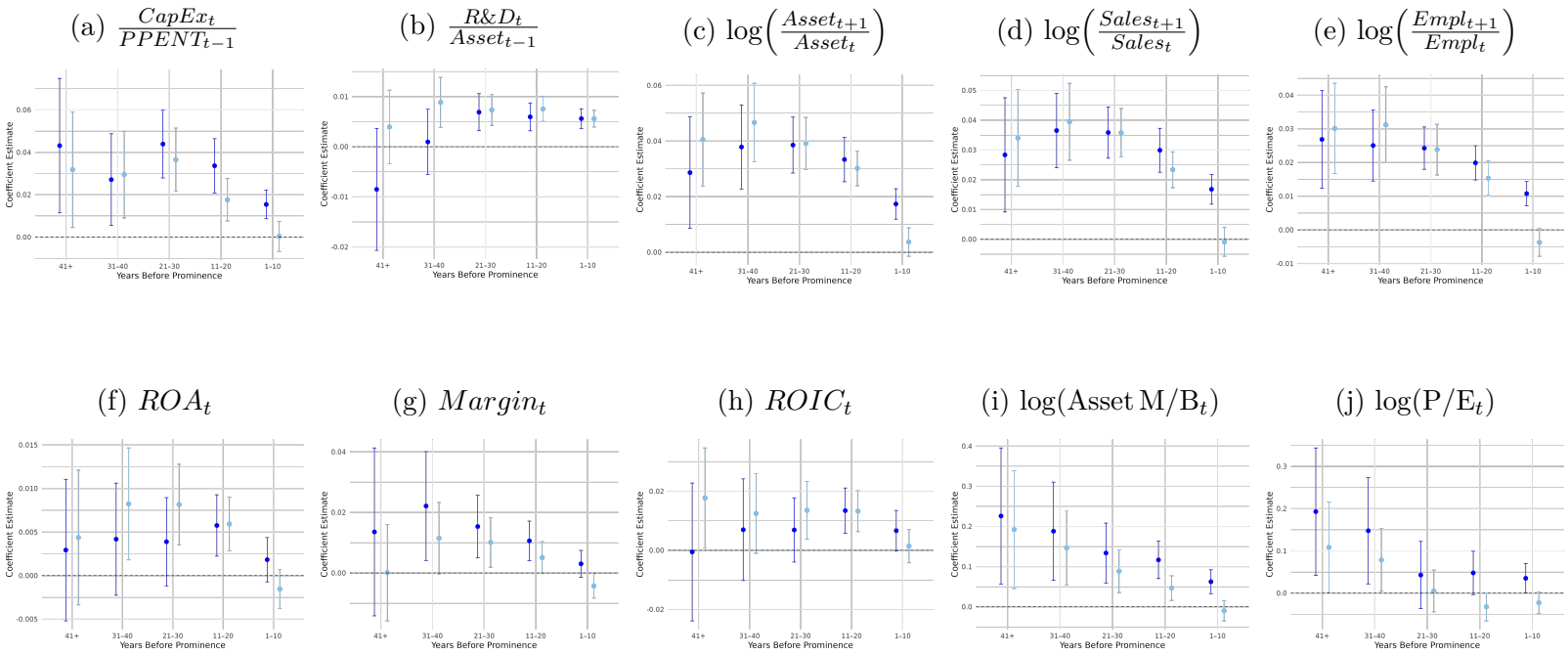
where i, j, and t represent firm, technology and year, respectively. The dependent variable $Y_{i,t}$ corresponds to the outcome indicated in the title of each panel. The dummy variable $\text{Assignee}_{\text{PreEmergence},i,j}$ ($\text{Assignee}_{\text{PreProminence},i,j}$) indicates whether firm i is a pre-emergence (pre-prominence) assignee of technology j. The dummy variable $\text{Assignee}^{\text{Active}}_{\text{PostProminence},i,j}$ ($\text{Assignee}^{\text{Inactive}}_{\text{PostProminence},i,j}$) equals one if firm i has at least one (zero) post-prominence patent in technology j. $Decade_{group,j,t}$ is dummy variable indicating decades before prominence, as shown by the x-axis of each figure. We control for firm fixed effects (α_i), technology-by-year fixed effects ($\mu_{j,t}$) and firm-year level variables ($X_{i,t}$) including firm age and assets, along with their squared terms. We double-cluster standard errors at the firm and technology levels. The red (pink) points present the estimated coefficient for $\beta_{group,1}$ ($\beta_{group,2}$) for the specific decade, with vertical lines indicating the 95% confidence interval.

Figure A9: Firm Outcomes Before Prominence: Pre-emergence Innovators (Changes)



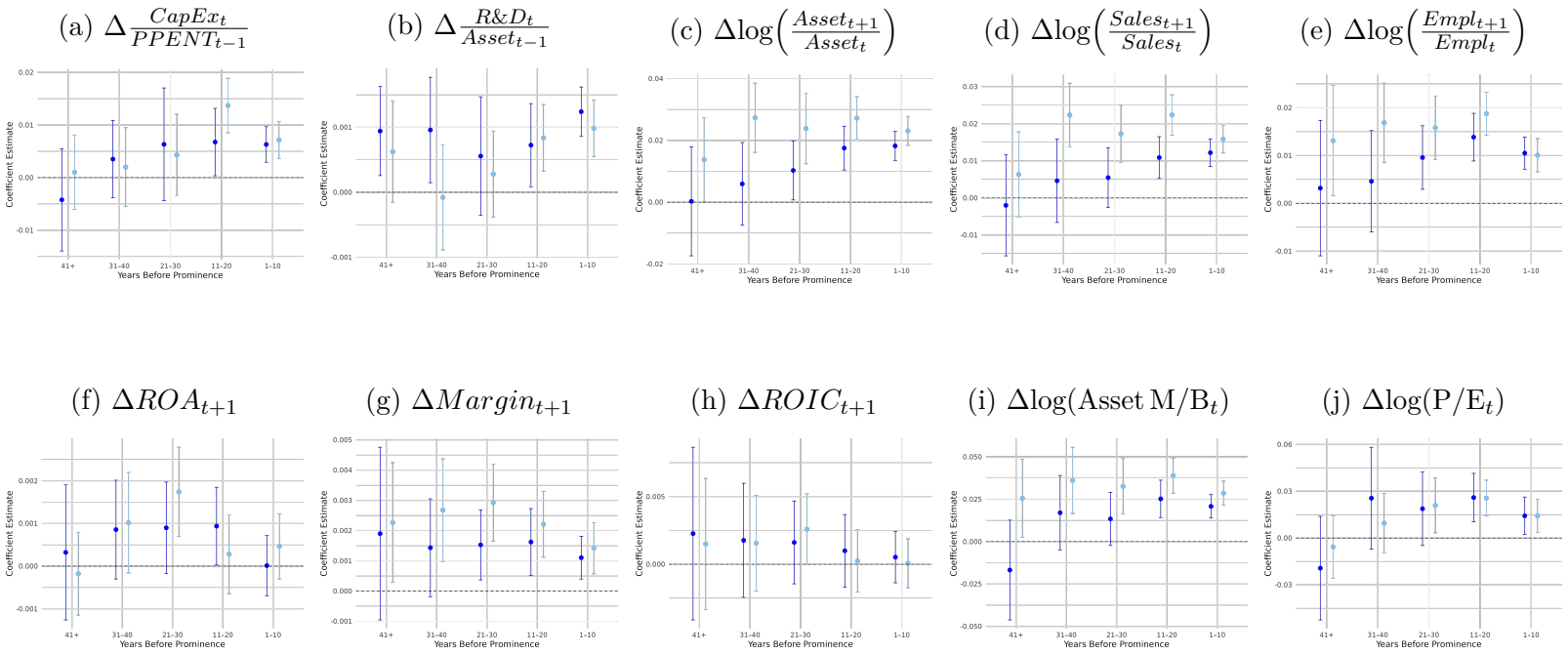
Note: This figure reports coefficient estimates from equation (7). The dependent variable $Y_{i,t}$ corresponds to the outcome indicated in the title of each panel. The red (pink) points present the estimated coefficient for $\beta_{group,1}$ ($\beta_{group,2}$) for the specific decade, with vertical lines indicating the 95% confidence interval.

