

Patents' Long-term Potentials

Chen Yang

Ying Jiang

Yuxin Zhang*

January 2026

Abstract

This paper discovers the long-term value of patents using a newly constructed indicator of technological potential. The index combines a patent's initial dormant nature (Cohen et al., 2024) with its subsequent activation by technology trends. Specifically, it rewards patents in niche technology spaces that align with rising technological trend after their grant. By using news topics as a proxy for technological trends and leveraging USPTO patent records integrated with financial data from 1976 to 2023, we develop a novel index to identify high-potential patents. At the patent level, high-potential patents exhibit accelerating forward citations over a 20-year horizon and a higher probability of being traded in the patent market. The index is distinct from existing patent metrics (e.g., KPSS and InnOrig) and can identify "hidden gems" patents that outperform in the long run despite receiving low KPSS valuations. At the firm level, we find that a company's portfolio of high-potential patents predicts higher future profitability and reduced idiosyncratic volatility. Our findings are robust across different industries and time periods.

Keywords: Patent Valuation, Late-Blooming Patents, Innovation, Firm Performance, Intellectual Property, Patent Transactions

JEL Codes: G14, G32, O31, O34

*University of Nottingham Ningbo China, Nottingham University Business School China. Emails: {Chen.Yang, Ying.Jiang, Yuxin.Zhang}@nottingham.edu.cn.

1 Introduction

Patents represent one of the most important intangible outputs of innovation, yet their economic value often follows unpredictable temporal patterns that challenge conventional valuation approaches. While some patents generate immediate market recognition and commercial success, others remain dormant for years before revealing their true economic and technological potential (Cohen et al., 2024). This phenomenon of late-blooming patents suggests that traditional patent valuation methods, which emphasize immediate market reactions and short-term indicators, may underestimate the long-term value embedded in patents. Understanding the mechanisms that drive delayed patent value realization has implications for intellectual property valuation, assessment for innovation-intensive company, and long-term technological risk (Baruffaldi and Simeth, 2020).

Traditional patent valuation centers on immediate market impact and short-term indicators of private and scientific worth. For private value, Kogan et al. (2017) employs the event study method with short event windows to demonstrate correlations between patent grants and contemporaneous stock price movements. Scientific value is typically measured through forward citations, originality, and generality. Among these metrics, however, only forward citations operate on an extended time horizon, accumulating up to 30 years post-grant to fully capture an invention’s lifetime impact (Trajtenberg, 1990; Hall et al., 2001, 2005). By contrast, originality (constructed from backward citations observed at grant) and generality (typically computed from the distribution of early forward citations within limited post-grant windows) are implemented as short-horizon indicators (Hall et al., 2001; Hirshleifer et al., 2018). While a body of literature including Cohen et al. (2013); Hirshleifer et al. (2013, 2018) has examined patent value and innovation outcomes, a significant gap exists in understanding the long-term evolution of patent value and its determinants. The emerging evidence of late-blooming patents - inventions that attract limited early attention but later generate significant technological and economic impact (Cohen et al., 2024), suggests that

traditional valuation approaches overlooks a distinction between an immediate "hit" and a "late-bloomer."

We bridge this gap by proposing a novel index that links patent value realization to the interaction between patent's dormant nature and technological shifts. In our framework, the initial dormancy of an invention is a rational response to high search and integration costs (Chamley and Gale, 1994; Cohen et al., 2024; Glode and Ordóñez, 2025). This dormancy ends when shifting technological trends reduce such frictions, facilitating the patent's subsequent activation.

To capture dormancy component, we build our measure on the strand of literature in delayed recognition of innovation (Teixeira et al., 2017; Van Raan, 2017; Cohen et al., 2024). A dormant patent is an invention that experiences an extended "sleeping period" following its grant, during which it is initially overlooked or neglected. Using the framework of Cohen et al. (2024), such patents are distinguished by two key characteristics: 1) *technological non-centrality* (measured by *Tech-class weight*) as they fall outside the inventor's core domain, and 2) *low competitive intensity* (measured by *Competition*), as they operate in sparsely populated technological fields. By deviating from established trajectories, these patents are less immediately exploitable. Cohen et al. (2024) interpret this late-blooming phenomenon as a rational "allocation-and-matching" process, where firms strategically decide whether to internally "write" or externally "hunt" for patents based on their core capabilities. Consistent with models of strategic delay (Chamley and Gale, 1994) and surplus appropriation (Glode and Ordóñez, 2025), this framework suggests that rational agents may initially neglect high-potential patents when the costs of search and integration are high. Consequently, these late-bloomers are identified and built upon only when the expected returns from commercialization exceed these specific search costs.

However, we argue that such ex-ante dormant nature alone is insufficient for value realization; a patent's potential must be unlocked by a catalyst of public recognition. This

aligns with the 'Sleeping Beauty' literature, where radical innovations remain dormant for years before a sudden surge in recognition, often catalyzed by shifting technological trends (Van Raan, 2004, 2017; Hou and Yang, 2019). To capture this, we integrate a technology salience component into our index, defined as the surge in public attention following the grant. Specifically, we track the monthly attention of technology-related topics in the Wall Street Journal by mapping patent Cooperative Patent Classification (CPC) to the news topics from Bybee et al. (2024). Our reliance on news attention is theoretically grounded in its dual role as both a macro-level signal and a micro-level friction reducer. First, following Alexopoulos (2011), we utilize public discourse as a measure for realized technological trends as it captures the commercial demand for innovation that raw patent counts fail to reflect. Second, consistent with Fang and Peress (2009), we argue that the surge in media coverage resolves information asymmetry. By directing limited investor attention to previously niche domains, this coverage eliminates the 'information discount' associated with dormant assets. Methodologically, this novel linkage between technological classifications and market discourse ensures that our index captures not just the existence of a technology, but the public's recognition to its commercial potential.

The *Potential* index is defined as the news attention received by a patent in the five years post-grant, net of its pre-grant news attention, *Tech-class weight*, and *Competition* level. This construction relies on a temporal separation to identify the late-blooming patents. First, the dormancy components (pre-grant news attention, *Tech-class weight* and *Competition*) are measured ex-ante at the moment of patent grant. By entering with negative signs, these terms reward patents in peripheral and uncrowded domains that received little attention prior to issuance. Second, the technology salience component captures the ex-post realization. The *Potential* index thus identifies patents whose dormant nature is unlocked by evolving technological trends.

The realization of long-term patent value is typically manifested through two channels: sci-

entific impact and market monetization. Prior research establishes accumulated forward citations as the primary proxy for scientific influence (Trajtenberg, 1990; Hall et al., 2005), while patent transactions reflect the recognition of economic worth in the ‘market for ideas’ (Arora et al., 2004; Serrano, 2010). If our *Potential* index effectively captures latent quality, we expect high-scoring patents to demonstrate superior performance in both sustained citation growth and market tradability.

To test our hypotheses, we integrate United States Patent and Trademark Office (USPTO) patent records and the KPSS patent-value database (Kogan et al., 2017) with financial information from Compustat and Center for Research in Security Prices (CRSP). Our final sample spans the 1976–2023 period and encompasses 1.4 million granted patents. Our technology salience measure leverages the news-topic attention data from Bybee et al. (2024). The empirical analysis yields several key findings into the realization of long-term innovation value. At the patent level, we identify two realization pathways. First, the index serves as a powerful predictor of scientific impact; the citation advantage of high-potential patents over their peers more than doubles as the time horizon extends from the first 5 years to 20 years post-grant. Crucially, we find that this measure provides incremental predictive power when compared with KPSS (Kogan et al., 2017) and innovation originality (InnOrig) (Hirshleifer et al., 2018). It successfully identifies ‘hidden gems’—patents with initially low-KPSS values that eventually surpass high-KPSS patents in long-term scientific impact. Second, we find that long-term potential enhances the probability of patent trade, suggesting that the market for technology effectively recognizes and monetizes high-quality assets as their latent value is validated. Aggregating these characteristics to the firm level, we find that higher *Potential* is positively associated with increased profitability for up to two decades. Moreover, these high-potential assets significantly reduce firm’s idiosyncratic volatility (IVOL). These findings are robust across diverse industrial sectors and different sample periods.

Our findings contribute to several streams of literature in innovation economics and corpo-

rate finance. First, we advance innovation literature by identifying patent characteristics that predict long-term scientific impact and patent transactions. We distinguish our measure from established innovation metrics, demonstrating that long-term technological value is overlooked by immediate market recognition. We document a "scissors effect" in value realization: whereas the predictive power of KPSS decays in the extended period, our index strengthens and becomes the dominant predictor of scientific impact in the long run. Also, our metric exhibits advantage in value discovery by identifying "hidden gems"—patents initially have relatively lower KPSS that eventually outperform even the KPSS’s highest-valued "stars." Second, we contribute to the growing literature on innovation efficiency and firm performance by identifying a new characteristic of innovation output that predicts sustained future firm profitability. Third, we establish a novel methodological linkage between patent taxonomies and business news topics, marking the first attempt to align technical categories with public narratives. This approach allows us to identify the specific early signals of technology trends.

While drawing inspiration from the late-blooming patterns identified by [Cohen et al. \(2024\)](#), our study introduces several key distinctions. First, [Cohen et al. \(2024\)](#) construct their measures (e.g., *tech-class-weight*) using the firm’s *entire* patent history over the full sample period, which introduces an ex-post bias in defining a patent’s initial state. In contrast, our index components (*Tech-class-weight* and *Competition*) are calculated strictly using information available at the moment of patent grant. This ensures that the dormancy aspect of our index is captured from a true ex-ante perspective. Second, we argue that dormancy measures reflects only latent potential, while the evolution of technological trends serves as the essential catalyst for its activation. Our technological salience component plays a key role in capturing this activation mechanism; by utilizing public attention as a proxy, it identifies the condition when a dormant domain transitions into a recognized growth phase.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature

and develops our hypotheses. Section 3 describes our data sources, variable construction, and summary statistics. Section 4 presents our main empirical results. Section 5 explains robustness checks. Finally, Section 6 concludes the paper.

2 Literature Review and Hypothesis Development

2.1 Measurement of patent value and its limitations

Patents serve as a core output of innovation, influencing a company’s market value and long-term economic performance (Hall et al., 2005). Existing literature has explored patent value from two main perspectives, including private value (Kogan et al., 2017), and scientific value (Hirshleifer et al., 2018; Kelly et al., 2021)

Numerous studies have attempted to measure and estimate patents’ private economic value using various methodologies. Among these approaches, the method developed by Kogan et al. (2017) (KPSS) has become the most widely adopted. KPSS estimates patent value by analyzing excess stock returns at the time of patent issuance, reflecting investors’ assessment of the patent’s worth. Beyond the KPSS measure, other approaches include direct surveys of patent holders (de Rassenfosse and Jaffe, 2018) or analysis of renewal decisions (Bessen, 2008; Hughes, 2024). However, much like KPSS, these methods are less effective at capturing the potential of patents whose value is not yet revealed or emerges gradually over time (Arora et al., 2024).

Patents’ scientific value provides another perspective on their worth. This approach is based on the view that more frequently cited patents have greater value (Griliches, 1958; Trajtenberg, 1990). Building on this foundation, Hall et al. (2001, 2005) introduced more sophisticated citation-based metrics, such as originality and generality, to establish that future citations reflect a patent’s impact on technological evolution. Moving beyond these citation patterns to analyze textual content, Kelly et al. (2021) define a patent’s significance

score through textual similarity comparisons with both prior and subsequent patents. They revealed a pattern that significant patents with high scientific value typically exhibit low textual similarity to previous patents but high textual similarity to subsequent patents. While powerful, it remains a retrospective metric of technical follow-up, as it requires the actual emergence of subsequent patents to validate a breakthrough.

2.2 Late-blooming patents: theoretical framework for long-term value

2.2.1 Dormancy of late-blooming patents

Existing literature has examined patent value from various perspectives, while discussions regarding the long-term value of patents remain limited. The long-term potential of a patent refers to its capacity to generate significant influence, economic value, or technological impact over an extended period after its grant ([Harhoff et al., 2003](#)). [Cohen et al. \(2024\)](#) analyzes USPTO patents issued from 1976 to 1999 and shows that many late-blooming patents reach the top 5% of citations within their CPC class after an unusually long delay (≈ 14 years). These patents become at least as influential as, and on average more influential than, those recognized early. They also identify the key organizational actors, such as patent hunters, who can spot and develop such overlooked ideas. For patent hunters, utilizing (i.e., by citing and building upon) late-blooming patents yields economically sizable gains: a 6.4% rise in sales growth, a 2.2% increase in Tobin’s Q, and a 2.2% increase in new-product introductions.

[Cohen et al. \(2024\)](#) finds that these late-blooming patents share distinct characteristics: they often lie far from the original patent writers’ core technological domains and typically exist in less competitive technological spaces with fewer contemporaneous patents. Specifically, late-bloomer status is negatively associated with tech-class-weight and negatively associated with competing patent stock. They interpret late blooming as a rational allocation-and-matching phenomenon, in which firms decide whether to internally ‘write’ or externally ‘hunt’ for

a patent based on how well the technology matches their core capabilities. Their "write-or-hunt" framework rooted in [Weitzman \(1998\)](#) concept of recombinant growth, viewing innovation as the reconfiguration of existing knowledge. Consistent with models of strategic delay ([Chamley and Gale, 1994](#)) and surplus appropriation ([Glode and Ordonez, 2025](#)), [Cohen et al. \(2024\)](#) suggest that rational agents may initially overlook or neglect high-potential patents when the costs of search and integration are high. Under this approach, these "late-bloomer" patents are identified and built upon only when the expected returns from commercialization exceed specific search costs.

2.2.2 Technological salience: a catalyst for value activation

While the search-and-matching framework of [Cohen et al. \(2024\)](#) explains the rational incentives of firm-level "hunters", this supply-side focus (on developers and "patent hunters") overlooks an alternative driver: the technology trends. As argued by [Hou and Yang \(2019\)](#), a patent may initially lack immediate applications or complementary technologies, but its latent value can be unlocked over time, coinciding with the growth phase of its corresponding technology cycle. We define this value activation process as the technological transition where a patent's underlying domain moves from a dormant niche into a prominent growth cycle. The key challenge lies in how to measure these macro-level technology shifts. [Alexopoulos \(2011\)](#) suggests that indicators derived from the title of publications can track the commercialization and diffusion of new technologies more accurately than R&D expenditures or patent counts. This finding implies that public discourse can capture the technological trends. In addition, [Fang and Peress \(2009\)](#) shows that assets with low media coverage suffer from an 'information discount' because investors demand a premium to compensate for the higher information uncertainty. In the context of late-blooming patents, the initial dormancy can be viewed as a period of information asymmetry where search costs prohibit value discovery ([Cohen et al., 2024](#)). When the underlying technology trend gains prominence in the news, it acts as a mechanism that resolves these frictions.

Taken together, news attention emerges as the ideal proxy for technology salience. On one hand, consistent with [Alexopoulos \(2011\)](#), it reflects the macro-level migration of technological hot-spots; on the other hand, aligned with [Fang and Peress \(2009\)](#), the surge in attention help resolve information asymmetry, driving the value realization of late-blooming patents. [Bybee et al. \(2024\)](#) employ topic modeling (Latent Dirichlet Allocation) on the full text of the Wall Street Journal. They identify 180 topics and tracks the monthly attention given to each. By identifying the technology-related themes within this set, we move beyond simple keyword counts to measure the salience of technological concepts in public and financial discourse. The elongated value realization of a late-blooming patent is therefore triggered when the patent’s underlying technological theme (as proxied by its associated news topics) matures and gains prominence in public discourse.

2.3 Realization of Long-Term Patent Value

Scientific Value (Forward Citations) - A primary channel through which a patent’s latent potential is realized over the long term is its scientific influence on subsequent innovation, a process observable through the accumulation of forward citations. When a patent introduces a valuable idea, other inventors and firms adopt, build upon, and extend it, creating a trail of new inventions that cite the original work ([Hall et al., 2005](#); [Kogan et al., 2017](#)). Unlike the private value measure (KPSS), scientific value measured by citations inherently unfolds over a long horizon. Late-blooming patents exhibit a distinct pattern of sustained or accelerating citation growth over extended periods. This phenomenon stems from their nature as foundational or cross-disciplinary innovations, which often require time for the scientific community to fully recognize their applicability ([Van Raan, 2017](#); [Cohen et al., 2024](#)). This temporal dimension also aligns with the [Weitzman \(1998\)](#), where peripheral yet high-potential ideas serve as building blocks for future developments. Consequently, for a dormant patent to achieve this high-impact trajectory, it should possesses inherent latent quality and its underlying technological domain should transition into a growth phase to trig-

ger wider diffusion. We hypothesize that this process requires the conjunction of patent’s latent potential and the catalyst of technology salience to unlock its scientific value:

H1: Patent-level long-term potential (proxied by *Potential* index) is positively associated with future citation counts.

Patent’s Marketability - Prior literature has emphasized that patents are not only legal protections for innovation but also tradable assets whose value can be realized through market transactions ([Arora et al., 2004](#); [Serrano, 2010](#); [Galasso et al., 2013](#)). [Serrano \(2010\)](#) investigates the dynamics of patent transfers and renewals by extending the classic renewal framework of [Pakes and Schankerman \(1984\)](#) to incorporate the possibility of surplus-enhancing transfers. They find that patents are more likely to be traded when the expected gains from trade outweigh the costs of technology adoption. [Akcigit et al. \(2016\)](#) also highlights the role of patent transfers as a mechanism for technology owners to extract value from innovations that are not directly useful to them. In their endogenous growth model, firms invest in R&D but may generate ideas that are misaligned with their core technological capabilities. Rather than letting such ideas go unused, firms can sell the associated patents on the market.

We examine this market adoption pathway through our measure of patent potential. High potential implies that the patent represent foundational or breakthrough ideas but not in line with the holding firm’s core area; thus, the market demand for such assets is naturally higher. Whether driven by the originator’s desire to monetize a valuable asset or an acquirer’s intent to secure promising technology, patents with high potential are expected to be more frequently traded as their latent value is eventually recognized by the public. We therefore have the following hypothesis:

H2: Patents with higher long-term potential (proxied by *Potential* index) are positively associated with future patent transactional activities.

Innovation Activities and Corporate Value - Innovation economics theory suggests that innovation often results from ”recombinant search”—creating new products or processes by

combining different technological fields (Weitzman, 1998; Singh and Fleming, 2010; Schumpeter and Swedberg, 2021). Building on these literature, Hirshleifer et al. (2018) introduced the concept of innovation originality (InnOrig) to capture a company’s ability to integrate diverse technological knowledge. Unlike previous research that focused on innovation inputs (such as R&D), this study examines innovation output by measuring originality through the number of unique technology categories cited in patents. The underlying assumption is that a patent incorporating knowledge from many different technological fields is more original. Empirically, they found that companies with high originality exhibited higher and more stable profitability. Furthermore, the stocks of these highly original firms achieved significant excess returns (9-13% annualized). This research reinforces the view that innovation activities have a positive and tangible impact on corporate value.

