Patents' Long-term Potentials

Chen Yang* Ying Jiang† Yuxin Zhang‡

July 2026

Abstract

This paper proposes a measure of long-horizon patent value, *Potential*, that integrates a patent's structural marginality, inventor team composition, and the post-grant trajectory of technology salience. 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. At the firm level, *Potential* forecasts higher profitability, lower idiosyncratic volatility, and reduced financing constraints. We further show that *Potential* supplements KPSS (Kogan et al., 2017) by capturing the long-horizon component of patent value: within both low- and high-KPSS groups, high-*Potential* patents generate stronger long-run citation performance than low-*Potential* patents. Consistent with the value-discovery role, a higher share of high-*Potential* patents in a firm's recent patent portfolio predicts stronger long-run firm outcomes.

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

JEL Codes: G14, G32, O31, O34

*Nottingham University Business School China, University of Nottingham Ningbo China. Email: Chen.Yang@nottingham.edu.cn.

†Nottingham University Business School China, University of Nottingham Ningbo China. Email: Ying.Jiang@nottingham.edu.cn.

‡Nottingham University Business School China, University of Nottingham Ningbo China. Email: Yuxin.Zhang@nottingham.edu.cn.

1 Introduction

Patents represent key innovation outputs, yet their economic value often follows unpredictable patterns that challenge conventional valuation. While some patents generate immediate market recognition and firm profits realization, others remain sleep for years before revealing their potential (Cohen et al., 2024). This phenomenon of late-blooming patents suggests that traditional patent valuation methods, emphasizing immediate market reactions and short-term indicators, may underestimate long-term value. Understanding the mechanisms that drive delayed patent value realization has implications for intellectual property valuation, firm assessment, and long-term technological risk (Baruffaldi and Simeth, 2020).

Traditional patent valuation centers on private and scientific worth. For private value, Kogan et al. (2017) link patent grants to stock returns via short-window event studies. Equity markets are inherently forward-looking. KPSS, the value measure introduced by Kogan et al. (2017), may already reflect these expectations in stock prices. However, the extent to which such expectations remain informative over long horizons is unclear, especially for technologies whose economic relevance emerges only gradually as subsequent innovations and applications develop over time. 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 long-run impact (Trajtenberg, 1990; Hall et al., 2001, 2005). By contrast, originality and generality are relatively short-horizon indicators (Hall et al., 2001; Hirshleifer et al., 2018).

While literature including Cohen et al. (2013); Hirshleifer et al. (2013, 2018) has examined patent value and innovation outcomes, a gap remains regarding the long-term evolution of patent value and its determinants. Patent value does not always emerge immediately. Some patents are recognized early, whereas others attract limited initial attention and generate significant technological and economic impact only over time. The existence of such late-

blooming patents (Cohen et al., 2024) indicates that traditional valuation methods may miss the distinction between immediate hits and patents whose value is realized with delay.

We propose such a predictor by linking delayed value realization to three elements: a patent's structural marginality at an early stage, the composition of its inventor team, and subsequent shifts in technology salience. Structural marginality captures the initial condition under which a patent can be overlooked. Such inventions are weakly aligned with their assignee's prevailing technological trajectory and often lie outside the typical capability set of potential adopters, making expected payoffs uncertain and complementary assets scarce at grant (Cohen et al., 2024; Van Raan, 2017; Hou and Yang, 2019). However, marginality alone does not imply late-blooming value: some marginal patents remain peripheral and never diffuse widely. Delayed value realization should instead depend on whether the patent also contains knowledge foundations that support broader downstream applications and whether its technological environment later becomes more favorable. Inventor-team composition captures the diversity of inventors' prior experience, which can broaden the range of applications that emerge from the invention (Singh and Fleming, 2010; Huo et al., 2019). Technology salience captures the activation stage: under sparse information and coordination frictions, rational agents may postpone search, integration, and commercialization until the surrounding technological environment evolves in the patent's favor (Chamley and Gale, 1994; Glode and Ordonez, 2025). Thus, patents that combine initial structural marginality with diverse inventor experience and a subsequent increase in technology salience should exhibit accelerating long-horizon impact.

We operationalize this idea through the *Potential* index, constructed from five ex-ante and early post-grant components. We measure structural marginality in two dimensions following Cohen et al. (2024): The first, *Tech-class-weight*, captures technological non-centrality within the assignee's existing patent portfolio. The second, *Competition*, captures the degree of crowding in the patent's technology space, measured by the concentration of patents in its

Cooperative Patent Classification (CPC) class. Unlike Cohen et al. (2024), we compute both measures using only information available at grant, which avoids ex-post bias in defining a patent's initial state. The third component, *vintage_diversity*, captures differences in inventors' prior patenting experience. Teams that combine inventors with different experience profiles are more likely to produce inventions whose relevance extends beyond their immediate use environment (Weitzman, 1998; Fleming, 2001; Wuchty et al., 2007; Jones, 2009). Finally, we integrate two technology-salience components: pre-grant attention ($Attention_{L5}$) in the five-year window before grant, and post-grant attention ($Attention_{W5}$) in the five-year window after grant. Low pre-grant and high post-grant attention together identify the activation phase emphasized in the "Sleeping Beauty" literature (Van Raan, 2004, 2017; Hou and Yang, 2019). The five components are aggregated through Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) (Hwang and Yoon, 2012) relative-closeness score so that a higher *Potential* value indicates greater proximity to the profile of a high-potential late-blooming patent.