Figure A10: Firm Outcomes Before Prominence: Pre-prominence Innovators (Levels)



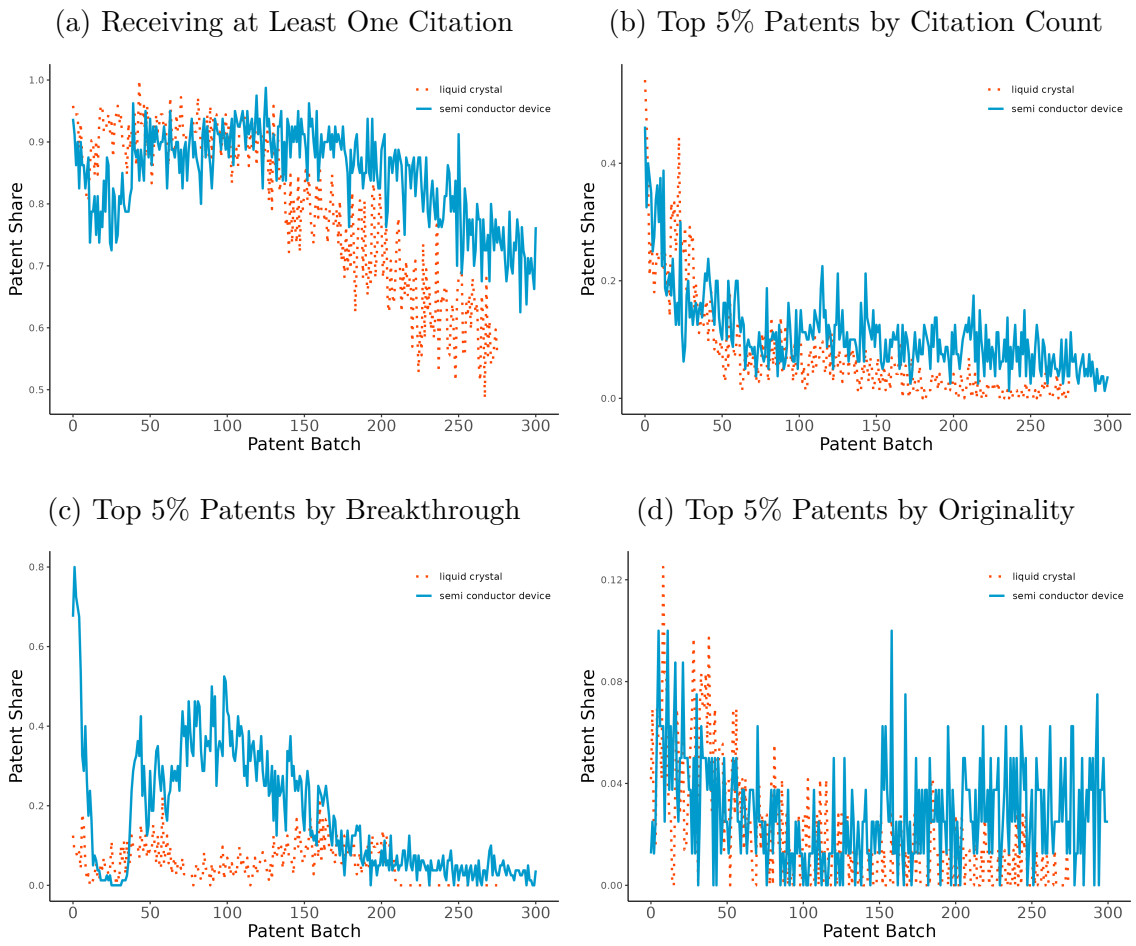
Note: This figure reports coefficient estimates from equation (7). The dependent variable $Y_{i,t}$ corresponds to the outcome indicated in the title of each panel. The blue (light blue) points present the estimated coefficient for $\gamma_{group,1}$ ($\gamma_{group,2}$) for the specific decade, with vertical lines indicating the 95% confidence interval.

Figure A11: Firm Outcomes Before Prominence: Pre-prominence Innovators (Changes)



Note: This figure reports coefficient estimates from equation (7). The dependent variable $Y_{i,t}$ corresponds to the outcome indicated in the title of each panel. The blue (light blue) points present the estimated coefficient for $\gamma_{group,1}$ ($\gamma_{group,2}$) for the specific decade, with vertical lines indicating the 95% confidence interval.

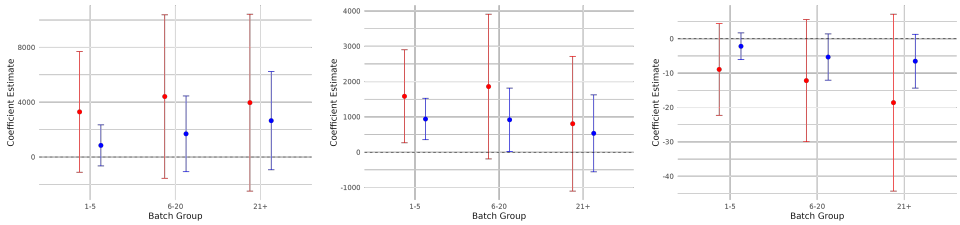
Figure A12: Patent Quality Across the Technology Life Cycle



Note: This figure tracks the evolution of patent quality, including citations, breakthrough level, and originality, over the life cycles of two representative technologies (*liquid crystal* and *semi conductor device*). We divide each technology's patents into sequential batches, with each batch's size determined by the number of patents granted before/in the technology's year of prominence. Panel (a) displays the fraction of patents in each batch that received at least one citation within five years of grant. Panel (b) presents the share of patents in the top 5% of five-year citation counts relative to all patents granted in the same year. Panel (c) presents the share of patents ranking in the top 5% based on the breakthrough measure by Kelly et al. (2021). Panel (d) presents the share of patents ranking in the top 5% based on originality.