Extending this idea to our framework, when patent-level characteristics are aggregated to the firm level, they represent the long-term resilience and latent quality of a company’s innovation portfolio. Under the resource-based view (Barney, 1991), a portfolio rich in patents with high long-term potential constitutes a strategic asset that builds competitive barriers. The delayed value realization mechanism suggests that firms accumulating these assets will experience gradual but sustained improvements in operational efficiency and financial returns as their portfolios mature (Chan et al., 2001; Hirshleifer et al., 2018). As the latent value of these innovations is unlocked and integrated into the market, we expect a translation into financial performance:

H3a: Firm-level aggregated long-term potential (proxied by *Potential* index) is positively associated with long-term profitability (ROE).

Complementing the profitability hypothesis, we conjecture that a portfolio of high-potential patents acts as a volatility dampener to mitigate the firm-specific risk (idiosyncratic volatility). While early-stage R&D inputs are typically associated with high uncertainty (Chan et al., 2001), the validated outputs captured by our index have the opposite effect. Drawing

on the Resource-Based View (RBV), assets that are rare and inimitable form the basis of sustained competitive advantage (Barney, 1991). Because our index selects patents originating in peripheral and non-competitive domains, these assets inherently face fewer direct rivals and have high uniqueness. Such unique assets create high barriers to entry, preventing the erosion of economic rents (Hirshleifer et al., 2018). These barriers serve as the source of market power, which functions as a "natural hedge" that allows firms to pass on cost shocks and stabilize cash flows (Peress, 2010). Consistent with Gaspar and Massa (2006) and Irvine and Pontiff (2009), who demonstrate that firms with strong market power exhibit significantly lower idiosyncratic volatility, we expect that the accumulation of these high-barrier assets will stabilize the firm's performance and reduce risk.

H3b: Firm-level aggregated long-term patent potential is negatively associated with future idiosyncratic volatility (IVOL).

3 Data and Variables

3.1 Data

Our sample comprises firms at the intersection of the Compustat, CRSP, and USPTO patent databases. Patent data are primarily sourced from the USPTO's PatentsView database and the KPSS Extended Data constructed by (Kogan et al., 2017). The PatentsView database provides CPC information for patents granted since 1976, along with citation relationships among patents and key structural variables, including the number of claims, citations, and inventors. Furthermore, we collect data from the USPTO Patent Assignment Dataset to track patent transactions. This dataset contains records of all publicly recorded patent assignments since 1980, identifying the assignor (seller) and assignee (buyer). To link patents to publicly traded firms, we utilize the KPSS Extended Data, accessed via Noah Stoffman's personal website, which matches patents to firms' PERMNO codes. Additionally, the KPSS Extended Data provides the market value of each patent. The initial collection includes

2,677,061 granted patents from 1976 to 2023. The requirement for non-missing information across all independent variables yields a final sample of 1,405,625 patents for our empirical analysis.

For technology trend measure, we utilize the news topic data developed by [Bybee et al. \(2024\)](#). This dataset is derived from a latent Dirichlet allocation (LDA) topic model estimated on the full text of approximately 800,000 Wall Street Journal articles spanning from 1984 to 2017. Following the authors’ optimal specification, the model identifies 180 distinct news topics. We obtained the data directly from the authors’ project website, www.structureofnews.com. For this study, we specifically use the authors’ topic labels, the monthly time series of topic attention, and the topic-term probability distributions (i.e., word weights) for each topic. Although the original dataset provides an overview of the business news, many of the 180 topics include general market conditions, corporate administration, or geopolitical events that do not directly reflect technological trajectories. To ensure our focus of shifts in public attention toward innovation, we implement a manual filtering process to select 40 topics that are highly relevant to technological fields and industries (e.g., *Internet*, *Pharma*, *Microchips*, and *Aerospace/defense*). We exclude topics that act as “noise” in the context of tracking technology trends, primarily categorizing them into four types: (1) abstract business descriptors (e.g., *Changes*, *Key role*), which lack substantive technical content; (2) macro-financial indicators (e.g., *Bond yields*, *IPOs*, *Earnings forecasts*), which reflect capital market fluctuations rather than R&D evolution; (3) corporate governance and administrative issues (e.g., *Executive pay*, *Management changes*), which focus on organizational behavior; and (4) purely geographic or political entities (e.g., *Russia*, *US Senate*, *Clintons*). By narrowing our focus to these 40 technology-relevant topics, we ensure that the resulting attention metric captures the public’s recognition of a patent’s underlying technical value rather than artifacts of general business cycles or news reporting processes.

We obtain accounting data of firms from Compustat and stock return data from the CRSP.

Following [Hirshleifer et al. \(2018\)](#), our sample excludes financial firms, identified by four-digit Standard Industrial Classification (SIC) codes ranging from 6000 to 6999 (encompassing finance, insurance, and real estate sectors), and utility firms, identified by SIC codes beginning with 49. Additionally, following [Fama and French \(1993\)](#), we exclude closed-end funds, trusts, American depository receipts, real estate investment trusts, and units of beneficial interest from the sample. The final sample comprises 3,506 publicly traded firms from 1982 to 2023. The start year is determined by our use of a five-year rolling window to compute firm-level metrics based on patent data available from 1976.

3.2 Construction of Long-Term Patent Potential Index

To capture a patent’s long-term potential, we construct a composite index that integrates both its ex-ante dormant nature and its ex-post trajectory of technology salience.

3.2.1 Dormant nature

Our dormant measures extend the framework of [Cohen et al. \(2024\)](#). We identify three key dimensions of a dormant patent: (i) it falls outside the inventor’s primary technology space, (ii) it resides in a relatively novel technology space with few competitors ([Cohen et al., 2024](#)), and (iii) it receives relatively lower public attention prior to its issuance. We construct the following variables to capture these attributes at the time of patent issuance.

Tech-class-weight - This is defined as the fraction of patents in the CPC technology class of a given patent relative to all patents held by its assignee up to the grant date of that patent. This metric reflects the degree of emphasis placed on the patent’s technology class within the firm’s overall patent portfolio at the time of the patent’s issuance. A higher tech-class-weight indicates greater firm-level focus on the patent’s technology class. Notably, different from the [Cohen et al. \(2024\)](#), we compute the tech-class-weight based solely on the firm’s known patent stock at the time of the patent’s grant.

Competition - This is quantified as the logarithm of the total number of patents from all firms within the same CPC class as the given patent, accumulated up to the grant date of that patent. Patents in CPC classes with fewer competing patents are considered to reside in relatively novel technology spaces.

Pre-Grant Attention ($Attention_{L5}$) - Unlike "hot" technologies that are born into existing market hype, a late-bloomer originates in a neglected or nascent technological domain. To measure this, we construct a news-based attention index. We utilize the news topic data from [Bybee et al. \(2024\)](#). We first map news topics to relevant CPC classifications. This process begins by employing a Large Language Model (Gemini 2.5 Pro) to match each topic to specific 4-digit CPC codes. We conduct this initial matching at a finer granularity because it compels the LLM to identify more precise technological sub-domains. The details of this CPC-topic mapping including the LLM prompts are provided in the Appendix. This procedure establishes a many-to-many relationship, where each topic can match multiple CPC codes, and each CPC code can be associated with multiple topics. Leveraging this mapping, we compute CPC-level attention by summing the monthly topic scores from [Bybee et al. \(2024\)](#), where each score represents the average probability (or share of words) assigned to a topic across all articles published that month. This process yields a monthly attention time-series for every 3-digits CPC class. Based on these monthly scores, we calculate for each patent the cumulative attention received by its technical field (defined by its CPC) over time. In instances where a patent is assigned to multiple CPC classes, we take the average of the attention scores across those classes following the methodology used by [Cohen et al. \(2024\)](#). $Attention_{L5}$ is defined as the cumulative public attention received by the patent's CPC class in the 5-year window prior to the grant. A low $Attention_{L5}$ value signifies that the patent belongs to a technology class that was overlooked by the public during its development phase.

3.2.2 Technological salience

Late-blooming patents are defined not just by their humble origins, but by their eventual transition to high-impact status. While the dormant nature captures the "quiet" phase, technological salience captures the subsequent activation phase where information frictions decrease and the technology's relevance becomes apparent to the broader public.

Post-Grant Attention ($Attention_{W5}$) - To capture this trajectory, we calculate the cumulative attention received by the patent's CPC class in the 5-year window following the grant. This period serves as the early signal window for the realization of the patent's potential. A high $Attention_{W5}$ indicates that the technological field in which the patent resides is trending.

To validate that our attention metric captures real-world technological trajectories and societal responses, we present the aggregate attention time series for three representative fields: *Information Technology*, *Health & Biotech*, and *Terrorism & Defense* (see Figure 1). It shows a high degree of synchronicity between the attention spikes and major historical events. In the information technology sector, the metric identifies surges in public attention surrounding the release of Windows 95 (1995), the peak of the Dot-com bubble (2000), and the launch of the iPhone (2007). Similarly, in the healthcare domain, the peaks correspond closely to global health crises such as the SARS outbreak (2003), the H1N1 pandemic (2009), and the Ebola crisis (2014). In the defense sector, the 9/11 attacks (2001) and the London bombings (2005) triggered immediate increases in the attention. These alignments provide evidence that our technology-relevant attention metric captures shifts in technology trends.

3.2.3 The Composite Potential Index

The dormant potential and technology salience are integrated into a single measure *Potential* consists two components to isolate patents that exhibit a late-blooming trajectory. The index

is defined as:

$$Potential = Attention_{W5} - (Tech-class-weight + Competition + Attention_{L5}) \quad (1)$$

To ensure comparability across components with different units and scales, each term in the *Potential* index formula is standardized using an expanding-window z-score based on its historical distribution up to the patent’s grant date. Specifically, the *Attention_{W5}*, *Attention_{L5}*, *Tech-class-weight*, and *Competition* are individually transformed to have a mean of zero and a unit standard deviation relative to the historical sample available at the time of issuance.

The Dormant Component (Sum of *Tech-class-weight*, *Competition*, and *Attention_{L5}*): Consistent with [Cohen et al. \(2024\)](#), these terms act as proxies for the patent’s dormant potential. A late-bloomer is likely to be a ”non-core” innovation for the firm (low *Tech-class-weight*), located in an uncrowded technology space (low *Competition*), and ignored by the public (low *Attention_{L5}*). By subtracting these terms in Equation (1), our index rewards patents that exhibit a high degree of dormancy at birth.

The Technology Salience Component (*Attention_{W5}*): This term captures the realization of value. A high positive value identifies a field that gained significant traction after the patent’s issuance.

3.2.4 Firm-Level Measure Construction

By calculating the *Potential* index for each individual patent, we then aggregate this measure to the firm level to capture the overall quality and late-blooming trajectory of a company’s innovation portfolio. Following [Hirshleifer et al. \(2018\)](#), we compute the average of *Potential* across all patents granted to the firm over the preceding five years.

Our methodology avoids look-ahead bias. By definition, if our cumulative post-grant attention measure is based on a 5-year window, the firm-level *Potential* for year *t* will inherently incorporate future information from the period up to *t*+5. Given that our study investigates

the long-term value of patents, our focus is on the power of the *Potential* to predict firm outcomes significantly further in the future, specifically in the window ranging from 6 to 20 years after year t . This temporal separation ensures that the future information embedded within the post-grant attention is not overlapped with the realization of dependent variables.

Table 1 presents the summary statistics for patent-level (Panel A) and firm-level (Panel B) variables. In Panel A, the summary statistics for the key composite index, *Potential*, reveal significant heterogeneity across the sample. The index shows a mean of -0.83 and a standard deviation of 2.00. This suggests that most patents follow a conventional innovation path rather than a late-blooming one. The Trade (binary) equals 1 when a patent experiences its first transaction more than five years after being granted. With a mean of 0.15, this indicates that 15% of patents in our sample show this delayed trading pattern. Regarding patent-level characteristics, most control variables are log-transformed to account for the highly skewed nature of patent data. The forward citation variables Citation5_10, Citation10_15, and Citation15_20 measure log-transformed citations accrued in sequential five-year windows. The mean of these log-transformed citation measures generally decrease over time, from 1.13 to 0.99 in the 15-20 year window. The standard deviation for these log-transformed variables remains relatively stable, ranging from 1.10 to 1.14, indicating that the dispersion of log-citation counts widens as patents age.

Table 1 Panel B presents the summary statistics for the firm-level variables, which are constructed as five-year rolling averages of the patent-level measures. The firm-level *Potential* shows a mean of 0.78 and a standard deviation of 1.52. While the aggregation of patents into firm-level naturally reduces the volatility observed at the patent level (where the standard deviation was 2.00), the firm-level measure retains substantial variation. The range spans from a minimum of -10.37 to a maximum of 5.19, indicating heterogeneity of corporate patent stocks.

We incorporate a set of control variables at both the patent and firm levels, following estab-

lished literature. Patent-level controls include the Innovation Originality (InnOrig), private valuation (KPSS), total number of claims in a patent (num-claims), the number of backward citations (num-citations), the count of backward citations to non-patent literature (num-nonpat-citations), the number of inventors (num-inventors), the number of distinct International Patent Classification (IPC) codes assigned to the patent (num-IPC), and the time in days from the patent’s application date to its grant date (grant-lag) (Harhoff et al., 2003; Hall et al., 2009; Kogan et al., 2017; Hirshleifer et al., 2018; Yang, 2023). Firm level controls include Innovation Originality (InnOrig) from Hirshleifer et al. (2018) and two measures of innovative efficiency, Citations-based Innovative Efficiency(CIE) and Patents-based Innovative Efficiency(PIE), from Hirshleifer et al. (2013). Other financial control variables include firm asset (Asset), sales growth (Salesgrowth), property, plant, and equipment (PPE), the debt-to-asset ratio (LEV), the book-to-market ratio (BTM), return on assets (ROA), return on equity (ROE), advertising expenses (ADV), capital expenditures (CAPEX), and R&D expenditures (R&D). Detailed definitions and construction methods for all variables are provided in Appendix A.

Before using these variables in our main analyses, we implement two data processing steps to mitigate the influence of outliers and enhance interpretability. We winsorize all continuous variables at the 1st and 99th percentiles and subsequently standardized them to have a mean of zero and a unit standard deviation.

4 Results

4.1 Long-term patent potential, forward Citations, and patent trades

Forward Citations - To provide preliminary evidence of the predictive power of our *Potential*, we conduct a non-parametric analysis using univariate sorts. We sort all sample

patents into quintiles based on their *Potential* scores. We then examine the average forward citations across these quintiles over three distinct long-term horizons: 5–10 years, 10–15 years, and 15–20 years after the grant date. Table 2 Panel A reports the results of this sorting procedure. We define the "High Potential" group as the top quintile and the "Low Potential" group as the bottom quintile. Consistent with Hypothesis 1, we find a positive relationship between the *Potential* and future scientific impact across all time windows. Notably, this difference gap widens as the time extends. It increases from 0.32 in the 5–10 year window to 0.68 in the 15–20 year window (t -statistic = 6.72). This escalating divergence suggests that the *Potential* is effective at identifying late-blooming patents.

To further validate Hypothesis 1, Table 2 Panel B presents the results where we regress log-transformed forward citations over three distinct long-term horizons (5–10 years, 10–15 years, and 15–20 years post-grant) on the *Potential* with a set of control variables. The model is specified as follows:

$$\begin{aligned}
\text{Citation}_{i,h} = & \alpha + \beta_1 \text{Potential}_i + \beta_2 \text{InnOrig}_i + \beta_3 \text{num_claims}_i + \beta_4 \text{num_citations}_i \\
& + \beta_5 \text{num_nonpat_citations}_i + \beta_6 \text{num_inventors}_i + \beta_7 \text{num_IPC}_i \\
& + \beta_8 \text{grant_lag}_i + \beta_9 \text{KPSS}_i + \beta_{10} \text{Asset}_{j,t} + \beta_{11} \text{Salesgrowth}_{j,t} \\
& + \beta_{12} \text{PP\&E}_{j,t} + \beta_{13} \text{R\&D}_{j,t} + \beta_{14} \text{CAPEX}_{j,t} + \beta_{15} \text{BTM}_{j,t} \\
& + \beta_{16} \text{LEV}_{j,t} + \beta_{17} \text{ROA}_{j,t} + \delta_t + \gamma_j + \epsilon_i
\end{aligned} \tag{2}$$

Where variables are constructed as detailed in Section 3 and Appendix A. The regression includes year fixed effects (δ_t) and firm fixed effects (γ_j). Standard errors are clustered at the firm level.

The empirical findings provide strong validation for our hypothesis. The coefficient on *Potential* is positive and statistically significant across all three specifications. In the earli-

est window (years 5–10), the coefficient stands at 0.0332 (t -statistic = 7.85). The predictive power of the *Potential* strengthens over longer horizons: the coefficient increases to 0.0674 (t -statistic = 9.38) for the 10–15 year window, and further rises to 0.0820 (t -statistic = 9.62) for the 15–20 year window. This escalating trend suggests that the *Potential* index is capable of capturing the "accrual" of scientific value over time. Collectively, these results support for our framework's ability to identify high-impact technological innovations.

Patent Trade - To test our hypothesis (H2) that patents with high long-term potential are more likely to be traded, we examine the relationship between the *Potential* index and the probability of patent reassignment using the following logit model specification:

$$\begin{aligned} \Pr(\text{Trade}_i = 1) = \Lambda \big(& \alpha + \beta_1 \text{Potential}_i + \beta_2 \text{InnOrig}_i + \beta_3 \text{num_claims}_i + \beta_4 \text{num_citations}_i \\ & + \beta_5 \text{num_nonpat_citations}_i + \beta_6 \text{num_inventors}_i + \beta_7 \text{num_IPC}_i \\ & + \beta_8 \text{grant_lag}_i + \beta_9 \text{KPSS}_i + \beta_{10} \text{Asset}_{j,t} + \beta_{11} \text{Salesgrowth}_{j,t} \\ & + \beta_{12} \text{PP\&E}_{j,t} + \beta_{13} \text{R\&D}_{j,t} + \beta_{14} \text{CAPEX}_{j,t} + \beta_{15} \text{BTM}_{j,t} \\ & + \beta_{16} \text{LEV}_{j,t} + \beta_{17} \text{ROA}_{j,t} + \delta_t + \eta_k + \epsilon_i \big) \end{aligned} \quad (3)$$

Where δ_t is the year fixed effect, and η_k is industry fixed effect.

Column (4) of Table 2 Panel B reports the results of this regression. The empirical results provide strong support for H2. We observe a positive and highly significant relationship between the *Potential* index and trading activity, with a coefficient of 0.0530 and a t -statistic of 21.17. Consistent with the framework of the 'market for ideas' (Gans and Stern, 2003; Arora et al., 2004), our results confirm that patents with high long-term potential are more likely to be traded.