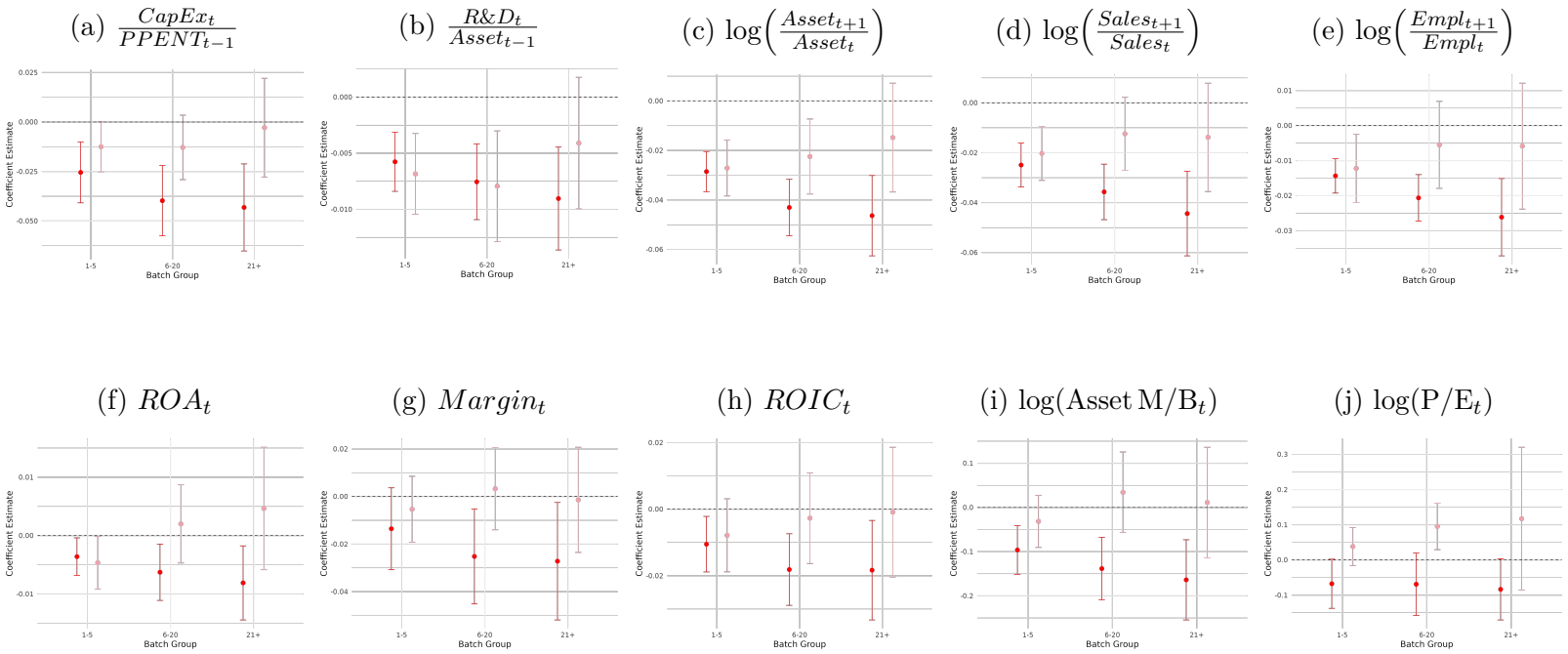
Figure A13: Firm Size After Prominence: Pre-emergence and Pre-prominence Innovators

(a) $Asset_t$ (b) $Sales_t$ (c) $Employment_t$



Note: The figure displays the coefficient estimates from equation (4). The dependent variable $Y_{i,t}$ corresponds to the outcome indicated in the title of each panel. The red (blue) points present the estimated coefficient for β_{group} (γ_{group}) for the specific batch group shown by the x-axis, with vertical lines indicating the 95% confidence interval.

Figure A14: Firm Outcomes After Prominence: Pre-emergence Innovators (Levels)