4.2 Comparing potential with existing metrics

4.2.1 Distinctiveness from existing patent metrics

To validate our index, we test whether it provides unique information beyond that of existing innovation metrics in forecasting patent’s scientific value. We compare our measure against two of the most prominent indicators in the literature: the private economic value of patents (KPSS), estimated from stock market reactions (Kogan et al., 2017), and innovation originality (InnOrig), which measures the breadth of technological knowledge cited by a patent (Hirshleifer et al., 2018).

Table 3 presents the relationship between the *Potential* index and these existing measures through correlation analysis and conditional double sorts. Panel A indicates that our index is virtually orthogonal to existing measures. The correlation between *Potential* and KPSS is negligible at 0.01, and the correlation with InnOrig is similarly low at 0.05. This near-zero correlation suggests that the late-blooming quality captured by our index is different from the immediate market value recognized at the time of grant (KPSS) or the combinatorial novelty of the invention (InnOrig).

To ensure that the predictive power of *Potential* is not driven by these known factors, we conduct conditional double sorts. Panel B of Table 3 reports the results of sorting patents first by KPSS (High vs. Low) and then by *Potential*. The column (1) of Panel B shows that within the same level of private value, patents with high *Potential* significantly outperform their peers in future scientific impact. The difference in citations is 0.78 in the 5-10 year window and expands to 1.37 in the 10-15 year window. Notably, the *Potential* index can identify ‘Hidden Gems’ within the ‘Low KPSS’ group. Figure 2 visualizes these citation trajectories. To provide a robust baseline, we include a line representing the average performance of the entire ‘High KPSS’ group. When comparing the ‘Hidden Gems’ (Low KPSS, High Potential) against this baseline and the ‘Stars’ (High KPSS, High Potential), a ‘catch-up effect’ emerges. In the initial 5-10 year window, the ‘Hidden Gems’ already perform on par with the High

KPSS average but lag behind the 'Stars'. However, their trajectory exhibits a significantly steeper slope. By the 10-15 year window, the 'Hidden Gems' not only break away from the High KPSS average but also cross over to surpass the 'Stars', becoming the highest-performing group. This visual trend is statistically supported by Panel B column (3): the positive differences (e.g., 1.55 in the 5–10 year window and 0.46 in the 10–15 year window) indicate that these 'Hidden Gems' actually accumulate more future citations than the 'Stars'. This finding highlights a distinct advantage of our metric: it discovers late-blooming patents that initially have low KPSS value and eventually outperform even the established high-value patents. Panel C of Table 3 repeats this analysis controlling for InnOrig, where the average difference remains positive and significant.

4.2.2 The role of technology salience in activating dormant value

To isolate the specific contribution of the *Technology Salience Component*, we employ a conditional double-sort strategy. We first partition the patent sample based on the *Dormant Component* as defined in Section 3. Patents below the median are classified as the dormant group. Within this group, we then sort by *Technology Salience Component*. Panel D of Table 3 reports the difference in forward citations between high- and low-activation patents. The difference is positive and statistically significant across all horizons. In the initial window (5–10 years), technology salience generates a citation difference of 1.60 (t -statistic = 35.01). This effect intensifies in the medium term (10–15 years), then reaching a peak difference of 2.76 (t -statistic = 33.68) in the long term (15–20 years). This confirms that once activated by relevant trends, these patents achieve substantial scientific recognition.

4.2.3 Incremental predictive power of potential for long-term citations

While the Table 3 provides evidence of the *Potential* index's validity, we now turn to a parametric evaluation. Table 4 reports the results of examining the incremental predictive power of the *Potential* index after controlling for the KPSS and InnOrig. Columns (1)–(3) present

the baseline models including only KPSS and InnOrig, while Columns (4)–(6) introduce our *Potential* index to test its marginal contribution.

The coefficient on *Potential* is not only statistically significant across all horizons but shows a distinct monotonic increase as the time window extends. Specifically, the coefficient magnitude more than doubles from 0.0332 (t -statistic = 7.85) in the 5–10 year window to 0.0820 (t -statistic = 9.62) in the 15–20 year window. In contrast, the predictive power of the control metrics tends to decay or stagnate over the long run. The KPSS measure sees its influence diminish significantly over time. Its coefficient drops by nearly two-thirds, falling from 0.0909 in the medium term (Column 4) to just 0.0337 in the long term (Column 6), with a corresponding decline in t -statistics (from 8.53 to 4.68). Similarly, while InnOrig remains a strong predictor, its significance weakens over the extended horizon. Furthermore, the inclusion of the *Potential* index consistently enhances the model’s explanatory power. Comparing the baseline models (Columns 1–3) with the full specifications (Columns 4–6), we observe a increase in adjusted R^2 across all time horizons.

4.2.4 The ‘Scissors Effect’: temporal dynamics of value realization

To visually dissect the evolution of predictive power, Figure 3 plots the term structure of the estimated coefficients for *Potential*, KPSS, and InnOrig on annual incremental citations over the horizon from year $t + 6$ to $t + 20$. This visualization reveals a pattern we term the “Scissors Effect,” which highlights the difference in how these metrics capture value realization over time. The trajectory of the *Potential* index (green line) is characterized by a steady, monotonic ascent. Starting with a modest coefficient of approximately 0.014 in year $t + 6$, its predictive power grows consistently, doubling to nearly 0.036 by year $t + 20$. This upward slope confirms the late-blooming mechanism: the value embedded in these patents is not static but accumulates momentum as the underlying technology trends activate.

In contrast, the KPSS measure (blue line) exhibits a “decay” trajectory. While it dominates in the years 6–12) with high coefficients, its predictive power collapses after year 12. The

curve enters a steep decline, intersecting with the rising *Potential* line around year 16. By year 20, the coefficient for KPSS drops to zero and becomes statistically insignificant, suggesting that the initial stock market reaction contains little information about a patent’s scientific relevance two decades later. The InnOrig metric (red line) exhibits a relatively stable but downward trend. Starting at approximately 0.025 in year 6, it gradually declines over the sample period. While it consistently predicts citations, it shows a slow erosion in predictive power over the long horizon.

Consequently, the Figure 3 reveals a "double scissors" effect. The first intersection occurs around year 12, where the rising *Potential* curve crosses above the slowly declining InnOrig line. This signals that by the start of the second decade, the late-blooming mechanism captured by our index becomes a more potent driver of scientific impact than innovation originality. The second, more dramatic crossover follows around year 16, where *Potential* overtakes the collapsing KPSS curve. These results reveal a clear temporal shift: KPSS and InnOrig predict early-to-mid-term impact, while the *Potential* index emerges as the primary driver of long-term technological legacy.

4.3 Long-term patent potential and firm profitability

4.3.1 Long-term patent potential and future profitability

Baseline Prediction Results - Table 5 presents the results from predictive regressions of future firm profitability, measured by return on equity (ROE), on our key variable of interest, *Potential*, and a set of standard control variables. We prioritize ROE as our primary dependent variable because it directly captures the net efficiency with which a firm generates returns for its shareholders from its innovation investments. ROA instead, can be sensitive to variations in capital structure (Nissim and Penman, 2001) and often distorted by the accounting treatment of intangible assets (Lev and Sougiannis, 1996)—a common issue in R&D-intensive sectors.

$$\begin{aligned}
ROE_{i,t+k} = & \alpha + \beta_1 \text{Potential}_{i,t} + \beta_2 \text{InnOrig}_{i,t} + \beta_3 \text{CIE}_{i,t} + \beta_4 \text{PIE}_{i,t} \\
& + \beta_5 \text{ADV}_{i,t} + \beta_6 \text{R\&D}_{i,t} + \beta_7 \text{BTM}_{i,t} + \beta_8 \text{Capex}_{i,t} \\
& + \beta_9 ROE_{i,t} + \delta_t + \eta_j + \epsilon_{i,t+k}
\end{aligned} \tag{4}$$

We employ two distinct econometric methodologies: Panel A reports the time-series averages of coefficients from annual Fama-Macbeth (FM) cross-sectional regressions ([Fama and MacBeth, 1973](#)) while Panel B presents results from a panel data model that includes both industry and year fixed effects.

The results in Panel A (Fama-Macbeth) provide evidence for our hypothesis 3a. *Potential* exhibits a positive and statistically significant coefficient across all forecast horizons from t+6 to t+20, indicating that a high-quality late-blooming patent portfolio is a strong predictor of long-term financial performance. The economic magnitude of this relationship intensifies in the long run, peaking at 0.0372 for ROA_t + 13 (t-statistic = 3.32). This increase in predictive power aligns with delayed value realization mechanism, where the latent value of dormant innovations requires an extended period to be fully integrated into firm's realized profitability. The Panel Regressions in Panel B further corroborate these findings, ensuring the results are robust to the inclusion of year and industry fixed effects. *Potential* continues to show a significant predictive power across all horizons. This consistency across both econometric specifications confirms that long-term corporate value is driven by the synergistic effect of possessing latent technological quality and its subsequent technology salience.

Out-of-Sample Predictive Performance - To assess the out-of-sample (OOS) predictive performance of our models incorporating *Potential* for future ROE levels over horizons from t+6 to t+20, we employ both expanding window (starting from a 10-year initial estimation period and cumulatively adding subsequent data) and rolling window (using a fixed 10-year estimation period that shifts forward) approaches. These OOS evaluations are conducted via

Fama-MacBeth and Pooled OLS regressions. To evaluate the incremental predictive value of our index, we define a baseline specifications (Benchmark Model) which includes all the control variables in Equation 4, but excludes the *Potential* variable. We then compare the forecast accuracy of the full model against this benchmark:

$$\begin{aligned}
ROE_{i,t+k} = & \alpha + \beta_2 \text{InnOrig}_{i,t} + \beta_3 \text{CIE}_{i,t} + \beta_4 \text{PIE}_{i,t} \\
& + \beta_5 \text{ADV}_{i,t} + \beta_6 \text{R\&D}_{i,t} + \beta_7 \text{BTM}_{i,t} + \beta_8 \text{Capex}_{i,t} \\
& + \beta_9 ROE_{i,t} + \delta_t + \eta_j + \epsilon_{i,t+k}
\end{aligned} \tag{5}$$

We also estimate the coefficients using two methodologies on the training set:

- **Pooled OLS:** The coefficients are estimated by running a single regression of Equation 4 on the training sample. This yields a single, fixed set of coefficients that is applied to out-of-sample predictions.
- **Fama-MacBeth (FM):** This two-step procedure first runs annual cross-sectional regressions of Equation 4 for each year in the training sample. The final coefficients used for prediction are then calculated as the time-series averages of these annual coefficient estimates.

The primary finding from Table 6 is that the *Potential* brings significant predictive power improvement in forecasting future ROE. In the expanding window, incorporating the *Potential* variable consistently improves forecasting performance across all horizons. The R^2 improvement is not only persistent but also increases at longer horizons, reaching at 0.0051 (FM) and 0.0059 (Pooled OLS) for the ROE_{t+20} . The statistical significance of this improvement is confirmed by the Clark and West (2007) MSPE-adjusted test. For all horizons and both models, the CW T-stats are highly significant. The rolling window panel provides consistent evidence, showing even slightly higher benchmark R^2 values and similarly signif-

icant improvements. Taken together, these results demonstrate that the *Potential* variable captures additional long-term predictive power.

4.3.2 Long-term patent potential and idiosyncratic volatility

To test Hypothesis 3b, we examine the impact of firm-level aggregated *Potential* on future idiosyncratic volatility (IVOL).

$$\begin{aligned}
IVOL_{i,t+k} = & \alpha + \beta_1 Potential_{i,t} + \beta_2 InnOrig_{i,t} + \beta_3 CIE_{i,t} + \beta_4 PIE_{i,t} \\
& + \beta_5 ADV_{i,t} + \beta_6 R\&D_{i,t} + \beta_7 BTM_{i,t} + \beta_8 Capex_{i,t} \\
& + \beta_9 ROE_{i,t} + \beta_{10} IVOL_{i,t} + \delta_t + \eta_j + \epsilon_{i,t+k}
\end{aligned} \tag{6}$$

The Fama-MacBeth regression models have the same dependent and independent variables as Equation 6 but do not include fixed effect terms.

Table 7 reports the results for horizons ranging from $t + 6$ to $t + 20$. Consistent with our hypothesis, the *Potential* index shows a negative and statistically significant relationship with future IVOL across both econometric specifications. For example in Panel A, the coefficient on *Potential* is negative across all horizons and grows in magnitude over time, starting from -0.0401 at $t + 6$ (t -stat = -8.52) and reaching -0.0412 at $t + 20$ (t -stat = -4.81). This downward trend suggests that as the latent value of a firm’s innovation portfolio is gradually recognized by the market, it serves as a stabilizing force that reduces firm-specific uncertainty. The Panel B further corroborate these findings after controlling for year and industry fixed effects. These findings support our argument that high-potential patents are not high-risk speculative bets; rather, they represent durable technological assets that strengthen a firm’s market power and competitive barriers. As these late-blooming assets mature, they stabilize the firm’s performance and lead to a reduction in idiosyncratic risk.

5 Robustness Checks

To ensure that our primary findings are not driven by specific variable constructions or sample selection biases, we conduct a series of robustness tests. We first examine whether our results are sensitive to alternative definitions of our key independent variables.

5.1 Alternative measures of long-term potential

In our baseline analysis, the construction of *Potential* relies on Wall Street Journal-based news attention and specific structural technology characteristic. To test the consistency of our results, we construct an alternative measure, *Potential_{alter}*, by redefining its core components:

Dormant potential measures - We first refine the *Tech-class-weight* by using a more granular classification. Specifically, we identify the primary CPC of a patent as the 4-digit CPC category that appears most frequently in its records; in cases of a tie, the category listed first is selected. This approach is different from our previous method in main results, where weights were calculated by averaging across all CPC classes assigned to a patent. By focusing on the primary classification, we isolate the patent’s dominant technological identity and avoid the noise introduced by diluting the measure across categories.

In addition, we replace the aggregate *Competition* (stock of existing patents) with a dynamic indicator, *NewEntrants*. We define a ‘new entrant’ as a firm that either enters a specific CPC subclass for the first time or returns after a hiatus of at least five years. At the CPC-year level, *NewEntrants* is calculated as the ratio of new-entrant firms to the total number of unique firms active in that subclass during the year. For each patent, the final *NewEntrants* score is the average share of new entrants across all CPC subclasses to which the patent is assigned. Contrary to the traditional view that high entry rates imply fierce rivalry, we argue that a high *NewEntrants* signals a ‘Blue Ocean’ or a nascent stage of the technological trajectory. According to the Industry Life Cycle theory ([Gort and Klepper, 1982](#)), a surge in

entry typically occurs during the growth phase where market expansion outpaces competitive crowding. In this phase, firms compete for future dominance rather than current market share. Furthermore, Organizational Ecology theory (Hannan and Carroll, 1992) suggests that in the early stages of a field, the influx of new actors serves to legitimize the new technology rather than deplete resources, creating a positive-sum environment. Therefore, a high share of new entrants signifies a 'technological frontier', which is the ideal habitat for late-blooming innovations.

Broadening Public Attention (Google Ngram) - To address potential concerns regarding the limited scope of business news, we utilize the Google Books Ngram data to capture a broader spectrum of technological trend. This corpus aggregates word frequencies from millions of digitized books, serves as a proxy for the long-term intellectual salience of technological concepts. The use of Google Books Ngram data to track the evolution of ideas and public discourse is well-established in recent economic literature. For instance, Shiller (2017) leverages the Ngram corpus to model the "epidemiology" of economic narratives. Similarly, Magness and Makovi (2023) utilize Ngram data to quantify the intellectual influence of Karl Marx over the 19th and 20th centuries. Also, Jha et al. (2025) apply computational linguistics to the Google Books Ngram corpus to trace the semantic evolution of the finance industry.

Following the topics identified by Bybee et al. (2024), we extract the top 50 high-frequency terms for each topic. We then employ LLM (Gemini-2.5) to select the most relevant keywords (up to 10 words per topic) based on their semantic alignment with the technological domain. The specific LLM prompts and keyword lists are provided in Appendix. The attention score for a topic is then defined as the mean frequency of these selected keywords in the Ngram dataset. The attention of Google Ngram is provided in year frequency, to achieve a precise 60-month alignment with each patent's issuance, we implement a linear weighting formula. Let m be the month of patent issuance and $f = (m - 1)/12$ represent the fraction of the year

elapsed. The pre-grant and post-grant cumulative attention ($L5$ and $W5$) are calculated as:

$$L5 = (1 - f) \cdot Attn_{T-5} + \sum_{k=1}^4 Attn_{T-k} + f \cdot Attn_T \quad (7)$$

$$W5 = (1 - f) \cdot Attn_T + \sum_{k=1}^4 Attn_{T+k} + f \cdot Attn_{T+5} \quad (8)$$

Our alternative potential index is defined as:

$$Potential_alter = AttentionBook_{W5} - (Tech-Class-Weight_{prime} - NewEntrants + AttentionBook_{L5}) \quad (9)$$

Panel A of Table 8 reports the results using *Potential_alter*. Across all specifications, the coefficients remain positive and statistically significant. Specifically, the predictive power for forward citations (*Citation5_10*, *Citation10_15*, and *Citation15_20*) exhibits the same increasing trend observed in our baseline results. The impact on patent trade also remains robust ($\beta = 0.0471, t = 23.94$), confirming that our findings are not contingent on a specific source of attention data or patent metrics.

5.2 Exclusion of pharmaceutical firms

Another potential concern is that our results can be disproportionately driven by the pharmaceutical sectors. In these industries, innovation is characterized by exceptionally long development cycles, and a late-blooming value realization pattern that is intrinsic to the nature of drug discovery (Grabowski, 2002). Prior literature has documented that pharmaceutical patents exhibit significantly longer citation lags and more persistent long-term value compared to other technological domains (Hall et al., 2001). If our *Potential* metric merely captures the idiosyncratic life cycle of the drug industry, its generalizability to broader innovation contexts would be limited. To address this, we re-estimate our baseline models by excluding all patents belonging to firms from pharmaceutical industries. Following the

Fama-French 49-industry classification, we remove observations categorized under Pharmaceutical Products (Industry 13), which primarily consists of firms with SIC codes 2830, 2831, 2833, 2834, 2835, and 2836.

The results are presented in Panel B of Table 8. Even after the removal of these 'natural' late-blooming firms, the coefficients on *Potential* remain positive and highly significant across all horizons. For instance, the predictive power for long-term scientific impact continues to escalate over time, with the coefficient rising from 0.0358 for Citation5_10 to 0.0833 for Citation15_20 (t -stat = 8.88). Similarly, the relationship between Potential and future patent trading activities remains robust ($\beta = 0.0468$, t -stat = 18.34). These findings confirm that our *Potential* metric captures a fundamental economic mechanism of value discovery that applied across industries.