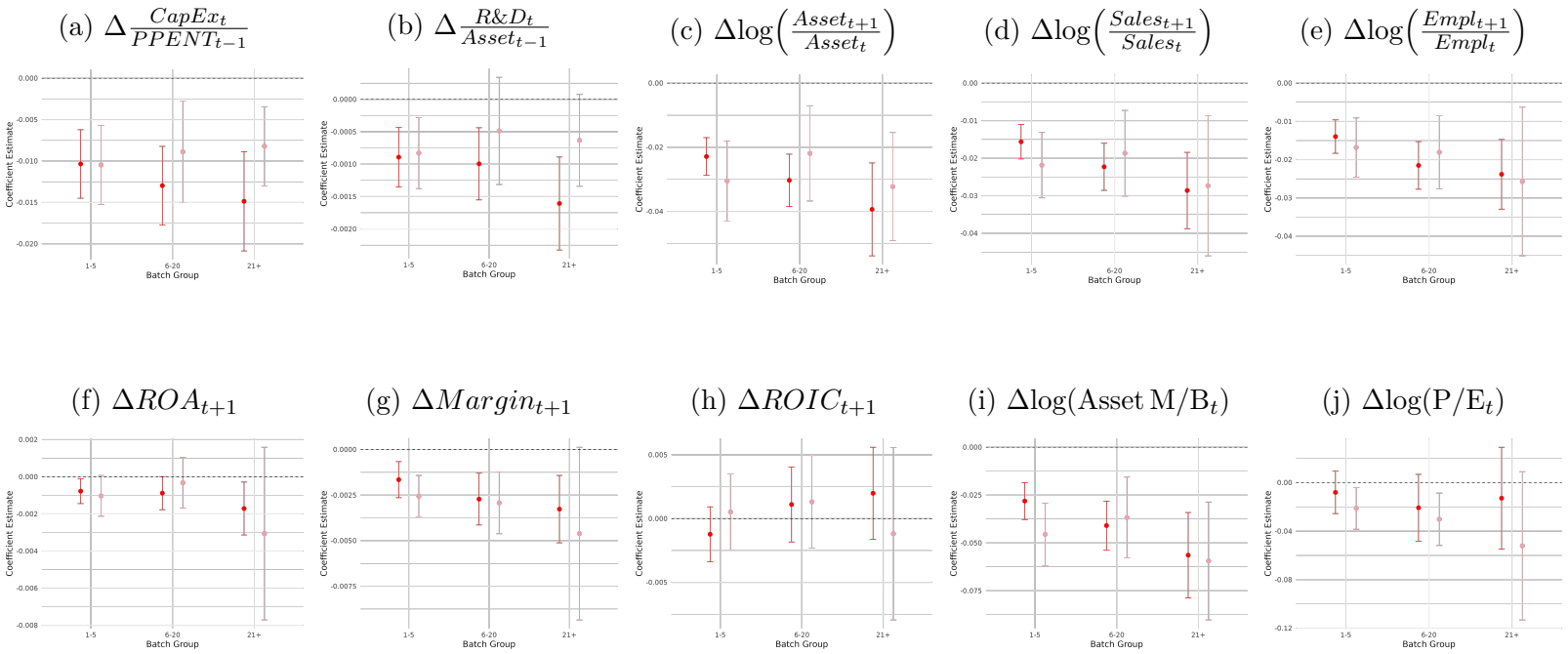


Note: This figure reports coefficient estimates from linear regression

$$
\begin{aligned}
Y_{i,t} = & \sum_{group \in \{1\text{–}5, 6\text{–}20, 21+\}} \beta_{group,1} \, \text{Assignee}_{\text{PreEmergence},i,j} \times \text{Assignee}^{\text{Active}}_{\text{PostProminence},i,j} \times Batch_{group,j,t} \\
& + \sum_{group \in \{1\text{–}5, 6\text{–}20, 21+\}} \beta_{group,2} \, \text{Assignee}_{\text{PreEmergence},i,j} \times \text{Assignee}^{\text{Inactive}}_{\text{PostProminence},i,j} \times Batch_{group,j,t} \\
& + \sum_{group \in \{1\text{–}5, 6\text{–}20, 21+\}} \gamma_{group,1} \, \text{Assignee}_{\text{PreProminence},i,j} \times \text{Assignee}^{\text{Active}}_{\text{PostProminence},i,j} \times Batch_{group,j,t} \\
& + \sum_{group \in \{1\text{–}5, 6\text{–}20, 21+\}} \gamma_{group,2} \, \text{Assignee}_{\text{PreProminence},i,j} \times \text{Assignee}^{\text{Inactive}}_{\text{PostProminence},i,j} \times Batch_{group,j,t} \\
& + \delta X_{i,t} + \alpha_i + \mu_{j,t} + \epsilon_{i,j,t},
\end{aligned}
\tag{8}
$$

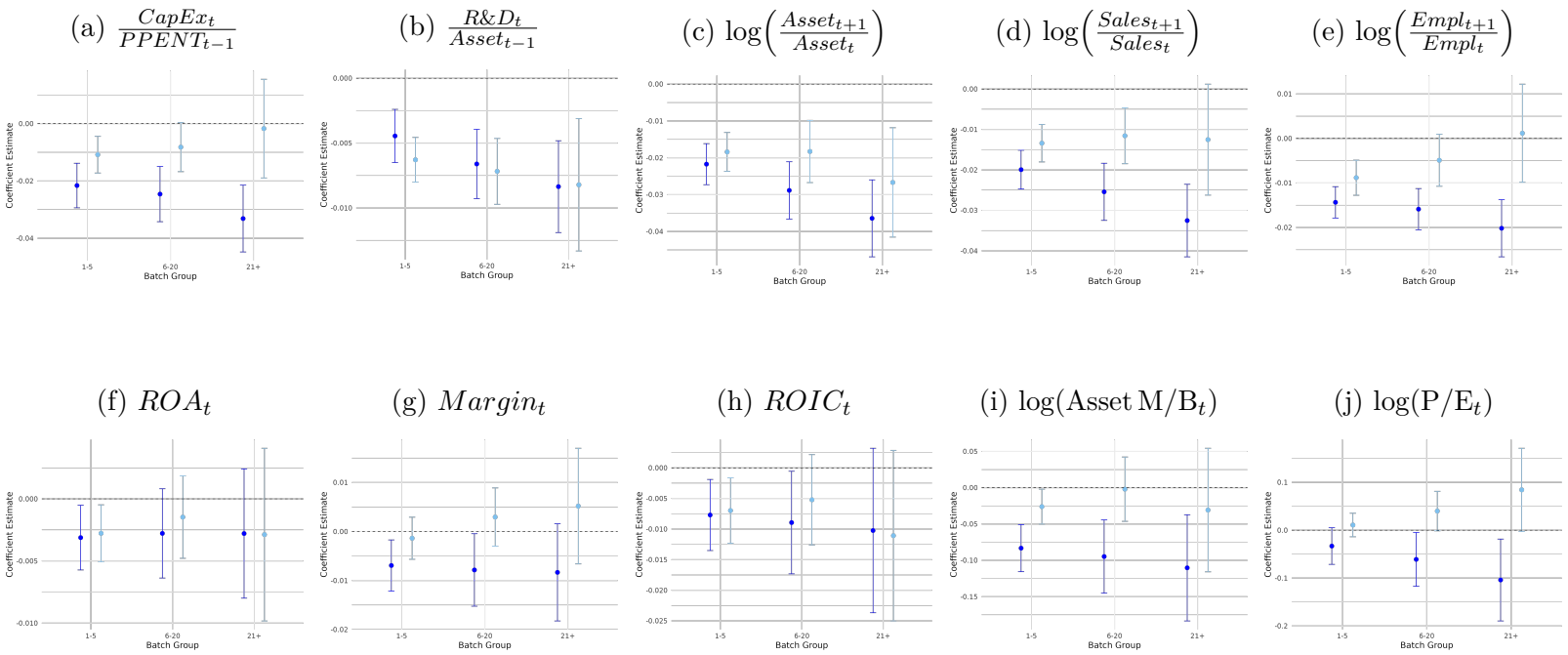
where i, j, and t represent firm, technology and year, respectively. The dependent variable $Y_{i,t}$ corresponds to the outcome indicated in the title of each panel. The dummy variable $\text{Assignee}_{\text{PreEmergence},i,j}$ ($\text{Assignee}_{\text{PreProminence},i,j}$) indicates whether firm i is a pre-emergence (pre-prominence) assignee of technology j. The dummy variable $\text{Assignee}^{\text{Active}}_{\text{PostProminence},i,j}$ ($\text{Assignee}^{\text{Inactive}}_{\text{PostProminence},i,j}$) equals one if firm i has at least one (zero) post-prominence patent in technology j. $Batch_{Group,j,t}$ is dummy variable indicating batch groups, as shown by the x-axis of each figure. We control for firm fixed effects (α_i), technology-by-year fixed effects ($\mu_{j,t}$) and firm-year level variables ($X_{i,t}$) including firm age and assets, along with their squared terms. We double-cluster standard errors at the firm and technology levels. The red (pink) points present the estimated coefficient for $\beta_{group,1}$ ($\beta_{group,2}$) for the specific batch group, with vertical lines indicating the 95% confidence interval.

Figure A15: Firm Outcomes After Prominence: Pre-emergence Innovators (Changes)