5.3 Subsample analysis: pre- and post-2000

The final robustness test addresses potential structural shifts in the technological and economic landscape. We divide our sample into two sub-periods using the year 2000 as a breakpoint for two primary reasons. First, the turn of the millennium was marked by the collapse of the Dot-com bubble, which induced significant structural changes in innovation activities, particularly in internet-related fields. As shown in Figure 4, patents in IT-related CPC categories - namely G06 (Computing and Data Processing) and H04 (Electric Communication Technique) - accounted for an average of 38% of all filings, peaking in 2000 before experiencing a three-year decline following the bubble's burst. Second, our study is motivated by Cohen et al. (2024), which utilizes a patent sample spanning 1976 to 1999. By splitting the sample, we can verify whether our findings remain consistent with prior literature while also testing the validity in the post-2000 era.

Panel C of Table 8 presents the results for the two sub-periods. Sub-panel C2 (Post-2000) confirms that the predictive power of our model persists in the modern innova-

tion environment. Notably, while the coefficients for forward citations remain robust, the impact on patent transactions (Trade) is substantially stronger in the post-2000 period ($\beta = 0.0739, t = 23.62$) compared to the pre-2000 era ($\beta = 0.0346, t = 7.23$). This heightened sensitivity likely reflects the maturation of the 'Market for Technology' in the 21st century, characterized by increased patent liquidity, lower information frictions in the digital age, and a more systematic recognition of latent technological value by market participants. Overall, the subsample analysis reinforces the temporal stability and modern relevance of our late-blooming identification framework.

5.4 Firm-Level Robustness: Alternative Metric and Subsamples

Panel A of Table 9 presents the panel regression results using the alternative potential measure (*Potential_alter*) constructed in Section 5. The coefficients for *Potential_alter* are positive and statistically significant across the entire forecast horizon. In Panel B of Table 9, we exclude pharmaceutical firms to ensure that the firm-level results are not artifacts of industry-specific R&D cycles. After removing the pharmaceutical sector, the magnitude and statistical significance of the *Potential* coefficients remains in all forecast horizons (e.g., $\beta = 0.0164, t = 4.10$ at $t + 20$). Panel C of Table 9 conducts a subsample analysis by splitting the sample at the year 2000. In both the pre-2000 period and the post-2000 period, *Potential* remains a significant predictor of future ROE. The supplemental tests from Table 10 confirm the core findings.

6 Conclusion

This paper introduces a novel framework that integrates ex-ante latent potential with ex-post technology salience to map the long-term value trajectory of innovations. By distinguishing between the initial dormancy of an invention and its subsequent salience of technology trends, we provide a robust tool that is distinct from traditional valuation metrics. Our

empirical analysis uncovers a "Scissors Effect" in the temporal dynamics of value realization. While the predictive power of stock market-based measures (KPSS) decays rapidly in the extended period, our *Potential* index exhibits a monotonic increase in significance, eventually becoming the dominant predictor of scientific impact in the long run. In addition, this metric excels at value discovery by identifying "Hidden Gems"—innovations located in niche, non-competitive domains that initially have relatively lower KPSS value. We find that these overlooked assets not only "catch up" but eventually surpass high-profile "Star" patents as their underlying technological themes gain salience. Furthermore, the "Market for Technology" appears to efficiently recognize this delayed value, as high-potential patents demonstrate a significantly higher probability of being traded once their latent value is validated by information signals. Aggregating these insights to the firm level, we demonstrate that a patent portfolio rich in latent potential serves as a strategic asset that drives durable financial performance. Firms holding these assets experience sustained profitability (ROE) and benefit from significantly reduced idiosyncratic volatility (IVOL). These findings are robust across diverse industries and subsample periods.

Our study conclude that the economic significance of innovation is temporally dispersed, relying on the synergy between an invention's intrinsic quality and the catalyst of evolving technological trends. For investors and managers, this suggests that long-term technological alpha lies not in chasing current "hot" topics, but in identifying the dormant building blocks of future cycles.

References

- Akcigit, U., Celik, M. A., and Greenwood, J. (2016). Buy, keep, or sell: Economic growth and the market for ideas. *Econometrica*, 84(3):943–984.
- Alexopoulos, M. (2011). Read all about it!! what happens following a technology shock? *American Economic Review*, 101(4):1144–1179.
- Arora, A., Belenzon, S., Ferracuti, E., and Nagar, J. P. (2024). Revisiting the private value of scientific inventions. Working Paper 32806, National Bureau of Economic Research.
- Arora, A., Fosfuri, A., and Gambardella, A. (2004). *Markets for Technology: The Economics of Innovation and Corporate Strategy*. MIT Press, Cambridge, MA.
- Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1):99–120.
- Baruffaldi, S. H. and Simeth, M. (2020). Patents and knowledge diffusion: The effect of early disclosure. *Research Policy*, 49(4):103927.
- Bessen, J. (2008). The value of u.s. patents by owner and patent characteristics. *Research Policy*, 37(5):932–945.
- Bybee, L., Kelly, B., Manela, A., and Xiu, D. (2024). Business news and business cycles. *The Journal of Finance*, 79(5):3105–3147.
- Chamley, C. and Gale, D. (1994). Information revelation and strategic delay in a model of investment. *Econometrica: Journal of the Econometric Society*, pages 1065–1085.
- Chan, L. K., Lakonishok, J., and Sougiannis, T. (2001). The stock market valuation of research and development expenditures. *The Journal of Finance*, 56(6):2431–2456.
- Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of econometrics*, 138(1):291–311.
- Cohen, L., Diether, K., and Malloy, C. (2013). Misvaluing innovation. *The Review of Financial Studies*, 26(3):635–666.
- Cohen, L., Gurun, U., Moon, K., and Suh, P. (2024). Patent hunters. Working Paper 32965, National Bureau of Economic Research.
- de Rassenfosse, G. and Jaffe, A. B. (2018). Econometric evidence on the depreciation of innovations. *European Economic Review*, 101:625–642.
- Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.
- Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636.
- Fang, L. and Peress, J. (2009). Media coverage and the cross-section of stock returns. *The journal of finance*, 64(5):2023–2052.
- Galasso, A., Schankerman, M., and Serrano, C. J. (2013). Trading and enforcing patent rights. *The RAND Journal of Economics*, 44(2):275–312.
- Gans, J. S. and Stern, S. (2003). The product market and the market for “ideas”: commercialization strategies for technology entrepreneurs. *Research Policy*, 32(2):333–350.
- Gaspar, J.-M. and Massa, M. (2006). Idiosyncratic volatility and product market competition. *The Journal of Business*, 79(6):3125–3152.
- Glode, V. and Ordóñez, G. (2025). Technological progress and rent seeking. *The Review of Financial Studies*, 38(4):1259–1289.
- Gort, M. and Klepper, S. (1982). Time paths in the diffusion of product innovations. *The*

- economic journal*, 92(367):630–653.
- Grabowski, H. (2002). Patents, innovation and access to new pharmaceuticals. *Journal of International Economic Law*, 5(4):849–860.
- Griliches, Z. (1958). Research costs and social returns: Hybrid corn and related innovations. *Journal of Political Economy*, 66(5):419–431.
- Hall, B. H., Jaffe, A., and Trajtenberg, M. (2005). Market value and patent citations. *The RAND Journal of Economics*, 36(1):16–38.
- 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., Thoma, G., and Torrisi, S. (2009). Financial patenting in europe. *European Management Review*, 6(1):45–63.
- Hannan, M. T. and Carroll, G. R. (1992). *Dynamics of organizational populations: Density, legitimation, and competition*. Oxford University Press.
- Harhoff, D., Scherer, F. M., and Vopel, K. (2003). Citations, family size, opposition and the value of patent rights. *Research Policy*, 32(8):1343–1363.
- Hirshleifer, D., Hsu, P.-H., and Li, D. (2013). Innovative efficiency and stock returns. *Journal of Financial Economics*, 107(3):632–654.
- Hirshleifer, D., Hsu, P.-H., and Li, D. (2018). Innovative originality, profitability, and stock returns. *The Review of Financial Studies*, 31(7):2553–2605.
- Hou, J. and Yang, X. (2019). Patent sleeping beauties: evolutionary trajectories and identification methods. *Scientometrics*, 120(1):187–215.
- Hughes, R. (2024). Examining u.s. patent values: Renewal choices, citation dynamics, and economic growth. Economic Working Paper 2024-5, USPTO.
- Irvine, P. J. and Pontiff, J. (2009). Idiosyncratic return volatility, cash flows, and product market competition. *The Review of Financial Studies*, 22(3):1149–1177.
- Jha, M., Liu, H., and Manela, A. (2025). Does finance benefit society? a language embedding approach. *The Review of Financial Studies*, page hhaf012.
- 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.
- 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.
- Lev, B. and Sougiannis, T. (1996). The capitalization, amortization, and value-relevance of r&d. *Journal of Accounting and Economics*, 21(1):107–138.
- Magness, P. W. and Makovi, M. (2023). The mainstreaming of marx: Measuring the effect of the russian revolution on karl marx’s influence. *Journal of Political Economy*, 131(6):1507–1545.
- Nissim, D. and Penman, S. H. (2001). Ratio analysis and equity valuation: From research to practice. *Review of Accounting Studies*, 6(1):109–154.
- Pakes, A. and Schankerman, M. (1984). The rate of obsolescence of patents, research gestation lags, and the private rate of return to research resources. In *R&D, Patents, and Productivity*, pages 73–88. University of Chicago Press, Chicago, IL.
- Peress, J. (2010). Product market competition, insider trading, and stock market efficiency. *The Journal of Finance*, 65(1):1–43.
- Schumpeter, J. A. and Swedberg, R. (2021). *The Theory of Economic Development*. Rout-

- ledge.
- Serrano, C. J. (2010). The dynamics of the transfer and renewal of patents. *The RAND Journal of Economics*, 41(4):686–708.
- Shiller, R. J. (2017). Narrative economics. *American Economic Review*, 107(4):967–1004.
- Singh, J. and Fleming, L. (2010). Lone inventors as sources of breakthroughs: Myth or reality? *Management Science*, 56(1):41–56.
- Teixeira, A. A., Vieira, P. C., and Abreu, A. P. (2017). Sleeping beauties and their princes in innovation studies. *Scientometrics*, 110(2):541–580.
- Trajtenberg, M. (1990). Product innovations, price indices and the (mis)measurement of economic performance. Working Paper 3261, National Bureau of Economic Research, Cambridge, MA.
- Van Raan, A. F. (2004). Sleeping beauties in science. *Scientometrics*, 59(3):467–472.
- Van Raan, A. F. (2017). Sleeping beauties cited in patents: Is there also a dormitory of inventions? *Scientometrics*, 110(3):1123–1156.
- Weitzman, M. L. (1998). Recombinant growth. *The Quarterly Journal of Economics*, 113(2):331–360.
- Yang, S. (2023). Predictive patentomics: Forecasting innovation success and valuation with chatgpt. *arXiv preprint arXiv:2307.01202*.

Tables

Table 1: Summary Statistics of Patent-level and Firm-level Variables

Panel A: Patent-level					
	N	Mean	SD	Min	Max
Potential	1405561	-0.83	2.00	-10.62	6.26
Citation5_10	1405561	1.13	1.10	0	7.85
Citation10_15	991222	1.07	1.14	0	7.51
Citation15_20	658455	0.99	1.13	0	7.62
Trade(binary)	768,976	0.15	0.36	0	1.00
InnOrig	1405561	0.54	0.30	0	2.94
KPSS	1405561	1.60	1.23	0.01	4.94
Num_claims	1405561	2.74	0.64	0.69	6.77
Num_citations	1405561	2.24	0.95	0.69	7.82
Num_nonpat_citations	1405561	0.90	1.14	0	7.95
Num_inventors	1405561	1.22	0.44	0	4.34
Num_ipc	1405561	0.92	0.37	0	5.36
Grant_lag	1405561	6.85	0.54	5.50	8.00
Panel B: Firm-level					
	N	Mean	SD	Min	Max
Potential	27504	0.78	1.52	-10.37	5.19
InnOrig	27504	0.73	0.34	0	2.00
CIE	27504	0.61	1.21	0	9.30
PIE	27504	0.19	0.36	0	3.02
ADV	27504	0.01	0.03	0	0.25
CAPEX	27504	0.05	0.05	0	0.43
R&D	27504	0.08	0.13	0	1.08
BTM	27504	0.56	0.48	0.02	7.81
ROA	27504	0.06	0.23	-1.63	0.48
ROE	25533	-0.01	0.38	-2.84	0.87
IVOL	27504	-0.19	0.89	-1.44	4.74
Asset	25850	5.96	2.00	3.56	13.26
Sales growth	25850	0.15	0.31	-0.40	0.89
PP&E	25850	0.43	0.31	0.05	1.87
LEV	25850	0.43	0.22	0.11	1.09

Notes: This table presents summary statistics for patent-level (Panel A) and firm-level (Panel B) variables for the sample period 1976–2023. *Potential* is a composite index designed to capture a patent’s late-blooming trajectory, defined as $Attention_{W5} - (Tech-class-weight + Competition + Attention_{L5})$ according to Equation (1). Detailed definitions and construction of all variables are provided in Section 3. Citation5_10, Citation10_15, and Citation15_20 represent log-transformed forward citations (excluding self-citations) accrued during the respective post-grant time intervals. Trade is a binary indicator of a patent’s first transaction occurring more than five years post-grant. InnOrig measures innovative originality (Hirshleifer et al., 2018). KPSS measures patent market value (Kogan et al., 2017). Num_claims, num_citations, and num_nonpat_citations denote the number of claims and citations to other patents and non-patent references (in logs). Num_inventors is the number of inventors, num_ipc measures technological breadth, and Grant_lag is the log-transformed days from application to grant. In Panel B, *Potential* and *InnOrig* are five-year rolling averages of patent-level measures. CIE and PIE measure innovative efficiency (Hirshleifer et al., 2013). Financial controls include: ADV (advertising to assets), CAPEX (capital expenditures to assets), R&D (R&D to assets), BTM (book-to-market ratio), ROA (return on assets), ROE (return on equity), IVOL (idiosyncratic volatility), Asset (log of total assets), Sales growth (annual revenue growth), PP&E (property, plant, and equipment to assets), and LEV (total debt-to-assets). The number of observations (N), mean, standard deviation (SD), minimum (min), and maximum (max) are reported.

Table 2: Patent Potential, Long-term Citations, and Market Transactions

Panel A: Univariate Sorts of Forward Citations by Patent Potential

	High_Potential	Low_Potential	Difference	T-statistic
Citation5_10	6.33	6.01	0.32***	4.91
Citation10_15	7.17	6.77	0.39***	4.97
Citation15_20	7.36	6.68	0.68***	6.72

Panel B: Long-term Potential, Forward Citations and Trades

	(1) Citation5_10	(2) Citation10_15	(3) Citation15_20	(4) Trade
Potential	0.0332*** (7.85)	0.0674*** (9.38)	0.0820*** (9.62)	0.0530*** (21.17)
Controls	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
Firm Fixed Effect	Yes	Yes	Yes	No
Industry Fixed Effect	No	No	No	Yes
Observations	920,017	636,498	437,344	768,665
Adjusted R-squared	0.2157	0.2290	0.2296	0.1751

Notes: **Panel A** reports the results of univariate sorts based on the patent-level *Potential*. All sample patents are sorted into quintiles according to their *Potential* scores. “High_Potential” refers to the top quintile (Group 5), and “Low_Potential” refers to the bottom quintile (Group 1). *Difference* represents the mean difference in forward citations (excluding self-citations) between the high and low potential groups. The forward citations are calculated over three distinct post-grant windows: 5–10 years, 10–15 years, and 15–20 years. **Panel B** reports the regression results of long-term patent outcomes on the *Potential*. The dependent variables in Columns (1)–(3) are log-transformed forward citations (excluding self-citations) accrued during the 5–10, 10–15, and 15–20 years post-grant, respectively. The dependent variable in Column (4) is a binary indicator for a patent’s first transaction occurring more than five years after its grant date. *Potential* is a composite index measuring a patent’s long-term potential, as defined in Section 3. Control variables include both patent-level characteristics (KPSS, InnOrig, number of claims, backward citations, non-patent citations, inventors, IPC classes, and grant lag) and firm-level financial controls (Assets, Sales growth, PP&E, R&D, Capex, Book-to-market ratio, Leverage ratio, and ROA). Standard errors are clustered at the firm level, and t-statistics are reported in parentheses. *** indicates statistical significance at the 1% level.

Table 3: Distinctiveness from Existing Patent Metrics

Panel A: Correlation Matrix

	Potential	InnOrig	KPSS
Potential	1		
InnOrig	0.05	1	
KPSS	0.01	0.03	1

Panel B: Long-term Citations by Patent Potential and KPSS Groups

	Average_Diff (HighPot-LowPot) (1)	vs. LowKPSS _LowPot (2)	vs. HighKPSS _HighPot (3)
Citation5_10	0.78***	1.00***	1.55***
Citation10_15	1.37***	1.52***	0.46***
Citation15_20	1.01***	1.09***	0.07

Panel C: Long-term Citations by Patent Potential and InnOrig Groups

	Average_Diff (HighPot-LowPot) (1)	vs. LowInnOrig _LowPot (2)	vs. HighInnOrig _HighPot (3)
Citation5_10	0.75***	1.18***	1.05***
Citation10_15	1.35***	1.67***	0.31***
Citation15_20	0.98***	1.08***	-0.41***

Panel D: Long-term Citations by Activation Levels

	HighActivation	LowActivation	Difference	T-statistic
Citation6_10	6.77	5.17	1.60***	35.01
Citation11_15	7.98	5.26	2.72***	42.70
Citation16_20	7.87	5.11	2.76***	33.68

Notes: Panel A reports Pearson correlation coefficients among three patent value measures. Panels B and C report results from a conditional double-sorting procedure. The sorting variables (Potential, KPSS, and InnOrig) are residualized measures, obtained by regressing the respective values on firm and grant-year fixed effects. In Panel B (Panel C), patents are first divided into two groups based on the sample median of KPSS (InnOrig). Within each resulting group, patents are further classified into High-Potential and Low-Potential categories using the median of the *Potential* index. The column (1) represents the mean difference in citations between the High and Low Potential groups, calculated as the average of the differences within each of the two control subsamples. The columns (2) and (3) show comparisons between the LowControl (KPSS and InnOrig) _HighPotential group and the respective baseline groups. Panel D reports the activation effect of the technology salience component within the *Dormant* patent group. The sample is restricted to patents with a dormant component (defined in Section 3) below the sample median. Within this subsample, patents are sorted into High-Activation and Low-Activation categories based on the median of the technology salience measure. The columns report the mean citation for the High and Low activation subgroups. Citations are measured over three distinct post-grant periods: 5–10 years, 10–15 years, and 15–20 years. Statistical significance at the 1%, 5%, and 10% levels is indicated by ***, **, and *, respectively.