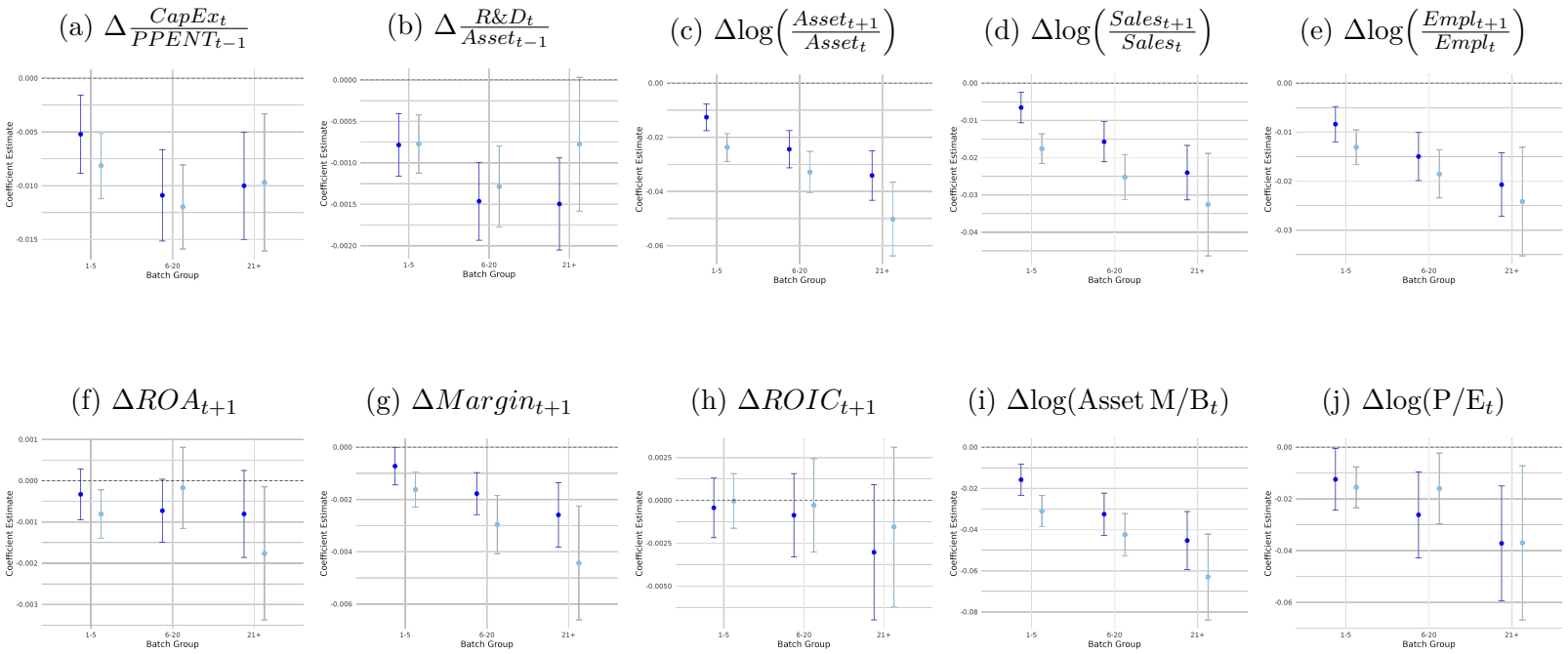
Note: This figure reports coefficient estimates from equation (8). The dependent variable $Y_{i,t}$ corresponds to the outcome indicated in the title of each panel. The red (pink) points present the estimated coefficient for $\beta_{group,1}$ ($\beta_{group,2}$) for the specific batch group, with vertical lines indicating the 95% confidence interval.

Figure A16: Firm Outcomes After Prominence: Pre-prominence Innovators (Levels)



Note: This figure reports coefficient estimates from equation (8). The dependent variable $Y_{i,t}$ corresponds to the outcome indicated in the title of each panel. The blue (light blue) points present the estimated coefficient for $\gamma_{group,1}$ ($\gamma_{group,2}$) for the specific batch group, with vertical lines indicating the 95% confidence interval.

Figure A17: Firm Outcomes After Prominence: Pre-prominence Innovators (Changes)



Note: TThis figure reports coefficient estimates from equation (8). The dependent variable $Y_{i,t}$ corresponds to the outcome indicated in the title of each panel. The blue (light blue) points present the estimated coefficient for $\gamma_{group,1}$ ($\gamma_{group,2}$) for the specific batch group, with vertical lines indicating the 95% confidence interval.

F Supplementary Tables

Table A1: Prominent Technologies with the Earliest and Most Recent Initiation Years

Panel A: Ten Prominent Technologies with the Earliest Initiation Years

Technology	Initiation	Emergence	Prominence
automotive vehicle	1920	1923	1932
cellulose ether composition	1920	1923	1923
combine typewrite	1920	1920	1920
handle system	1920	1933	1973
heat transfer	1920	1927	1966
antenna system	1921	1927	1946
electron tube	1921	1925	1947
froth flotation	1921	1931	1947
heat resistant	1921	1950	1982
land gear	1921	1924	1945

Panel B: Ten Prominent Technologies with the Most Recent Initiation Years

Technology	Initiation	Emergence	Prominence
unman aerial vehicle	1992	2007	2023
virtual reality	1994	1995	2019
augment reality	1995	2004	2018
medium content	1998	2002	2016
computer readable record medium	1999	2000	2015
computer readable storage medium	1999	2001	2013
additive manufacture	2000	2013	2020
organic light emit display	2000	2004	2014
social network	2006	2008	2013
deep learn	2016	2017	2023

Note: Panel A lists the ten prominent technologies with the earliest initiation years; Panel B lists the ten with the most recent initiation years. Initiation is the grant year of the technology's first patent. "Emergence" is the first year with at least five granted patents, and "Prominence" is the first year the technology ranks among the annual top 100 by patent count.

Table A2: Top Pre-emergence Assignees

Panel A: Top Pre-emergence Assignees by Patent Count

Assignee	NumPatent
RCA CORP	214
INTL BUSINESS MACHINES CORP	204
AT&T CORP	179
GENERAL ELECTRIC CO	148
EASTMAN KODAK CO	106
PARAMOUNT GLOBAL	87
EXXON MOBIL CORP	83
DUPONT DE NEMOURS INC	72
HITACHI LTD	58
XEROX HOLDINGS CORP	57

Panel B: Top Pre-emergence Assignees by Technology Count

Assignee	NumTech
AT&T CORP	80
INTL BUSINESS MACHINES CORP	77
RCA CORP	72
GENERAL ELECTRIC CO	62
PARAMOUNT GLOBAL	48
ITT CORPORATION	33
EASTMAN KODAK CO	27
HITACHI LTD	26
EXXON MOBIL CORP	24
SPERRY CORP	23

Note: This table presents leading pre-emergence innovators across 289 prominent technologies in our sample. Panel A and Panel B show the top ten pre-emergence innovators by counts of pre-emergence patents and technologies, respectively.

Table A3: Average Patent Quality by Batch Group

	Batch 0	Batch 1-5	Batch 6-10	Batch 11-50	Batch 51+
Avg. $Citations$	5.08	3.75	3.44	3.28	3.20
Avg. $\mathbb{1}(Citations > 0)$	0.78	0.70	0.68	0.62	0.62
Avg. $\mathbb{1}(Breakthrough)$	0.14	0.08	0.05	0.05	0.06
Avg. $Originality$	0.77	0.78	0.79	0.78	0.79

Note: This table presents average patent quality across batches over technology life cycle measured by patent batches.