Table 4: Incremental Predictive Power of Potential for Long-term Citations

	(1) Citation5_10	(2) Citation10_15	(3) Citation15_20	(4) Citation5_10	(5) Citation10_15	(6) Citation15_20
Potential				0.0332*** (7.85)	0.0674*** (9.38)	0.0820*** (9.62)
KPSS	0.0857*** (8.43)	0.0803*** (6.75)	0.0462*** (5.53)	0.0909*** (8.53)	0.0818*** (6.63)	0.0337*** (4.68)
InnOrig	0.1594*** (11.13)	0.1784*** (10.82)	0.1717*** (9.29)	0.1326*** (7.45)	0.1469*** (6.22)	0.1493*** (5.05)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,181,830	890,530	684,705	920,017	636,498	437,344
Adj. R-squared	0.2147	0.2246	0.2160	0.2157	0.2290	0.2296

Notes: This table reports the regression results of long-term forward citations on the *Potential* index, alongside with KPSS (Kogan et al., 2017) and InnOrig (Hirshleifer et al., 2018). The dependent variables are log-transformed forward citations (excluding self-citations) accrued during 5-10, 10-15, and 15-20 years post-grant. (i) Columns (4)-(6) present regression all three measures while Columns (1)-(3) provide baseline results excluding the *Potential* index for comparison. Control variables include both patent-level characteristics and firm-level financial controls as defined in Table 2. T-statistics are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 5: Long-term Patent Potential and Future ROE

Panel A: Fama-Macbeth															
	ROE _{t+6}	ROE _{t+7}	ROE _{t+8}	ROE _{t+9}	ROE _{t+10}	ROE _{t+11}	ROE _{t+12}	ROE _{t+13}	ROE _{t+14}	ROE _{t+15}	ROE _{t+16}	ROE _{t+17}	ROE _{t+18}	ROE _{t+19}	ROE _{t+20}
Potential	0.0142*** (2.71)	0.0116*** (1.65)	0.0154*** (2.74)	0.0189*** (4.81)	0.0233*** (4.84)	0.0290*** (4.55)	0.0356*** (4.09)	0.0372*** (3.32)	0.0227*** (3.32)	0.0291*** (4.35)	0.0285*** (3.96)	0.0302*** (3.35)	0.0440*** (2.29)	0.0363*** (3.35)	0.0358*** (3.41)
InnOrig	-0.0132*** (-3.39)	-0.0073*** (-1.19)	-0.0100*** (-1.87)	-0.0156*** (-4.14)	-0.0045*** (-0.41)	-0.0102*** (-2.43)	-0.0093*** (-1.93)	-0.0110*** (-2.45)	-0.0112*** (-2.55)	-0.0091*** (-2.12)	-0.0076*** (-1.39)	-0.0094*** (-1.38)	-0.0080*** (-1.32)	-0.0033*** (-0.57)	-0.0041*** (-0.77)
CIE	-0.0067*** (-1.61)	-0.0032*** (-0.56)	-0.0063*** (-1.23)	-0.0038*** (-0.78)	-0.0083*** (-1.18)	-0.0091*** (-1.61)	-0.0089*** (-1.82)	-0.0086*** (-1.54)	-0.0080*** (-1.79)	-0.0017*** (-0.32)	0.0036*** (0.62)	0.0019*** (0.24)	0.0094*** (0.81)	-0.0002*** (-0.01)	0.0016*** (0.14)
PIE	-0.0145*** (-4.06)	-0.0208*** (-5.11)	-0.0200*** (-5.11)	-0.0220*** (-5.18)	-0.0085*** (-2.03)	-0.0098*** (-2.33)	-0.0098*** (-2.33)	-0.0169*** (-4.32)	-0.0202*** (-5.29)	-0.0254*** (-6.40)	-0.0315*** (-7.35)	-0.0199*** (-5.17)	-0.0213*** (-5.38)	-0.0213*** (-5.38)	-0.0201*** (-5.38)
ADV	0.0196*** (6.24)	0.0199*** (6.03)	0.0179*** (3.34)	0.0279*** (6.07)	0.0287*** (6.79)	0.0262*** (7.28)	0.0286*** (8.65)	0.0313*** (7.17)	0.0278*** (7.17)	0.0280*** (6.95)	0.0296*** (6.43)	0.0290*** (5.94)	0.0293*** (4.51)	0.0226*** (5.13)	0.0261*** (5.35)
R&D	-0.1247*** (-11.16)	-0.1286*** (-8.62)	-0.1244*** (-9.41)	-0.1153*** (-8.14)	-0.1001*** (-6.54)	-0.1002*** (-6.89)	-0.0814*** (-6.17)	-0.0754*** (-4.80)	-0.1124*** (-7.20)	-0.0913*** (-7.20)	-0.0865*** (-5.28)	-0.0768*** (-4.75)	-0.0627*** (-3.04)	-0.0627*** (-3.39)	-0.0421*** (-2.47)
BTM	-0.0344*** (-4.18)	-0.0343*** (-4.13)	-0.0325*** (-4.00)	-0.0356*** (-4.71)	-0.0342*** (-3.84)	-0.0378*** (-4.73)	-0.0328*** (-4.38)	-0.0326*** (-5.29)	-0.0333*** (-5.29)	-0.0307*** (-3.91)	-0.0277*** (-4.20)	-0.0377*** (-4.26)	-0.0359*** (-3.94)	-0.0362*** (-4.59)	-0.0222*** (-2.74)
Capex	0.0042*** (1.00)	-0.0005*** (-0.12)	-0.0015*** (-0.36)	0.0041*** (1.06)	0.0079*** (2.27)	0.0050*** (1.07)	0.0090*** (2.64)	0.0070*** (2.06)	0.0070*** (2.06)	-0.0017*** (-0.45)	0.0004*** (0.15)	-0.0053*** (-0.73)	-0.0100*** (-1.38)	-0.0051*** (-0.71)	-0.0059*** (-0.79)
ROE _t	0.1136*** (6.80)	0.1085*** (7.13)	0.0940*** (4.52)	0.0805*** (7.24)	0.0914*** (6.26)	0.0865*** (8.65)	0.0668*** (8.23)	0.0654*** (6.06)	0.0638*** (4.84)	0.0671*** (5.87)	0.0636*** (7.31)	0.0723*** (5.15)	0.0720*** (6.00)	0.0619*** (4.95)	0.0799*** (5.06)
Intercept	-0.0064*** (-0.59)	0.0033*** (0.36)	0.0125*** (1.61)	0.0180*** (2.51)	0.0246*** (3.80)	0.0290*** (4.11)	0.0362*** (5.52)	0.0436*** (5.58)	0.0464*** (4.67)	0.0553*** (5.95)	0.0603*** (6.41)	0.0650*** (7.32)	0.0667*** (7.32)	0.0749*** (7.50)	0.0825*** (7.78)
Industry Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Avg. Adj. R-squared	0.2043	0.2197	0.1961	0.1779	0.1709	0.1627	0.1448	0.1384	0.1807	0.1496	0.1327	0.1462	0.1531	0.1269	0.1376
Total Firm-Year Obs.	25130	22924	20834	18947	17233	15720	14339	13064	11881	10782	9759	8803	7940	7140	6371

Panel B: Panel Regression															
	ROE _{t+6}	ROE _{t+7}	ROE _{t+8}	ROE _{t+9}	ROE _{t+10}	ROE _{t+11}	ROE _{t+12}	ROE _{t+13}	ROE _{t+14}	ROE _{t+15}	ROE _{t+16}	ROE _{t+17}	ROE _{t+18}	ROE _{t+19}	ROE _{t+20}
Potential	0.0054*** (2.28)	0.0080*** (3.37)	0.0110*** (4.31)	0.0137*** (5.19)	0.0156*** (5.44)	0.0185*** (6.04)	0.0205*** (7.28)	0.0191*** (6.42)	0.0165*** (5.23)	0.0174*** (5.17)	0.0180*** (4.96)	0.0173*** (4.75)	0.0201*** (5.25)	0.0225*** (5.71)	0.0219*** (5.33)
InnOrig	-0.0044*** (-1.07)	-0.0028*** (-0.69)	-0.0053*** (-1.23)	-0.0094*** (-2.17)	-0.0087*** (-1.87)	-0.0074*** (-1.49)	-0.0087*** (-1.88)	-0.0101*** (-2.07)	-0.0049*** (-0.94)	-0.0078*** (-1.41)	-0.0073*** (-1.23)	-0.0014*** (-0.23)	0.0004*** (0.06)	0.0011*** (0.17)	-0.0052*** (-0.79)
CIE	-0.0033*** (-1.36)	-0.0010*** (-0.41)	-0.0008*** (-0.32)	0.0017*** (0.66)	0.0001*** (0.04)	-0.0039*** (-1.34)	-0.0024*** (-0.91)	-0.0017*** (-0.61)	-0.0025*** (-0.83)	-0.0005*** (-0.14)	0.0020*** (0.57)	-0.0035*** (-1.01)	-0.0019*** (-0.52)	0.0010*** (0.26)	-0.0035*** (-0.88)
PIE	-0.0078*** (-3.12)	-0.0131*** (-5.37)	-0.0119*** (-4.64)	-0.0129*** (-4.99)	-0.0107*** (-3.88)	-0.0048*** (-1.63)	-0.0093*** (-3.46)	-0.0121*** (-4.28)	-0.0124*** (-4.16)	-0.0159*** (-4.99)	-0.0182*** (-5.37)	-0.0136*** (-3.99)	-0.0142*** (-3.94)	-0.0183*** (-4.95)	-0.0128*** (-3.34)
ADV	0.0266*** (7.04)	0.0270*** (7.23)	0.0284*** (7.29)	0.0313*** (7.96)	0.0318*** (7.59)	0.0307*** (6.90)	0.0332*** (8.15)	0.0361*** (8.35)	0.0365*** (7.90)	0.0361*** (7.36)	0.0359*** (6.84)	0.0316*** (6.00)	0.0304*** (5.52)	0.0308*** (5.51)	0.0317*** (5.49)
R&D	-0.1138*** (-25.10)	-0.1054*** (-23.60)	-0.1052*** (-22.37)	-0.1009*** (-21.11)	-0.0913*** (-17.77)	-0.0811*** (-14.88)	-0.0776*** (-15.49)	-0.0807*** (-15.19)	-0.0909*** (-16.04)	-0.0918*** (-15.30)	-0.0886*** (-13.68)	-0.0789*** (-12.10)	-0.0808*** (-11.74)	-0.0716*** (-10.05)	-0.0535*** (-7.19)
BTM	-0.0452*** (-10.02)	-0.0421*** (-9.39)	-0.0465*** (-9.87)	-0.0481*** (-10.08)	-0.0494*** (-9.77)	-0.0499*** (-9.23)	-0.0462*** (-9.30)	-0.0431*** (-8.27)	-0.0371*** (-6.74)	-0.0393*** (-6.73)	-0.0373*** (-5.97)	-0.0419*** (-6.02)	-0.0471*** (-7.19)	-0.0426*** (-6.32)	-0.0224*** (-3.18)
Capex	0.0064*** (1.27)	-0.0025*** (-0.51)	0.0057*** (1.10)	0.0149*** (2.88)	0.0148*** (2.70)	0.0119*** (2.03)	0.0113*** (2.11)	0.0085*** (1.53)	0.0043*** (0.73)	0.0073*** (1.17)	0.0057*** (0.86)	-0.0020*** (-0.30)	-0.0096*** (-1.37)	0.0048*** (0.99)	0.0073*** (0.99)
ROE _t	0.1160*** (26.82)	0.1102*** (25.96)	0.0895*** (20.05)	0.0888*** (19.68)	0.0884*** (18.39)	0.0928*** (18.27)	0.0749*** (15.88)	0.0695*** (13.79)	0.0771*** (14.18)	0.0816*** (14.11)	0.0753*** (12.01)	0.0844*** (13.38)	0.0837*** (12.39)	0.0786*** (11.24)	0.0963*** (12.88)
Intercept	0.0161*** (2.55)	0.0245*** (3.97)	0.0373*** (5.76)	0.0471*** (7.20)	0.0486*** (6.96)	0.0489*** (6.58)	0.0615*** (8.93)	0.0720*** (9.87)	0.0704*** (9.03)	0.0768*** (9.26)	0.0801*** (9.02)	0.0713*** (8.02)	0.0700*** (7.48)	0.0792*** (8.27)	0.0875*** (8.76)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.1826	0.1835	0.1650	0.1698	0.1567	0.1479	0.1608	0.1525	0.1592	0.1635	0.1496	0.1566	0.1555	0.1467	0.1419
Total Obs.	24886	22697	20622	18749	17048	15545	14175	12912	11742	10655	9644	8699	7847	7058	6299

Notes: This table reports regression results predicting future Return on Equity (ROE) using the *Potential*. The dependent variable is the firm-level ROE in year $t+k$, where k ranges from 6 to 20. For the Fama-Macbeth (FM) models, the table shows the time-series average slopes from annual cross-sectional regressions, with Newey-West autocorrelation-adjusted and heteroskedasticity-robust t-statistics in parentheses. For the panel models, coefficients from pooled regressions with industry and year fixed effects are reported. FM models include industry dummies, though their coefficients are omitted for brevity. All independent variables are winsorized at the 1% and 99% levels.

Table 6: Out-of-sample Forecasting Performance

Panel A: Expanding window																
Model	Metric	ROE_{t+6}	ROE_{t+7}	ROE_{t+8}	ROE_{t+9}	ROE_{t+10}	ROE_{t+11}	ROE_{t+12}	ROE_{t+13}	ROE_{t+14}	ROE_{t+15}	ROE_{t+16}	ROE_{t+17}	ROE_{t+18}	ROE_{t+19}	ROE_{t+20}
FM	Benchmark R^2	0.1665	0.1713	0.1510	0.1547	0.1418	0.1372	0.1471	0.1369	0.1462	0.1495	0.1357	0.1454	0.1429	0.1303	0.1317
	With Potential R^2	0.1668	0.1719	0.1523	0.1566	0.1442	0.1406	0.1525	0.1419	0.1499	0.1534	0.1396	0.1494	0.1480	0.1364	0.1368
	Improvement	0.0003***	0.0006***	0.0013***	0.0019***	0.0024***	0.0034***	0.0054***	0.0050***	0.0038***	0.0039***	0.0039***	0.0040***	0.0051***	0.0061***	0.0051***
	CW T-stat	2.85	3.26	4.46	5.30	6.02	5.49	7.49	7.31	6.00	5.51	5.00	4.74	5.11	4.86	3.48
Pooled OLS	Benchmark R^2	0.1677	0.1727	0.1525	0.1561	0.1431	0.1380	0.1482	0.1381	0.1492	0.1531	0.1397	0.1494	0.1460	0.1355	0.1350
	With Potential R^2	0.1680	0.1732	0.1536	0.1579	0.1456	0.1417	0.1540	0.1434	0.1531	0.1570	0.1437	0.1540	0.1519	0.1425	0.1409
	Improvement	0.0003***	0.0005***	0.0011***	0.0018***	0.0025***	0.0037***	0.0058***	0.0052***	0.0039***	0.0039***	0.0040***	0.0046***	0.0059***	0.0069***	0.0059***
	CW T-stat	2.88	3.12	4.05	4.92	6.02	5.38	7.59	7.30	5.94	5.44	4.98	5.02	5.39	5.11	3.74
Pooled OLS	N_{obs}	16688	15045	13498	12100	10837	9674	8605	7603	6685	5845	5057	4337	3682	3069	2480
Panel B: Rolling window																
Model	Metric	ROE_{t+6}	ROE_{t+7}	ROE_{t+8}	ROE_{t+9}	ROE_{t+10}	ROE_{t+11}	ROE_{t+12}	ROE_{t+13}	ROE_{t+14}	ROE_{t+15}	ROE_{t+16}	ROE_{t+17}	ROE_{t+18}	ROE_{t+19}	ROE_{t+20}
FM	Benchmark R^2	0.1766	0.1804	0.1601	0.1629	0.1473	0.1414	0.1532	0.1421	0.1530	0.1597	0.1456	0.1536	0.1477	0.1369	0.1358
	With Potential R^2	0.1771	0.1813	0.1614	0.1648	0.1498	0.1450	0.1589	0.1472	0.1564	0.1629	0.1491	0.1574	0.1530	0.1432	0.1412
	Improvement	0.0006***	0.0009***	0.0012***	0.0020***	0.0024***	0.0037***	0.0057***	0.0052***	0.0035***	0.0037***	0.0034***	0.0037***	0.0053***	0.0064***	0.0054***
	CW T-stat	3.61	3.74	4.48	5.27	5.73	5.34	6.87	6.52	5.06	4.66	4.38	4.15	4.74	4.62	3.47
Pooled OLS	Benchmark R^2	0.1765	0.1805	0.1603	0.1630	0.1478	0.1419	0.1534	0.1426	0.1547	0.1612	0.1472	0.1553	0.1495	0.1401	0.1382
	With Potential R^2	0.1771	0.1813	0.1614	0.1648	0.1503	0.1457	0.1592	0.1476	0.1580	0.1642	0.1506	0.1592	0.1550	0.1469	0.1441
	Improvement	0.0006***	0.0008***	0.0011***	0.0018***	0.0025***	0.0038***	0.0058***	0.0051***	0.0034***	0.0031***	0.0034***	0.0039***	0.0055***	0.0068***	0.0058***
	CW T-stat	3.62	3.73	4.24	5.01	5.74	5.19	6.92	6.50	5.05	4.62	4.38	4.29	4.85	4.79	3.61
Pooled OLS	N_{obs}	16688	15045	13498	12100	10837	9674	8605	7603	6685	5845	5057	4337	3682	3069	2480

Notes: This table reports the out-of-sample forecasting performance for future ROE comparing a benchmark model (defined by Equation 5) and a model incorporating *Potential*. We evaluate two forecasting schemes: an expanding window and a rolling window. For each scheme, we compare Fama-Macbeth (FM) and Pooled OLS models. Metrics include the Benchmark OOS- R^2 , With Potential OOS- R^2 , the Improvement (R^2 difference), and the [Clark and West \(2007\)](#) MSPE-adjusted test statistic (CW T-stat). Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: Long-term Patent Potential and Idiosyncratic Volatility