Table A4: Early Innovators' Patent Shares Across Technology Life Cycle: Initiation Period

	(1)	(2)	(3)	(4)	(5)	(6)
$\text{Assignee}^{\text{Large}}_{\text{PreEmergence}} \times \mathbb{1}(\text{Initiation} \geq 1980)$	0.005	0.010	0.005	0.011	0.009	0.013*
	(0.011)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
$\text{Assignee}^{\text{Large}}_{\text{PreEmergence}} \times \text{Batch} \times \mathbb{1}(\text{Initiation} \geq 1980)$		−0.027***		−0.025***		−0.020**
		(0.010)		(0.008)		(0.009)
$\text{Assignee}^{\text{Large}}_{\text{PreEmergence}}$	0.018***	0.021***	0.012***	0.014***	0.010***	0.012***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$\text{Assignee}^{\text{Large}}_{\text{PreEmergence}} \times \text{Batch}$		−0.003***		−0.003***		−0.002***
		(0.000)		(0.000)		(0.000)
$\text{Assignee}^{\text{Non-Large}}_{\text{PreEmergence}} \times \mathbb{1}(\text{Initiation} \geq 1980)$	0.002	−0.001	0.000	−0.002	−0.002	−0.008
	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.008)
$\text{Assignee}^{\text{Non-Large}}_{\text{PreEmergence}} \times \text{Batch} \times \mathbb{1}(\text{Initiation} \geq 1980)$		0.094***		0.044		0.178
		(0.027)		(0.065)		(0.110)
$\text{Assignee}^{\text{Non-Large}}_{\text{PreEmergence}}$	0.004*	0.005**	0.008**	0.010***	0.009**	0.012***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)
$\text{Assignee}^{\text{Non-Large}}_{\text{PreEmergence}} \times \text{Batch}$		−0.003***		−0.004***		−0.005***
		(0.001)		(0.001)		(0.002)
Technology × Batch	X	X	X	X	X	X
Assignee			X	X		
Assignee × Year					X	X
Obs	1,169,156	1,169,156	1,123,417	1,123,417	1,032,865	1,032,865
Adj. R^2	0.115	0.116	0.251	0.252	0.277	0.278

Note: The result from estimating equation (3), with independent variables interacted with an indicator variable $\mathbb{I}(Initiation \geq 1980)$, which equals to one for technologies whose first patent is granted in 1980 or later. Each observation is at assignee-technology-batch level. The dependent variable $PatentShare$, is the fraction of patents in the batch that are assigned to a given assignee. $Assignee^{\text{Large}}_{\text{PreEmergence}}$ is an indicator for pre-emergence innovators that fall in the top 1% of the patent-count distribution in any year during the technology's pre-emergence period. $Assignee^{\text{NonLarge}}_{\text{PreEmergence}}$ is an indicator for pre-emergence innovators that are outside the top 1% of the patent-count distribution in all years during the technology's pre-emergence period. $Batch$ is the technology batch to which the patent is classified, scaled by 100. We include technology-batch fixed effects and assignee–batch start-year fixed effects. Standard errors are double-clustered at the technology and firm levels. *** denotes p-value < 0.01, ** denotes p-value< 0.05, and * denotes p-value< 0.1.

Table A5: Outcomes of Pre-emergence and Pre-prominence Innovators Before Prominence: North American Firms

Panel A.	(1) $\frac{CapEx_t}{PPENT_{t-1}}$	(2) $\frac{R\&DExp_t}{Asset_{t-1}}$	(3) $log(\frac{Asset_{t+1}}{Asset_t})$	(4) $log(\frac{Sales_{t+1}}{Sales_t})$	(5) $log(\frac{Empl_{t+1}}{Empl_t})$	(6) ROA_t	(7) $Margin_t$	(8) $ROIC_t$	(9) log(Asset M/B$_t$)	(10) $log(P/E_t)$
Assignee$_{\text{PreEmergence}}$ × PreProminence × PreEmergence	0.033***	0.003	0.022***	0.022***	0.016***	0.001	0.012	−0.001	0.082*	0.096**
	(0.012)	(0.002)	(0.007)	(0.006)	(0.004)	(0.002)	(0.011)	(0.006)	(0.044)	(0.041)
Assignee$_{\text{PreEmergence}}$ × PreProminence	0.008	0.003**	0.020***	0.015***	0.009***	0.003	0.008	0.006	0.055**	0.011
	(0.006)	(0.002)	(0.004)	(0.004)	(0.003)	(0.002)	(0.006)	(0.005)	(0.028)	(0.026)
Assignee$_{\text{PreEmergence}}$	−0.010***	−0.002**	−0.012***	−0.010***	−0.007***	−0.001	−0.006	−0.001	−0.034***	−0.021
	(0.004)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.004)	(0.002)	(0.012)	(0.014)
Assignee$_{\text{PreProminence}}$ × PreProminence × PreEmergence	0.020***	0.000	0.014***	0.017***	0.014***	0.002	0.006	−0.003	0.077***	0.082***
	(0.006)	(0.001)	(0.003)	(0.003)	(0.003)	(0.002)	(0.004)	(0.004)	(0.026)	(0.021)
Assignee$_{\text{PreProminence}}$ × PreProminence	0.017***	0.006***	0.019***	0.015***	0.012***	0.003**	0.003	0.006*	0.057***	0.033***
	(0.004)	(0.001)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.017)	(0.013)
Assignee$_{\text{PreProminence}}$	−0.011***	−0.003***	−0.010***	−0.009***	−0.007***	−0.002**	−0.002	−0.002	−0.029***	−0.022***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.009)	(0.006)
Log Asset	−0.144***	−0.026***	−0.095***	−0.083***	−0.056***	0.003	0.015*	0.009	−0.145***	−0.045
	(0.019)	(0.005)	(0.010)	(0.008)	(0.007)	(0.004)	(0.009)	(0.012)	(0.045)	(0.047)
Log Asset^2	0.005***	0.001***	0.002***	0.002***	0.001**	−0.001**	0.000	−0.002***	0.001	−0.005
	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.004)	(0.003)
Log Age	−0.561***	−0.097***	−0.259***	−0.206***	−0.148***	−0.006	−0.008	−0.062*	−0.889***	−0.642***
	(0.067)	(0.011)	(0.032)	(0.029)	(0.027)	(0.014)	(0.023)	(0.037)	(0.165)	(0.187)
Log Age^2	0.119***	0.029***	0.073***	0.052***	0.032***	0.001	−0.008	0.008	0.198***	0.155***
	(0.018)	(0.004)	(0.010)	(0.009)	(0.008)	(0.005)	(0.008)	(0.012)	(0.055)	(0.058)
Assignee	X	X	X	X	X	X	X	X	X	X
Technology × Year	X	X	X	X	X	X	X	X	X	X
Obs	1,002,959	1,024,947	1,009,188	1,007,080	983,081	1,022,390	1,019,562	898,703	921,538	786,243
Adj. R^2	0.415	0.716	0.251	0.261	0.236	0.413	0.586	0.396	0.704	0.636