Panel A: Fama-Macbeth		IVOL _{t+6}	IVOL _{t+7}	IVOL _{t+8}	IVOL _{t+9}	IVOL _{t+10}	IVOL _{t+11}	IVOL _{t+12}	IVOL _{t+13}	IVOL _{t+14}	IVOL _{t+15}	IVOL _{t+16}	IVOL _{t+17}	IVOL _{t+18}	IVOL _{t+19}	IVOL _{t+20}
Potential		-0.0401*** (-8.52)	-0.0443*** (-6.90)	-0.0425*** (-6.62)	-0.0454*** (-6.47)	-0.0427*** (-6.48)	-0.0392*** (-6.12)	-0.0383*** (-6.22)	-0.0436*** (-6.37)	-0.0493*** (-6.49)	-0.0610*** (-6.97)	-0.0576*** (-6.44)	-0.0686*** (-6.55)	-0.0600*** (-6.66)	-0.0506*** (-6.12)	-0.0412*** (-6.81)
InnOrig		0.0259*** (4.95)	0.0253*** (4.44)	0.0191** (1.93)	0.0192** (1.77)	0.0192** (1.77)	0.0162 (1.35)	0.0180 (1.35)	0.0274** (1.97)	0.0283*** (2.58)	0.0352** (2.53)	0.0315** (2.04)	0.0206 (1.44)	0.0140 (0.98)	-0.0013 (-0.09)	0.0110 (0.83)
CIE		0.0178*** (2.77)	0.0207*** (3.90)	0.0108 (1.42)	0.0158 (1.59)	0.0052 (0.54)	0.0052 (0.87)	0.0238** (2.90)	0.0309*** (3.22)	0.0317*** (2.61)	0.0411*** (2.65)	0.0252** (2.17)	0.0076 (0.51)	0.0087 (0.44)	0.0347* (1.77)	0.0353 (1.35)
PIE		0.0326*** (4.81)	0.0317*** (3.17)	0.0460*** (5.56)	0.0452*** (4.01)	0.0490*** (5.64)	0.0448*** (4.84)	0.0504*** (6.10)	0.0574*** (5.37)	0.0514*** (3.79)	0.0430*** (2.31)	0.0576*** (4.76)	0.0634*** (3.89)	0.0591*** (3.37)	0.0357*** (2.76)	0.0251 (1.60)
ADV		0.0044 (0.55)	-0.0006 (-0.08)	-0.0092 (-1.28)	-0.0104 (-1.33)	-0.0150* (-1.93)	-0.0136 (-1.06)	-0.0108 (-1.24)	-0.0123 (-1.66)	-0.0413 (-1.46)	-0.0041 (-0.46)	-0.0080 (-0.96)	-0.0057 (-0.71)	-0.0124 (-1.55)	-0.0114 (-0.91)	0.0109 (0.91)
R&D		0.0363** (2.27)	0.0401*** (3.05)	0.0492*** (3.00)	0.0435 (1.80)	0.0376*** (2.79)	0.0352* (2.14)	0.0429 (1.73)	0.0405 (2.21)	0.0081 (0.38)	0.0085 (-0.52)	0.0163 (1.23)	-0.0117 (-0.63)	0.0119 (0.62)	0.0411* (1.80)	0.0319 (1.38)
BTM		0.0708*** (4.34)	0.0626*** (3.77)	0.0688*** (3.74)	0.0642*** (2.92)	0.0579*** (4.24)	0.0611*** (3.86)	0.0736*** (4.79)	0.0618*** (6.23)	0.0652*** (9.84)	0.0509*** (5.12)	0.0516*** (3.97)	0.0460*** (2.72)	0.0622*** (3.70)	0.0586*** (3.41)	0.0394** (2.16)
Capex		0.0325*** (7.40)	0.0286*** (4.14)	0.0275*** (4.03)	0.0204*** (3.03)	0.0289*** (2.51)	0.0226** (2.43)	0.0140 (1.64)	0.0175 (1.84)	0.0238*** (2.88)	0.0224** (2.09)	0.0074 (0.66)	0.0118 (0.71)	0.0322* (1.93)	0.0290* (1.95)	0.0443* (1.92)
ROA _t		-0.2270*** (-12.53)	-0.2153*** (-12.51)	-0.2083*** (-13.04)	-0.1975*** (-12.12)	-0.1921*** (-10.72)	-0.2057*** (-11.86)	-0.2028*** (-11.04)	-0.2136*** (-12.23)	-0.2218*** (-12.92)	-0.2266*** (-14.34)	-0.2205*** (-12.92)	-0.2380*** (-9.37)	-0.2075*** (-8.05)	-0.2028*** (-16.22)	-0.1913*** (-9.57)
IVOL _t		0.4356*** (26.16)	0.4326*** (24.62)	0.4207*** (28.30)	0.4162*** (25.11)	0.4181*** (22.04)	0.4042*** (23.49)	0.3810*** (21.69)	0.3948*** (31.37)	0.3864*** (26.71)	0.3811*** (27.56)	0.3675*** (23.22)	0.3623*** (17.25)	0.3709*** (15.90)	0.3598*** (15.26)	0.3849*** (13.89)
Intercept		0.0880 (1.36)	0.0772 (1.03)	0.0705 (0.92)	0.0650 (0.82)	0.0452 (0.58)	0.0178 (0.24)	0.0442 (0.64)	0.0482 (0.80)	0.0265 (0.41)	0.0284 (0.36)	0.0697 (1.07)	0.0785 (0.88)	0.1418 (1.53)	0.1280 (0.90)	0.2137 (1.27)
Industry Control	Yes		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Avg. Adj. R-squared		0.4564	0.4469	0.4352	0.4186	0.4034	0.3945	0.3871	0.3828	0.3787	0.3669	0.3557	0.3461	0.3564	0.3366	0.3437
Total Firm-Year Obs.		27400	25809	23675	21589	19706	17995	16489	15104	13791	12578	11434	10385	9391	8499	7640

Panel B: Panel Regression		IVOL _{t+6}	IVOL _{t+7}	IVOL _{t+8}	IVOL _{t+9}	IVOL _{t+10}	IVOL _{t+11}	IVOL _{t+12}	IVOL _{t+13}	IVOL _{t+14}	IVOL _{t+15}	IVOL _{t+16}	IVOL _{t+17}	IVOL _{t+18}	IVOL _{t+19}	IVOL _{t+20}
Potential		-0.0173*** (-5.23)	-0.0215*** (-6.26)	-0.0256*** (-6.84)	-0.0304*** (-7.48)	-0.0300*** (-6.75)	-0.0292*** (-6.09)	-0.0289*** (-5.66)	-0.0283*** (-5.30)	-0.0287*** (-5.08)	-0.0328*** (-5.08)	-0.0320*** (-5.08)	-0.0335*** (-5.07)	-0.0276*** (-3.98)	-0.0314*** (-4.22)	-0.0328*** (-4.18)
InnOrig		0.0286*** (4.99)	0.0281*** (4.71)	0.0261*** (4.10)	0.0236*** (3.49)	0.0229*** (3.17)	0.0228*** (2.97)	0.0242*** (2.98)	0.0339*** (3.98)	0.0364*** (4.02)	0.0444*** (4.62)	0.0401*** (3.96)	0.0321*** (3.04)	0.0233*** (2.11)	0.0125 (1.06)	0.0156 (1.26)
CIE		0.0143*** (4.17)	0.0127*** (3.56)	0.0127*** (3.35)	0.0143*** (3.55)	0.0059 (1.36)	0.0084 (1.83)	0.0122 (2.54)	0.0107** (2.12)	0.0138** (2.57)	0.0196*** (3.45)	0.0073 (1.22)	0.0070 (1.12)	0.0035 (0.54)	0.0120* (1.70)	0.0014 (0.19)
PIE		0.0170*** (4.86)	0.0210*** (5.75)	0.0234*** (6.02)	0.0220*** (5.34)	0.0311*** (7.08)	0.0326*** (6.99)	0.0325*** (6.63)	0.0370*** (7.23)	0.0348*** (6.42)	0.0290*** (5.08)	0.0365*** (6.10)	0.0345*** (5.39)	0.0364*** (5.59)	0.0285*** (4.09)	0.0305*** (4.14)
ADV		0.0000 (0.01)	-0.0017 (-0.32)	-0.0109* (-1.89)	-0.0112* (-1.83)	-0.0123* (-1.88)	-0.0104 (-1.51)	-0.0042 (-0.57)	-0.0135 (-1.24)	-0.0135* (-1.67)	-0.0086 (-1.00)	-0.0102 (-1.14)	-0.0075 (-0.81)	-0.0081 (-0.82)	-0.0033 (-0.32)	-0.0009 (-0.08)
R&D		0.0435*** (6.65)	0.0442*** (6.47)	0.0534*** (7.25)	0.0457*** (5.84)	0.0405*** (4.83)	0.0416*** (4.70)	0.0401*** (4.28)	0.0202** (2.06)	0.0100 (0.96)	-0.0065 (-0.59)	-0.0011 (-0.09)	-0.0003 (-0.02)	0.0224* (1.76)	0.0389*** (2.86)	0.0145 (0.99)
BTM		0.0897*** (13.89)	0.0790*** (11.70)	0.0824*** (11.36)	0.0777*** (10.02)	0.0798*** (9.71)	0.0834*** (9.64)	0.0945*** (10.31)	0.0893*** (9.28)	0.0889*** (8.77)	0.0795*** (7.48)	0.0665*** (5.97)	0.0746*** (6.46)	0.0815*** (6.73)	0.0820*** (6.36)	0.0558*** (4.05)
Capex		0.0546*** (7.88)	0.0471*** (6.53)	0.0471*** (6.12)	0.0358*** (4.37)	0.0444*** (5.11)	0.0447*** (4.87)	0.0374*** (3.87)	0.0481*** (4.77)	0.0476*** (4.50)	0.0362*** (3.26)	0.0305*** (2.63)	0.0396*** (3.28)	0.0540*** (4.30)	0.0356*** (2.67)	0.0441*** (3.15)
IVOL _t		0.4521*** (70.49)	0.4500*** (67.38)	0.4353*** (60.52)	0.4440*** (58.05)	0.4299*** (52.96)	0.4296*** (50.23)	0.4116*** (45.42)	0.4357*** (45.78)	0.4258*** (42.24)	0.4115*** (38.30)	0.4127*** (36.24)	0.4194*** (34.89)	0.4353*** (34.26)	0.4085*** (29.81)	0.4233*** (28.67)
ROA _t		-0.2113*** (-29.66)	-0.2047*** (-27.60)	-0.2065*** (-25.88)	-0.1983*** (-23.44)	-0.1995*** (-22.14)	-0.1993*** (-21.04)	-0.2092*** (-20.95)	-0.2005*** (-19.13)	-0.2316*** (-19.95)	-0.2342*** (-19.88)	-0.2357*** (-18.97)	-0.2358*** (-18.09)	-0.2358*** (-14.42)	-0.2056*** (-13.87)	-0.2161*** (-13.74)
Intercept		-0.0254*** (-2.77)	-0.0245*** (-2.57)	-0.0177* (-1.74)	-0.0042 (-0.39)	0.0055 (0.48)	0.0137 (1.13)	0.0202 (1.56)	0.0228* (1.68)	0.0325** (2.24)	0.0232 (1.50)	0.0306* (1.87)	0.0520*** (3.04)	0.0622*** (3.47)	0.0684*** (3.57)	0.0843*** (4.16)
Industry FE	Yes		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared		0.4232	0.4158	0.4035	0.3955	0.3817	0.3785	0.3732	0.3733	0.3682	0.3537	0.3516	0.3526	0.3474	0.3368	0.3315
Total Obs.		27138	25565	23448	21377	19509	17810	16317	14945	13645	12444	11312	10274	9292	8411	7563

Notes: This table reports results from Fama-Macbeth (FM) and panel regressions on idiosyncratic volatility. The dependent variable (IVOL) is computed at the fiscal year-end as the standard deviation of the residuals from regressing daily excess stock returns on the Fama-French five-factor returns over the 12-month period ending at the fiscal year-end (requiring a minimum of 31 trading days). For the Fama-Macbeth (FM) models, the table shows the time-series average slopes from annual cross-sectional regressions, with Newey-West autocorrelation-adjusted and heteroskedasticity-robust t-statistics in parentheses. For the panel models, coefficients from pooled regressions with industry and year fixed effects are reported. Control variables include past idiosyncratic volatility (IVOL_t) and other firm characteristics. All independent variables are winsorized at the 1% and 99% levels.

Table 8: Robustness: Long-term Patent Potential, Forward Citations and Trades

Panel A: Alternative Measure of Potential				
	Citation5_10	Citation10_15	Citation15_20	Trade
Potential _{alter}	0.0371*** (6.29)	0.0522*** (8.44)	0.0527*** (8.09)	0.0471*** (23.94)
Controls	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
Firm Fixed Effect	Yes	Yes	Yes	No
Industry Fixed Effect	No	No	No	Yes
Observations	1,032,812	805,039	599,515	804,162
Adjusted R^2	0.1992	0.2187	0.2129	0.1431
Panel B: Exclude Pharmaceutical Firms				
	Citation5_10	Citation10_15	Citation15_20	Trade
Potential	0.0358*** (7.06)	0.0711*** (8.58)	0.0833*** (8.88)	0.0468*** (18.34)
Controls	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
Firm Fixed Effect	Yes	Yes	Yes	No
Industry Fixed Effect	No	No	No	Yes
Observations	866,654	595,083	406,564	725,080
Adjusted R^2	0.2121	0.2253	0.2241	0.1790
Panel C: Subsample Analysis				
<i>Sub-panel C1: Pre-2000</i>	Citation5_10	Citation10_15	Citation15_20	Trade
Potential	0.0694*** (5.64)	0.0985*** (7.75)	0.0956*** (7.90)	0.0346*** (7.23)
Controls	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
Firm Fixed Effect	Yes	Yes	Yes	No
Industry Fixed Effect	No	No	No	Yes
Observations	231,480	231,396	231,396	184,723
Adjusted R^2	0.1999	0.2189	0.2183	0.1341
<i>Sub-panel C2: Post-2000</i>	Citation5_10	Citation10_15	Citation15_20	Trade
Potential	0.0194*** (4.07)	0.0345*** (5.72)	0.0544*** (7.92)	0.0739*** (23.62)
Controls	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
Firm Fixed Effect	Yes	Yes	Yes	No
Industry Fixed Effect	No	No	No	Yes
Observations	688,345	404,659	205,613	581,954
Adjusted R^2	0.2166	0.2380	0.2389	0.2122

Notes: This table reports the robustness checks for the impact of patent potential on long-term outcomes. **Panel A** uses an alternative measure of patent potential (*Potential_{alter}*). **Panel B** excludes pharmaceutical firms, defined as Industry 13 in the Fama-French 49-industry classification, which encompasses SIC codes 2830, 2831, 2833, 2834, 2835, and 2836. **Panel C** conducts subsample analysis by splitting the sample into two periods: patents granted before 2000 (Sub-panel C1) and patents granted in or after 2000 (Sub-panel C2). In all panels, the dependent variables are log-transformed forward citations (excluding self-citations) accrued during the 5–10, 10–15, and 15–20 years post-grant, and a binary indicator for a patent's first transaction occurring more than five years after its grant date. *Potential* is a composite index measuring a patent's long-term potential, as defined in Section 3. Control variables include both patent-level characteristics (KPSS, InnOrig, number of claims, backward citations, non-patent citations, inventors, IPC classes, and grant lag) and firm-level financial controls (Assets, Sales growth, PP&E, R&D, Capex, Book-to-market ratio, Leverage ratio, and ROA). Standard errors are clustered at the firm level, and *t*-statistics are reported in parentheses. *** and ** indicate statistical significance at the 1% and 5% levels, respectively.

Table 9: Long-term Patent Potential and Future ROE: Robustness and Subsamples

	ROE_{t+6}	ROE_{t+7}	ROE_{t+8}	ROE_{t+9}	ROE_{t+10}	ROE_{t+11}	ROE_{t+12}	ROE_{t+13}	ROE_{t+14}	ROE_{t+15}	ROE_{t+16}	ROE_{t+17}	ROE_{t+18}	ROE_{t+19}	ROE_{t+20}
Panel A: Panel Regression (Alternative Measure)															
Potential.alter	0.0042*** (2.79)	0.0039*** (2.67)	0.0041*** (2.68)	0.0063*** (4.18)	0.0083*** (5.25)	0.0093*** (5.58)	0.0100*** (6.30)	0.0090*** (5.34)	0.0088*** (4.88)	0.0090*** (4.71)	0.0106*** (5.39)	0.0116*** (5.91)	0.0125*** (6.12)	0.0142*** (6.84)	0.0138*** (6.48)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.1704	0.1741	0.1553	0.1616	0.1503	0.1425	0.1490	0.1401	0.1445	0.1445	0.1379	0.1404	0.1421	0.1363	0.1358
Total Obs.	26,882	25,342	23,904	22,583	20,708	18,947	17,293	15,772	14,371	13,105	11,957	10,895	9,926	9,017	8,133
Panel B: Panel Regression (Exclude Pharmaceutical Firms)															
Potential	0.0048** (2.48)	0.0063*** (3.03)	0.0096*** (4.41)	0.0117*** (5.10)	0.0116*** (4.68)	0.0129*** (4.88)	0.0173*** (6.49)	0.0178*** (6.26)	0.0168*** (5.43)	0.0166*** (5.04)	0.0159*** (4.49)	0.0142*** (4.03)	0.0154*** (4.12)	0.0163*** (4.33)	0.0164*** (4.10)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.1715	0.1587	0.1352	0.1233	0.1169	0.1153	0.1149	0.1084	0.1083	0.1128	0.1075	0.1182	0.1173	0.1155	0.1124
Total Obs.	22,712	20,748	18,876	17,185	15,649	14,295	13,064	11,918	10,852	9,861	8,942	8,083	7,297	6,574	5,879
Panel C: Subsample Analysis by Time Period															
<i>Year < 2000</i>															
Potential	0.0110*** (3.60)	0.0140*** (4.67)	0.0131*** (3.30)	0.0155*** (5.02)	0.0127*** (3.83)	0.0120*** (3.49)	0.0137*** (4.25)	0.0130*** (3.84)	0.0111*** (3.26)	0.0146*** (3.98)	0.0160*** (4.12)	0.0139*** (3.56)	0.0157*** (3.86)	0.0191*** (4.60)	0.0202*** (4.75)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.1775	0.1720	0.1065	0.1538	0.1413	0.1344	0.1302	0.1194	0.1297	0.1306	0.1258	0.1359	0.1410	0.1288	0.1371
Total Obs.	9,305	8,683	8,094	7,543	7,061	6,666	6,327	6,012	5,713	5,447	5,171	4,911	4,692	4,490	4,298
<i>Year ≥ 2000</i>															
Potential	0.0057 (1.62)	0.0073** (2.00)	0.0127*** (3.55)	0.0166*** (3.88)	0.0227*** (4.77)	0.0293*** (5.56)	0.0303*** (6.26)	0.0290*** (5.57)	0.0271*** (4.67)	0.0251*** (3.94)	0.0262*** (3.64)	0.0298*** (3.97)	0.0381*** (4.57)	0.0383*** (4.25)	0.0322*** (3.16)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.1911	0.1955	0.2077	0.1859	0.1727	0.1629	0.1886	0.1852	0.1899	0.1960	0.1752	0.1835	0.1813	0.1810	0.1672
Total Obs.	15,580	14,014	12,528	11,205	9,987	8,879	7,848	6,900	6,029	5,208	4,473	3,787	3,155	2,567	2,000

Notes: This table reports robustness tests for the impact of Potential on future firm performance (ROE). The dependent variable is the firm-level Return on Equity (ROE) in year $t + k$, where k ranges from 6 to 20. Panel A uses an alternative measure of *Potential*. Panel B excludes pharmaceutical firms, defined as Industry 13 in the Fama-French 49-industry classification, which encompasses SIC codes 2830, 2831, 2833, 2834, 2835, and 2836. Panels C conduct subsample analysis by splitting the sample period at the year 2000 ($Year < 2000$ and $Year \geq 2000$, respectively). All columns report coefficients from panel regressions with industry and year fixed effects. Control variables (as specified in the main analysis) are included but their coefficients are omitted for brevity. t -statistics based on robust standard errors clustered by firm are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 10: Long-term Patent Potential and Idiosyncratic Volatility: Robustness and Subsamples

	$IVOL_{t+6}$	$IVOL_{t+7}$	$IVOL_{t+8}$	$IVOL_{t+9}$	$IVOL_{t+10}$	$IVOL_{t+11}$	$IVOL_{t+12}$	$IVOL_{t+13}$	$IVOL_{t+14}$	$IVOL_{t+15}$	$IVOL_{t+16}$	$IVOL_{t+17}$	$IVOL_{t+18}$	$IVOL_{t+19}$	$IVOL_{t+20}$
Panel A: Alternative measure of Potential															
Potential- <i>alter</i>	-0.0169*** (-7.59)	-0.0188*** (-8.14)	-0.0198*** (-8.17)	-0.0224*** (-8.93)	-0.0209*** (-7.93)	-0.0210*** (-7.58)	-0.0197*** (-6.65)	-0.0202*** (-6.40)	-0.0230*** (-6.86)	-0.0208*** (-7.57)	-0.0298*** (-8.03)	-0.0299*** (-7.87)	-0.0313*** (-7.96)	-0.0370*** (-8.91)	-0.0361*** (-8.34)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.4112	0.4029	0.3928	0.3885	0.3765	0.3732	0.3647	0.3581	0.3500	0.3388	0.3358	0.3412	0.3364	0.3267	0.3201
Total Obs.	29,065	27,426	25,903	24,520	23,256	21,448	19,673	18,032	16,484	15,072	13,781	12,612	11,508	10,526	9,559
Panel B: Excluding Pharmaceutical Firms															
Potential	-0.0134*** (-4.09)	-0.0170*** (-4.97)	-0.0210*** (-5.67)	-0.0247*** (-6.14)	-0.0233*** (-5.30)	-0.0225*** (-4.71)	-0.0221*** (-4.38)	-0.0212*** (-4.01)	-0.0242*** (-4.31)	-0.0283*** (-4.76)	-0.0275*** (-4.37)	-0.0287*** (-4.36)	-0.0188*** (-2.70)	-0.0225*** (-3.05)	-0.0197*** (-2.51)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.4075	0.3952	0.3812	0.3731	0.3571	0.3515	0.3479	0.3494	0.3352	0.3223	0.3174	0.3208	0.3096	0.2989	0.2984
Total Obs.	24,530	23,144	21,264	19,411	17,735	16,215	14,863	13,635	12,471	11,397	10,382	9,449	8,564	7,765	6,997
Panel C: Subsample Analysis by Time Period															
<i>Subsample: Year < 2000</i>															
Potential	-0.0250*** (-4.74)	-0.0274*** (-4.92)	-0.0281*** (-4.71)	-0.0299*** (-4.69)	-0.0257*** (-3.71)	-0.0247*** (-3.39)	-0.0235*** (-3.16)	-0.0183** (-2.40)	-0.0175** (-2.21)	-0.0271*** (-3.33)	-0.0239*** (-2.84)	-0.0272*** (-3.17)	-0.0171** (-1.98)	-0.0173* (-1.91)	-0.0251*** (-2.71)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.4385	0.4338	0.4139	0.3904	0.3492	0.3478	0.3493	0.3575	0.3633	0.3538	0.3562	0.3510	0.3638	0.3547	0.3512
Total Obs.	9,986	9,340	8,720	8,148	7,645	7,184	6,809	6,472	6,156	5,884	5,614	5,358	5,120	4,911	4,713
<i>Subsample: Year ≥ 2000</i>															
Potential	-0.0275*** (-6.07)	-0.0319*** (-6.82)	-0.0370*** (-7.19)	-0.0424*** (-7.48)	-0.0438*** (-7.02)	-0.0423*** (-6.12)	-0.0439*** (-5.78)	-0.0500*** (-6.19)	-0.0563*** (-6.45)	-0.0585*** (-6.19)	-0.0618*** (-6.03)	-0.0589*** (-5.29)	-0.0591*** (-4.79)	-0.0739*** (-5.37)	-0.0591*** (-3.86)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.4244	0.4156	0.4063	0.4045	0.4077	0.4047	0.3972	0.3958	0.3856	0.3687	0.3624	0.3673	0.3443	0.3317	0.3251
Total Obs.	17,151	16,225	14,728	13,229	11,863	10,626	9,508	8,473	7,489	6,560	5,698	4,916	4,172	3,500	2,850

Notes: This table reports robustness checks for the impact of Potential on future idiosyncratic volatility (IVOL). The dependent variable (IVOL) is computed at the fiscal year-end as the standard deviation of the residuals from regressing daily excess stock returns on the Fama-French five-factor returns over the 12-month period. Panel A uses an alternative measure of *Potential*. Panel B excludes pharmaceutical firms, defined as Industry 13 in the Fama-French 49-industry classification, which encompasses SIC codes 2830, 2831, 2833, 2834, 2835, and 2836. Panel C provide subsample analysis for the periods before and after the year 2000. All models report coefficients from panel regressions including industry and year fixed effects. Controls for past idiosyncratic volatility ($IVOL_t$) and other firm characteristics are included. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Figures

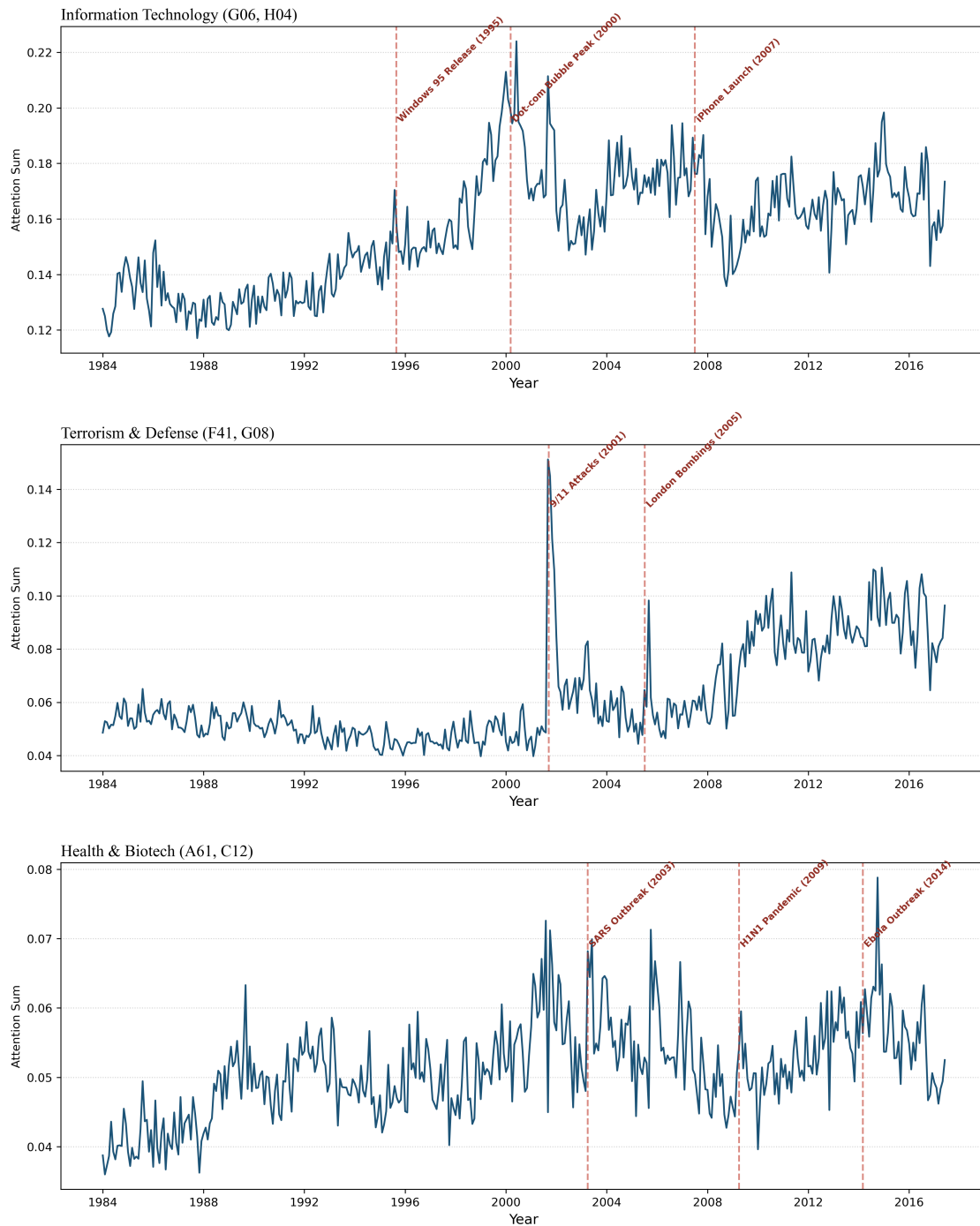


Figure 1: CPC Attention and Key Historical Events

Notes: This figure plots the aggregate attention time series for three representative technological fields: Information Technology (comprising G06: Computing; Calculating and H04: Electric Communication Technique), Terrorism & Defense (comprising F41: Weapons and G08: Signaling), and Health & Biotech (comprising A61: Medical or Veterinary Science; Hygiene and C12: Biochemistry and Genetic Engineering). Vertical dashed lines indicate major historical events, such as the Windows 95 release (1995), the peak of the Dot-com bubble (2000), the 9/11 attack (2001), and various global health crises.

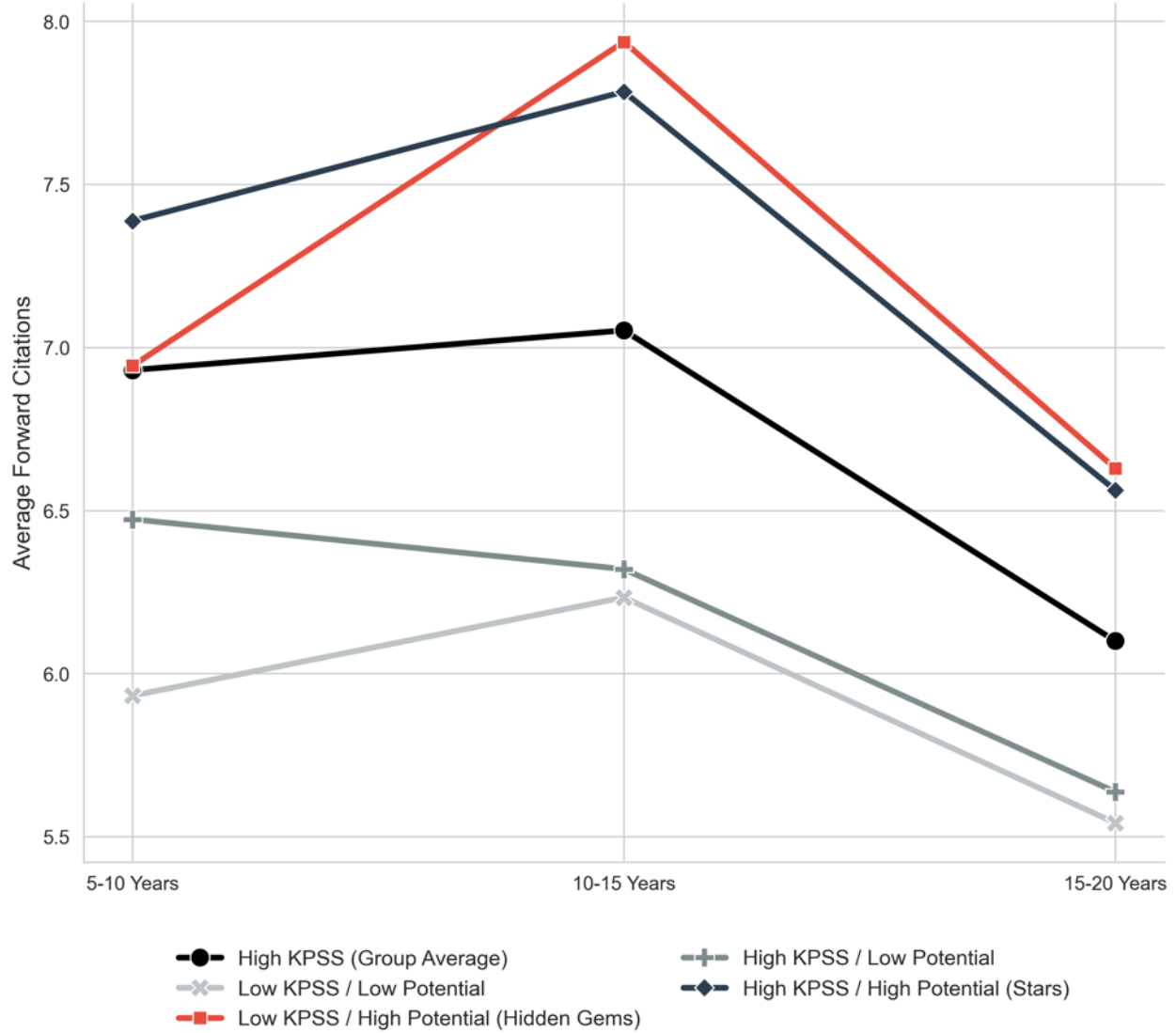


Figure 2: Long-term Citation Trajectory by Double-sorted Patent Groups

Notes: This figure plots the evolution of average forward citations across three post-grant windows (5–10, 10–15, and 15–20 years) for four patent portfolios. Portfolios are constructed using a conditional double sort procedure. Patents are first sorted into High and Low groups based on the residuals of market value (KPSS). Within each KPSS group, patents are further sorted into High and Low groups based on the residuals of the *Potential* index. All sorting variables are residualized with respect to firm and grant-year fixed effects.

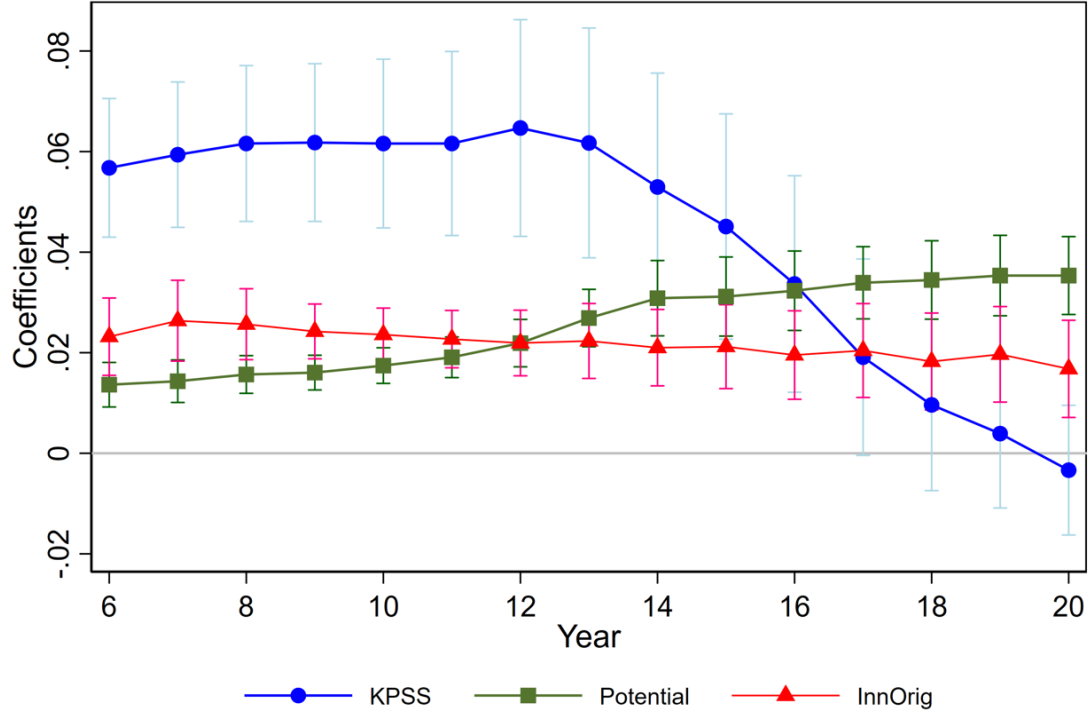


Figure 3: The 'Scissors' Effect: Term Structure of Coefficients

Notes: This figure illustrates the predictive power of different patent valuation metrics on forward citations. The dependent variable is the natural logarithm of annual incremental citations ($\ln(1 + \Delta \text{Citation}_t)$), defined as the flow of new citations received in each specific year from year $t + 6$ to $t + 20$ after the grant year. The key independent variables are: (1) *Potential*; (2) KPSS (Kogan et al., 2017); and (3) InnOrig (Hirshleifer et al., 2018). The plotted points represent the estimated coefficients (β) from separate regressions, controlling for firm characteristics and patent characteristics. Vertical bars represent 95% confidence intervals. All regressions include firm and year fixed effects, with standard errors clustered at the firm level.