Panel B.	(1) $\Delta\frac{CapEx_t}{PPENT_{t-1}}$	(2) $\Delta\frac{R\&DExp_t}{Asset_{t-1}}$	(3) $\Delta log(\frac{Asset_{t+1}}{Asset_t})$	(4) $\Delta log(\frac{Sales_{t+1}}{Sales_t})$	(5) $\Delta log(\frac{Empl_{t+1}}{Empl_t})$	(6) ΔROA_{t+1}	(7) $\Delta Margin_{t+1}$	(8) $\Delta ROIC_{t+1}$	(9) $\Delta log(\text{Asset M/B}_t)$	(10) $\Delta log(P/E_t)$
$\text{Assignee}_{\text{PreEmergence}} \times \text{PreProminence} \times \text{PreEmergence}$	−0.005*	−0.001**	0.001	−0.001	0.001	−0.001	0.001	−0.001	−0.002	−0.013
	(0.003)	(0.000)	(0.005)	(0.004)	(0.004)	(0.001)	(0.001)	(0.001)	(0.006)	(0.008)
$\text{Assignee}_{\text{PreEmergence}} \times \text{PreProminence}$	0.012***	0.001***	0.027***	0.017***	0.016***	0.001***	0.002***	0.001	0.037***	0.019*
	(0.002)	(0.000)	(0.004)	(0.002)	(0.003)	(0.000)	(0.001)	(0.001)	(0.006)	(0.011)
$\text{Assignee}_{\text{PreEmergence}}$	−0.004***	0.000***	−0.009***	−0.006***	−0.006***	−0.001***	−0.001***	0.000	−0.012***	−0.005
	(0.001)	(0.000)	(0.002)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.003)	(0.004)
$\text{Assignee}_{\text{PreProminence}} \times \text{PreProminence} \times \text{PreEmergence}$	−0.007***	−0.001***	−0.010***	−0.007***	−0.004*	0.000	0.000	0.000	−0.012**	−0.014**
	(0.002)	(0.000)	(0.003)	(0.002)	(0.002)	(0.000)	(0.001)	(0.001)	(0.005)	(0.006)
$\text{Assignee}_{\text{PreProminence}} \times \text{PreProminence}$	0.008***	0.001***	0.021***	0.014***	0.012***	0.001**	0.002***	0.001	0.027***	0.012***
	(0.002)	(0.000)	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)	(0.001)	(0.004)	(0.005)
$\text{Assignee}_{\text{PreProminence}}$	−0.002**	0.000***	−0.007***	−0.004***	−0.004***	0.000*	−0.001***	0.000	−0.008***	−0.003*
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.002)	(0.002)
Log Asset	−0.051***	−0.007***	−0.116***	−0.063***	−0.066***	−0.004**	−0.009***	−0.007	−0.205***	−0.066***
	(0.015)	(0.002)	(0.012)	(0.008)	(0.008)	(0.002)	(0.002)	(0.005)	(0.019)	(0.020)
Log Asset^2	0.002***	0.000***	0.003***	0.001**	0.001***	0.000**	0.000*	0.000	0.006***	0.002
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Log Age	0.375***	0.067***	0.437***	0.320***	0.115***	0.009	−0.009	0.003	0.645***	0.531***
	(0.103)	(0.010)	(0.054)	(0.048)	(0.043)	(0.007)	(0.007)	(0.019)	(0.085)	(0.151)
Log Age^2	−0.066**	−0.014***	−0.062***	−0.048***	0.001	−0.001	0.005**	0.003	−0.089***	−0.100***
	(0.026)	(0.002)	(0.014)	(0.012)	(0.011)	(0.002)	(0.002)	(0.006)	(0.022)	(0.038)
Assignee	X	X	X	X	X	X	X	X	X	X
Technology × Year	X	X	X	X	X	X	X	X	X	X
Obs	973,366	997,083	982,015	979,722	938,601	1,007,106	1,002,265	876,892	891,273	709,994
Adj. R^2	0.072	0.094	0.124	0.114	0.091	0.075	0.098	0.066	0.209	0.131

Note: This table reports estimates from equation (2) for North American subsample. The unit of observation is the firm–technology–year. All dependent variables are defined in Appendix A. The key explanatory variables identify early innovators: $Assignee_{\text{PreEmergence}}$ is an indicator equal to one if the firm's first patent in the technology is granted on or before the technology's emergence year; $Assignee_{\text{PreProminence}}$ is an indicator equal to one if the firm's first patent in the technology is granted after emergence but on or before the technology's prominence year. These variables are interacted with indicators *PreEmergence* and *PreProminence*, which equal one if the observation year is on or before the technology's emergence and prominence, respectively. We control for firm size and age, and include firm fixed effects and technology-by-year fixed effects. Standard errors are double-clustered at the technology and firm levels. *** denotes p-value < 0.01, ** denotes p-value< 0.05, and * denotes p-value< 0.1.

Table A6: Outcomes of Pre-emergence and Pre-prominence Innovators After Prominence: North American Firms

Panel A.	(1) $\frac{CapEx_t}{PPENT_{t-1}}$	(2) $\frac{R\&DExp_t}{Asset_{t-1}}$	(3) $log(\frac{Asset_{t+1}}{Asset_t})$	(4) $log(\frac{Sales_{t+1}}{Sales_t})$	(5) $log(\frac{Empl_{t+1}}{Empl_t})$	(6) ROA_t	(7) $Margin_t$	(8) $ROIC_t$	(9) log(Asset M/B$_t$)	(10) $log(P/E_t)$
Assignee$_{\text{PreEmergence}}$ × Batch$_{1-5}$	−0.016**	−0.004***	−0.023***	−0.019***	−0.012***	−0.003	−0.009	−0.004	−0.055**	−0.027
	(0.007)	(0.002)	(0.005)	(0.004)	(0.003)	(0.002)	(0.008)	(0.005)	(0.026)	(0.029)
Assignee$_{\text{PreEmergence}}$ × Batch$_{6-20}$	−0.024***	−0.005**	−0.033***	−0.025***	−0.016***	−0.003	−0.014	−0.006	−0.096**	−0.046
	(0.009)	(0.002)	(0.006)	(0.005)	(0.004)	(0.003)	(0.010)	(0.006)	(0.038)	(0.043)
Assignee$_{\text{PreEmergence}}$ × Batch$_{21+}$	−0.029**	−0.006*	−0.035***	−0.034***	−0.024***	−0.006*	−0.024*	−0.012	−0.158***	−0.083
	(0.012)	(0.003)	(0.009)	(0.008)	(0.006)	(0.004)	(0.012)	(0.008)	(0.052)	(0.053)
Assignee$_{\text{PreEmergence}}$	0.009**	0.003**	0.016***	0.012***	0.008***	0.002	0.006	0.004	0.045**	0.018
	(0.004)	(0.001)	(0.003)	(0.003)	(0.002)	(0.001)	(0.005)	(0.003)	(0.019)	(0.020)
Assignee$_{\text{PreProminence}}$ × Batch$_{1-5}$	−0.022***	−0.006***	−0.023***	−0.019***	−0.016***	−0.004***	−0.004	−0.006*	−0.071***	−0.044***
	(0.005)	(0.001)	(0.003)	(0.003)	(0.002)	(0.001)	(0.003)	(0.003)	(0.017)	(0.014)
Assignee$_{\text{PreProminence}}$ × Batch$_{6-20}$	−0.022***	−0.006***	−0.022***	−0.017***	−0.014***	−0.003	−0.003	−0.002	−0.070**	−0.062***
	(0.006)	(0.002)	(0.004)	(0.004)	(0.003)	(0.002)	(0.005)	(0.005)	(0.027)	(0.019)
Assignee$_{\text{PreProminence}}$ × Batch$_{21+}$	−0.024***	−0.007***	−0.030***	−0.025***	−0.020***	−0.005	−0.007	−0.009	−0.102**	−0.074**
	(0.007)	(0.002)	(0.007)	(0.005)	(0.005)	(0.004)	(0.008)	(0.009)	(0.050)	(0.037)
Assignee$_{\text{PreProminence}}$	0.012***	0.004***	0.013***	0.011***	0.009***	0.002**	0.003	0.004*	0.046***	0.030***
	(0.003)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.012)	(0.009)
Log Asset	−0.144***	−0.026***	−0.095***	−0.083***	−0.056***	0.003	0.015*	0.009	−0.146***	−0.045
	(0.019)	(0.005)	(0.010)	(0.008)	(0.007)	(0.004)	(0.009)	(0.012)	(0.045)	(0.047)
Log Asset^2	0.005***	0.001***	0.002***	0.002***	0.001**	−0.001**	0.000	−0.002***	0.001	−0.005
	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.004)	(0.003)
Log Age	−0.561***	−0.097***	−0.258***	−0.206***	−0.147***	−0.006	−0.008	−0.062*	−0.886***	−0.638***
	(0.067)	(0.011)	(0.032)	(0.029)	(0.027)	(0.014)	(0.023)	(0.037)	(0.164)	(0.187)
Log Age^2	0.119***	0.029***	0.073***	0.052***	0.032***	0.001	−0.008	0.008	0.197***	0.154***
	(0.018)	(0.004)	(0.010)	(0.009)	(0.008)	(0.005)	(0.008)	(0.012)	(0.054)	(0.058)
Assignee	X	X	X	X	X	X	X	X	X	X
Technology × Year	X	X	X	X	X	X	X	X	X	X
Obs	1,002,959	1,024,947	1,009,188	1,007,080	983,081	1,022,390	1,019,562	898,703	921,538	786,243
Adj. R^2	0.415	0.716	0.251	0.261	0.236	0.413	0.586	0.396	0.704	0.636