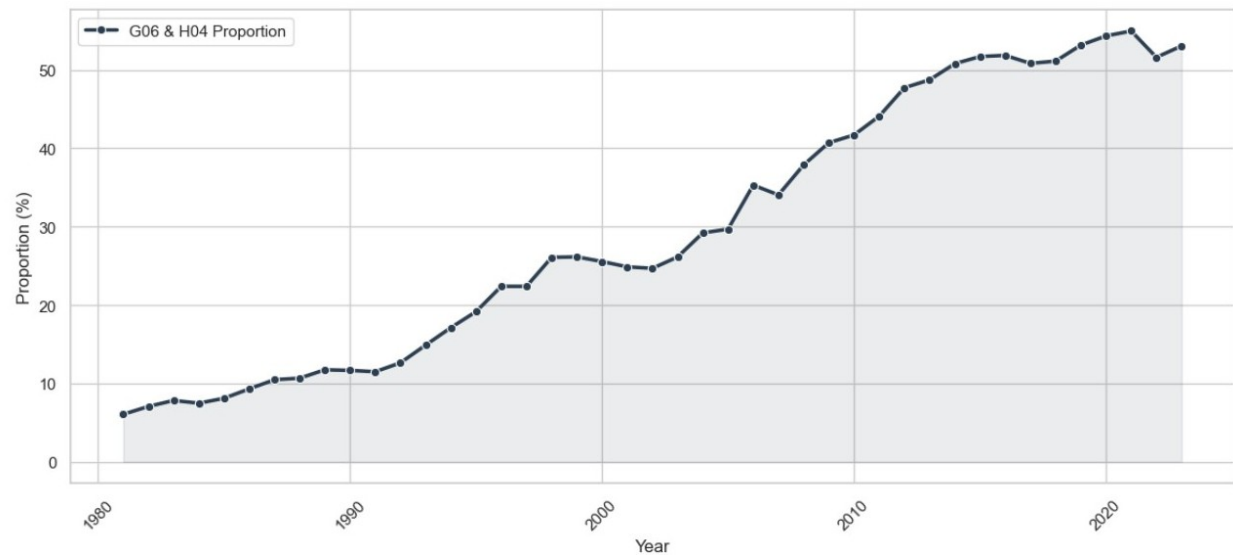


Figure 4: Annual Proportion of G06 and H04 Patents

Notes: This figure illustrates the annual proportion of patents granted in key information technology categories—specifically G06 (Computing; Calculating; Counting) and H04 (Electric Communication Technique)—relative to all granted patents from 1980 to 2023.

Appendix A: Variable Definitions

Variable	Definition
A. Core Patent-Level Variables	
Potential	The primary composite index of a patent’s long-term potential, defined as $Attention_{W5} - (Tech-class-weight + Competition + Attention_{L5})$.
Attention_W5	(Component of <i>Potential</i>) Cumulative public attention received by a patent’s technological domain (mapped to CPC classes) over the 5 years post-grant, based on the news topic attention series from Bybee et al. (2024) .
Attention_L5	(Component of <i>Potential</i>) Cumulative public attention received by a patent’s technological domain over the 5 years pre-grant (prior to the grant date).
Tech-class-weight	(Component of <i>Potential</i>) The fraction of patents in the Cooperative Patent Classification (CPC) technology class of a given patent relative to all patents held by its assignee up to the grant date.
Competition	(Component of <i>Potential</i>) The logarithm of the total number of patents from all firms within the same CPC class as the given patent, accumulated up to the grant date.
Potential_alter	Alternative measure used for robustness, substituting <i>Tech-class-weight</i> with <i>Tech-class-weight_prime</i> , <i>Competition</i> with <i>NewEntrants</i> and utilizing Google Ngram data for activation scores. Defined as $(AttentionBook_{W5} - AttentionBook_{L5}) - Tech-class-weight_prime + NewEntrants$.
AttentionBook_W5	Cumulative attention received by a patent’s technological domain (mapped to CPC classes) over the 5 years post-grant, based on the attention series from Google Book Ngram
AttentionBook_L5	Cumulative public attention received by a patent’s technological domain over the 5 years pre-grant (prior to the grant date), based on the attention series from Google Book Ngram
Tech-class-weight_prime	The fraction of patents in the Cooperative Patent Classification (CPC) technology class (Prime CPC) of a given patent relative to all patents held by its assignee up to the grant date.
NewEntrants	The proportion of active firms in a CPC subclass that are either first-time entrants or returning after a hiatus of at least five years.
B. Other Patent-Level Variables	
KPSS	Private value of a patent, calculated based on the KPSS method (Kogan et al., 2017), representing the market value of the patent.

Variable	Definition
Trade (binary)	Binary indicator that equals 1 if a patent is sold or transferred at any point after its grant date, and 0 otherwise.
Citation5	Natural logarithm of one plus the number of forward citations (excluding self-citations) accrued within 5 years after the patent's grant.
Citation5_10	Natural logarithm of one plus the number of forward citations (excluding self-citations) accrued between 5 and 10 years after the patent's grant.
Citation10_15	Natural logarithm of one plus the number of forward citations (excluding self-citations) accrued between 10 and 15 years after the patent's grant.
Citation15_20	Natural logarithm of one plus the number of forward citations (excluding self-citations) accrued between 15 and 20 years after the patent's grant.
Num_claims	Number of patent claims.
Num_citations	Number of other patents cited by the patent (backward citations), reflecting the patent's reliance on prior art.
Num_nonpat_citations	Number of non-patent citations (e.g., scientific literature) cited by the patent.
Num_inventors	Number of inventors listed on the patent.
Num_IPC	Number of International Patent Classification (IPC) codes assigned to the patent.
Grant_lag	Lag in days from the patent application date to the grant date.

C. Firm-Level Variables

Potential (firm-level)	Average <i>Potential</i> index across all patents granted to a firm over the preceding five-year rolling window.
InnOrig	Innovation Originality. The number of unique technology categories found in a patent's backward citations. Following Hirshleifer et al. (2018) .
CIE	Citations-based innovative efficiency measure. Adjusted patent citations received in year t-1 by patents granted in years t-2 to t-6, scaled by the sum of R&D expenses in years t-4 to t-8.
PIE	Patents-based innovative efficiency measure. Patents granted to a firm in year t-1 scaled by R&D capital in year t-3.
BTM	The ratio of book equity to market capitalization.
ROA	Return on assets.
ROE	Return on equity.
Assets	Total assets.
ADV	Advertising expenses.
Capex	Capital expenditure divided by assets.
R&D	R&D expenditure divided by assets.
Salesgrowth	Calculated as the year-over-year growth rate of revenue.

Variable	Definition
PPE	Property, plant, and equipment divided by assets.
LEV	Debt to assets ratio.

Appendix B: CPC-Topic Mapping

Topics	CPC1	CPC2	CPC3	CPC4	CPC5	CPC6	CPC7	CPC8
Natural disasters	G01W	G01V	G08B	E04H	E02B	A62D	A62B	
Internet	H04L	G06F	G06Q	H04W	H04N			
Soft drinks	A23L	B65D	B67D	C02F	B01F			
Mobile devices	H04W	G06F	H04M	H04B	G09G	H01M	H02J	
Police/crime	G06V	G08B	G01N	F41H	A61B	E05B		
Electronics	H01L	G06F	H05K	H03K	H02M	H03F	H01B	
Steel	C21C	C22C	C21D	B21B	C23C			
Cable	H01B	G02B	D07B	H02G	B65H			
Fast food	A47J	G06Q	B65D	B67D	A23L	A21B		
Disease	A61K	A61B	C12N	C12Q	A61P	G16H	A61M	G01N
Music industry	G11B	G06Q	G10H	G10L	H04S	H04L		
Chemicals/paper	D21C	D21H	D21F	C08B	B31B	C08L	C09D	
Environment	Y02E	C02F	B01D	B09C	G01N	E03B	Y02W	
Mining	E21C	B03B	E21B	E21D	E21F	G01V		
Pharma	A61K	A61P	C07D	C07K	C12Q	G01N	C12N	
Aerospace/defense	B64C	B64G	F02K	G01S	F41H	F42B	F41G	B64D
Phone companies	H04W	H04M	H04L	G06Q	G06F	H04B		
Computers	G06F	H01L	H04L	G06T	H05K	G11C	H03K	
Nuclear/North Korea	G21G	G21C	B01D	F42B	G21F	G21J	F41G	
Tobacco	A24B	A24F	A24C	A24D				
Biology/chemistry/physics	C12N	A61K	C07D	C08F	H01L	G21C	G01N	
Movie industry	H04N	G06T	G11B	G02B	H04S	G03B		
Automotive	B60W	B62D	B60K	B60L	F02D	H01M	F02B	
Machinery	F16H	F16C	F15B	B25J	F04B	B23B	F01	G05B
Space program	B64G	F02K	G05D	G01S	H02S	A62B	H04B	
Oil drilling	E21B	G01V	C09K	B63B	F16K			
Rail/trucking/shipping	B61F	B61L	B62D	B60P	B63B	B65G		
Airlines	G06Q	G08G	B64F	B64D	G07C			
Software	G06F	G06N	G06Q	H04L	G06T	G06V		
Agriculture	A01D	A01C	A01K	A01G	A01H	A01M	C12N	
Real estate	G06Q	E04B	E04G	E04H	G01C	G08B		
Currencies/metals	G06Q	B42D	G07D	C22B	G07F	G01N	C22C	
Credit cards	G06Q	G06K	H04L	G07F	B42D			
Terrorism	G01N	G06V	F41H	G08B	G01S	A62D		
Oil market	G06Q	C10G	B63B	F17D	E21B	B65G		
Couriers	G06Q	B65G	G06K	B60P	B65D	G08G		
Safety administrations	A62B	F16P	G08B	G01N	B60W	A23L		
Microchips	H01L	G03F	G06F	G11C	B82Y	H03K	H05K	
Luxury/beverages	C12G	C12H	B65D	B42D	A47J	G06K	B67D	
Retail	G06Q	A47F	G07G	G06K	B65G			

Appendix C: Large Language Model Prompts

CPC-Topic Mapping Prompt

You are now a senior patent analyst and technology classification researcher. Your core task is to accurately match a given cutting-edge technology topic (Topic) with relevant 4-digit Cooperative Patent Classification (CPC) codes.

You must strictly follow the instructions below:

- **Identify Core Classifications:**
 - Find the 4-digit CPC classifications that are directly related to the Topic.
 - Relevance is the primary criterion, and there is no limit to the number of codes. A Topic might correspond to only one core CPC or multiple CPCs.
 - Sort the identified CPCs in descending order of relevance.
- **Provide Official Descriptions:** For each identified CPC, provide its official and concise description.
- **Explain Rationale for Relevance:** For each CPC, clearly explain in 1-2 sentences why it is relevant to the input Topic.
- **Formatted Output:** You must present the results in two separate Markdown tables as specified below.
 - **Table 1 (Detailed Information):** Organize all findings into a clear Markdown table with three columns: *CPC (4-digit code)*, *CPC Description*, and *Rationale for Relevance*.
 - **Table 2 (Quick Copy):** Immediately following the first table, create a separate, single-row table. The first cell should contain the Topic name, followed by all the identified CPC codes you listed, in descending order of relevance. *Do not include column headers for this second table.*

Example

Input Topic: "Electric Vehicle Battery"

Expected Output:

CPC (4-digit code)	CPC Description	Rationale for Relevance
H01M	PROCESSES OR MEANS, E.G. BATTERIES, FOR THE DIRECT CONVERSION OF CHEMICAL ENERGY INTO ELECTRICAL ENERGY	This is the most core classification, directly covering the chemistry, structure, and manufacturing of batteries. As the power source for EVs is the battery pack, this is of primary relevance.
B60L	ELECTRICALLY-PROPELLED VEHICLES; PROPULSION OR CONTROL SYSTEMS THEREOF	This classification covers how batteries are integrated, managed, and controlled within the vehicle's powertrain (e.g., motors, controllers), directly reflecting their application in automobiles.
H02J	CIRCUIT ARRANGEMENTS FOR CHARGING OR DEPOLARISING BATTERIES OR FOR SUPPLYING LOADS FROM BATTERIES	This includes systems for battery charging, discharge management, balancing, and monitoring circuits. Patents related to the Battery Management System (BMS) in EVs fall into this category.

Electric Vehicle Battery	H01M	B60L	H02J
--------------------------	------	------	------

Ngram Anchor Word Selection Prompt

System Role: You are an expert in Text Analysis and Economic History. Your task is to select high-quality "Anchor Words" for Google Books Ngram analysis based on LDA topics.

Context & Constraints:

- **No POS Tags:** Use plain text only (e.g., "earthquake", not "earthquake_NOUN_").
- **Avoid Ambiguity:** Avoid common verbs or generic terms that could be misinterpreted (e.g., avoid "run", "time", "said").
- **Specificity:** Prefer specific nouns and bigrams (e.g., "tsunami", "interest rate").
- **Case Sensitivity:** Keep Proper Nouns capitalized (e.g., "Internet", "Marxism") and generic terms lowercase.

Input Format: === TOPIC: [Topic Name] ===

Candidates: word1, word2, ...

Output Format: Return ONLY a JSON object where keys are Topic Names and values are lists of selected strings.

Example Output:

```
{  
  "Natural disasters": ["earthquake", "tsunami", "hurricane", "flood"],  
  "Internet": ["Internet", "World Wide Web", "website", "online"]  
}
```

Appendix D: Topic-Specific Ngram Anchor Words

Topic Name	Selected Anchor Words (Keywords)
Natural disasters	natural disaster, earthquake, tsunami, hurricane, flood, Hurricane Katrina, evacuation, epicenter, debris, emergency management
Internet	Internet, World Wide Web, website, search engine, social media, cyberspace, electronic commerce, Facebook, Google, online
Soft drinks	soft drink, Coca-Cola, Pepsi, soda, cola, PepsiCo, bottler, Cadbury Schweppes, beverage, diet soda
Mobile devices	smartphone, mobile phone, cell phone, iPhone, BlackBerry, Android, mobile device, text message, handset, wireless technology
Police/crime	police officer, violent crime, homicide, robbery, drug trafficking, criminal justice, narcotics, cocaine, police department, drug dealer
Electronics	consumer electronics, digital camera, video game, PlayStation, Nintendo, compact disk, hard drive, Sony, Kodak, videocassette
Steel	steel industry, steelmaker, steel mill, aluminum, iron ore, steelmaking, smelter, Bethlehem Steel, Alcoa, Nucor
Cable	cable industry, Time Warner, Comcast, AOL, HBO, cable system, Cablevision, media conglomerate, Charter Communications, subscriber
Fast food	fast food, McDonald's, Burger King, Taco Bell, Pizza Hut, Starbucks, franchise, hamburger, pizza, restaurant chain
Disease	infectious disease, heart disease, breast cancer, HIV, epidemic, vaccine, Alzheimer's disease, pneumonia, immune system, virus
Music industry	music industry, record label, Warner Music, Universal Music, album, concert, EMI, copyright, iTunes, musician
Chemicals/paper	chemical industry, Union Carbide, DuPont, International Paper, Weyerhaeuser, specialty chemical, pulp, BASF, Georgia-Pacific, plastic
Environment	greenhouse gases, carbon dioxide, climate change, global warming, EPA, hazardous waste, Superfund, endangered species, air pollution, fossil fuel
Mining	mining industry, Rio Tinto, BHP Billiton, Newmont Mining, iron ore, gold mine, Anglo American, copper, coal, bullion
Pharma	pharmaceutical industry, Pfizer, Merck, GlaxoSmithKline, Eli Lilly, Novartis, generic drug, Bristol-Myers Squibb, AstraZeneca, antidepressant

Topic Name	Selected Anchor Words (Keywords)
Aerospace/defense	Lockheed Martin, Boeing, Airbus, Northrop Grumman, Raytheon, McDonnell Douglas, General Dynamics, defense contractor, aerospace industry, jet engine
Phone companies	telecommunications, AT&T, Verizon, Sprint, MCI, Baby Bell, Bell Atlantic, fiber optic, long distance, SBC Communications
Computers	personal computer, mainframe, IBM, Compaq, Dell, Hewlett-Packard, supercomputer, desktop computer, workstation, Digital Equipment
Nuclear/North Korea	North Korea, nuclear weapon, uranium, plutonium, Pyongyang, Kim Jong-il, nuclear program, economic sanctions, nonproliferation, International Atomic Energy Agency
Tobacco	tobacco industry, Philip Morris, cigarette, nicotine, Marlboro, R.J. Reynolds, lung cancer, smoker, Altria, health risk
Biology/chemistry/physics	DNA, gene, biology, physics, chemistry, Nobel Prize, molecule, protein, physicist, Massachusetts Institute of Technology
Movie industry	Hollywood, motion picture, box office, Disney, Paramount Pictures, DreamWorks, movie theater, blockbuster, animation, Columbia Pictures
Automotive	automotive industry, Ford Motor, Toyota, Chevrolet, Chrysler, sport utility vehicle, Honda, minivan, fuel economy, electric vehicle
Machinery	manufacturing, assembly line, Caterpillar, Deere, machine tool, construction equipment, supply chain, Whirlpool, automation, factory
Space program	NASA, Space Shuttle, astronaut, space station, satellite, spacecraft, rocket, Mars, orbit, National Aeronautics and Space Administration
Oil drilling	Exxon Mobil, offshore drilling, Chevron, oil spill, Royal Dutch Shell, natural gas, pipeline, Halliburton, oil exploration, Amoco
Rail/trucking/shipping	railroad, Union Pacific, Amtrak, freight, locomotive, shipping, trucking, CSX, Norfolk Southern, Conrail
Airlines	airline industry, American Airlines, Delta Air Lines, United Airlines, TWA, flight attendant, British Airways, Southwest Airlines, Federal Aviation Administration, pilot
Software	Microsoft, software industry, Bill Gates, Oracle, Windows, Linux, Netscape, web browser, spreadsheet, operating system
Agriculture	agriculture, genetically modified, farmer, Tyson Foods, agribusiness, Monsanto, poultry, USDA, ethanol, livestock
Real estate	real estate, commercial property, REIT, real estate market, property value, real estate developer, commercial real estate, housing market, foreclosure, Rockefeller

Topic Name	Selected Anchor Words (Keywords)
Currencies/metals	British pound, Japanese yen, Swiss franc, greenback, currency market, currency trader, Comex, gold price, exchange rate, foreign exchange
Credit cards	credit card, MasterCard, American Express, debit card, Visa, consumer credit, annual fee, credit union, delinquency, payment system
Terrorism	terrorism, Al Qaeda, Taliban, terrorist attack, suicide bomber, Osama bin Laden, Guantanamo Bay, insurgency, counterterrorism, jihad
Oil market	OPEC, crude oil, Brent crude, oil price, gasoline price, petroleum, oil supply, West Texas Intermediate, oil production, oil refinery
Couriers	Postal Service, FedEx, United Parcel Service, Post Office, Federal Express, direct mail, logistics, overnight delivery, parcel, customer service
Safety administrations	National Transportation Safety Board, OSHA, air bag, traffic safety, National Highway Traffic Safety Administration, occupational safety, plane crash, defective, seat belt, recall
Microchips	semiconductor, Intel, Silicon Valley, microchip, microprocessor, integrated circuit, Pentium, Cisco Systems, AMD, memory chip
Luxury/beverages	Louis Vuitton, Anheuser-Busch, Budweiser, champagne, luxury goods, LVMH, vodka, whiskey, Calvin Klein, Moet Hennessey
Retail	Wal-Mart, department store, Home Depot, Costco, Sears, Kmart, Macy's, retail industry, Nordstrom, discount store