Panel B.	(1) $\Delta\frac{CapEx_t}{PPENT_{t-1}}$	(2) $\Delta\frac{R\&DExp_t}{Asset_{t-1}}$	(3) $\Delta log(\frac{Asset_{t+1}}{Asset_t})$	(4) $\Delta log(\frac{Sales_{t+1}}{Sales_t})$	(5) $\Delta log(\frac{Empl_{t+1}}{Empl_t})$	(6) ΔROA_{t+1}	(7) $\Delta Margin_{t+1}$	(8) $\Delta ROIC_{t+1}$	(9) $\Delta log(\text{Asset M/B}_t)$	(10) $\Delta log(P/E_t)$
$\text{Assignee}_{\text{PreEmergence}} \times \text{Batch}_{1-5}$	−0.009***	−0.001***	−0.024***	−0.014***	−0.013***	−0.001***	−0.002***	−0.001	−0.030***	−0.013*
	(0.002)	(0.000)	(0.004)	(0.003)	(0.003)	(0.000)	(0.001)	(0.001)	(0.006)	(0.008)
$\text{Assignee}_{\text{PreEmergence}} \times \text{Batch}_{6-20}$	−0.012***	−0.001***	−0.032***	−0.020***	−0.021***	−0.001**	−0.003***	0.001	−0.042***	−0.021
	(0.003)	(0.000)	(0.005)	(0.003)	(0.004)	(0.001)	(0.001)	(0.001)	(0.007)	(0.015)
$\text{Assignee}_{\text{PreEmergence}} \times \text{Batch}_{21+}$	−0.013***	−0.002***	−0.041***	−0.024***	−0.023***	−0.002**	−0.004***	0.001	−0.059***	−0.013
	(0.003)	(0.000)	(0.009)	(0.005)	(0.005)	(0.001)	(0.001)	(0.002)	(0.013)	(0.024)
$\text{Assignee}_{\text{PreEmergence}}$	0.007***	0.001***	0.019***	0.011***	0.011***	0.001***	0.002***	0.000	0.025***	0.011*
	(0.002)	(0.000)	(0.003)	(0.002)	(0.002)	(0.000)	(0.000)	(0.001)	(0.004)	(0.006)
$\text{Assignee}_{\text{PreProminence}} \times \text{Batch}_{1-5}$	−0.004**	−0.001**	−0.015***	−0.009***	−0.009***	−0.001*	−0.001***	−0.001	−0.019***	−0.005
	(0.002)	(0.000)	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)	(0.001)	(0.004)	(0.004)
$\text{Assignee}_{\text{PreProminence}} \times \text{Batch}_{6-20}$	−0.010***	−0.002***	−0.024***	−0.016***	−0.013***	−0.001	−0.002***	0.000	−0.034***	−0.016**
	(0.002)	(0.000)	(0.004)	(0.003)	(0.002)	(0.000)	(0.000)	(0.001)	(0.005)	(0.007)
$\text{Assignee}_{\text{PreProminence}} \times \text{Batch}_{21+}$	−0.010***	−0.002***	−0.037***	−0.024***	−0.020***	−0.001	−0.002***	−0.003	−0.047***	−0.032**
	(0.003)	(0.000)	(0.005)	(0.004)	(0.004)	(0.001)	(0.001)	(0.002)	(0.008)	(0.015)
$\text{Assignee}_{\text{PreProminence}}$	0.004***	0.001***	0.012***	0.008***	0.007***	0.000**	0.001***	0.000	0.016***	0.006**
	(0.001)	(0.000)	(0.002)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.003)	(0.003)
Log Asset	−0.051***	−0.007***	−0.116***	−0.063***	−0.066***	−0.004**	−0.009***	−0.007	−0.205***	−0.066***
	(0.015)	(0.002)	(0.012)	(0.008)	(0.008)	(0.002)	(0.002)	(0.005)	(0.019)	(0.020)
Log Asset^2	0.002***	0.000***	0.003***	0.001**	0.001***	0.000**	0.000*	0.000	0.006***	0.002
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Log Age	0.375***	0.067***	0.440***	0.321***	0.116***	0.009	−0.009	0.003	0.649***	0.534***
	(0.103)	(0.010)	(0.054)	(0.048)	(0.043)	(0.007)	(0.007)	(0.019)	(0.085)	(0.151)
Log Age^2	−0.066**	−0.014***	−0.063***	−0.048***	0.000	−0.001	0.005**	0.003	−0.090***	−0.100***
	(0.026)	(0.002)	(0.014)	(0.012)	(0.011)	(0.002)	(0.002)	(0.006)	(0.022)	(0.038)
Assignee	X	X	X	X	X	X	X	X	X	X
Technology × Year	X	X	X	X	X	X	X	X	X	X
Obs	973,366	997,083	982,015	979,722	938,601	1,007,106	1,002,265	876,892	891,273	709,994
Adj. R^2	0.072	0.094	0.124	0.114	0.091	0.075	0.098	0.066	0.209	0.132

Note: This table reports estimates from equation (4) for North American subsample. The unit of observation is the firm–technology–year. All dependent variables are defined in Appendix A. The key explanatory variables identify early entrants: $Assignee_{\text{PreEmergence}}$ is an indicator equal to one if the firm's first patent in the technology is granted on or before the technology's emergence year; $Assignee_{\text{PreProminence}}$ is an indicator equal to one if the firm's first patent in the technology is granted after emergence but on or before the technology's prominence year. These variables are interacted with lifecycle indicators $Batch_{1-5}$, $Batch_{6-20}$, and $Batch_{21+}$, which equal one if the observation year corresponds to the respective batch groups of the technology. We control for firm size and age, and include firm fixed effects and technology-by-year fixed effects. Standard errors are double-clustered at the technology and firm levels. *** denotes p-value < 0.01, ** denotes p-value< 0.05, and * denotes p-value< 0.1.