Our empirical analysis uses patents granted between 1976 and 2023 that can be linked to publicly traded U.S. firms. The final sample includes 1,405,625 patents and 3,506 publicly traded firms. We evaluate patent value realization using forward citations and the probability of post-grant patent trading. At the patent level, *Potential* predicts two realization channels. First, the citation advantage of high-*Potential* patents grows with horizon. Second, *Potential* predicts a higher probability of post-grant trade, consistent with the market for ideas eventually recognizing latent value (Arora et al., 2004; Serrano, 2010; Akcigit et al., 2016). These effects are identified after controlling for KPSS and Hirshleifer et al. (2018)'s innovation originality (InnOrig). Firm level *Potential* reflects the long-run quality of a firm's innovation assets. Hirshleifer et al. (2018) suggests that the characteristics of innovation asset are associated with firm profitability and market outcomes. Consistent with these points, average *Potential* forecasts higher ROA and ROE, lower idiosyncratic volatility (IVOL), and lower financing constraints (SA index) over long horizons. The effects are robust across

industries and subperiods.

Most importantly, *Potential* carries information that complements KPSS. The predictive content of KPSS for future citations decays at long horizons, while that of *Potential* strengthens, producing a scissor-shaped pattern. To further illustrate this complementarity, we conduct a conditional double-sort on KPSS and *Potential*. This procedure identifies patents with low KPSS but high *Potential*, which we refer to as "HiddenGems." Within the low-KPSS group, HiddenGems generate stronger long-run citation performance than low-*Potential* patents. Within the high-KPSS group, patents with high *Potential* also outperform those with low *Potential*. At the firm level, patent portfolios rich in high-*Potential* patents predict higher ROA and ROE and lower IVOL and SA index.

Our findings contribute to three stands of literature. First, we add to the innovation-economics literature on patent value by proposing a novel forward-looking measure of long-term patent potential. It captures both the initial conditions under which a patent is overlooked and the factors that explain why some overlooked patents later become influential. This allows us to identify patents whose value is not immediately captured by commonly used patent-value measures, such as grant-date market valuation or citation-based indicators (Kogan et al., 2017; Hall et al., 2005; Hirshleifer et al., 2018). Our double-sort analysis shows that this predictive content is not a relabeling of KPSS: it identifies patents systematically missed by grant-date market valuation. Second, we contribute to the literature on innovation and firm performance (Chan et al., 2001; Hirshleifer et al., 2013, 2018) by showing that the composition of a firm's patent portfolio with respect to long-horizon potential predicts not only profitability but also future uncertainty and financing-constraint measures. Third, we provide a methodological contribution by mapping CPC classifications to technology-relevant news topics from Bybee et al. (2024), yielding a field-level measure of technology salience that can be used to study shifts in technological trends more broadly.

While our framework draws on Cohen et al. (2024), two distinctions are worth noting. First,

we move beyond structural marginality by incorporating inventor-team composition and the evolution of technology salience. Our framework seeks to explain why only some of overlooked patents later realize substantial value. Second, our structural-marginality components use only information available at grant and avoid look-ahead bias.

The remainder of this paper is organized as follows: Section 2 reviews literature and develops hypotheses. Section 3 describes data and variables. Section 4 presents results. Section 5 examines incremental information of our index. Section 6 presents robustness checks. Finally, Section 7 concludes the paper.

2 Literature Review and Hypothesis Development

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

Structural marginality

The long-term potential of a patent refers to its capacity to generate substantial technological influence or economic value over an extended period after grant (Harhoff et al., 2003). Cohen et al. (2024) show that many patents are recognized with substantial delay (14 years) and eventually become highly influential. These late-blooming patents tend to be technologically peripheral to the assignee's core domains and to arise in less crowded technology spaces. They interpret this through a "write-or-hunt" framework rooted in Weitzman (1998)'s recombinant-growth view: firms may write peripheral, option-like patents that become influential once better-matched downstream users incur the search, licensing, and adaptation costs to hunt and integrate them.

Inventor team composition

As the stock of knowledge expands, innovation increasingly requires teams that combine different forms of prior experience rather than relying on a single inventor's knowledge base (Wuchty et al., 2007; Jones, 2009, 2010). The reason is that inventive search involves both exploiting accumulated knowledge and exploring new combinations (March, 1991; Weitzman, 1998). Inventors with longer patenting histories bring established technical experience and accumulated problem-solving routines, but experience can also make search more local and path dependent (March, 1991). Inventors whose experience is concentrated in more recent periods can be closer to emerging technological trajectories, consistent with evidence that career stage shapes the use of new ideas and the nature of inventive output (Kaltenberg et al., 2023). This is especially relevant for late-blooming patents: their value is not exhausted by the immediate technological environment, but instead depend on whether the invention embeds knowledge combinations that remain useful as downstream applications evolve. Hence, patents produced by teams with more diverse prior experience should embed greater potential for delayed impact.

Technological salience

As Hou and Yang (2019) notes, a patent may initially lack complementary technologies but have its latent value unlocked as the corresponding technology enters a growth cycle. Drawing on Geels (2002), we view value activation as a process in which a patent's underlying domain moves from low visibility to broader technological relevance. Alexopoulos (2011) show that publications track commercialization more closely than R&D or patent counts, suggesting that public discourse contains information about the evolution of technology. In a related spirit, Bybee et al. (2024) apply latent Dirichlet allocation to business news text to identify a large set of topics and trace how attention to each topic evolves over time. This makes news attention a useful macro-level proxy for shifts in the salience of technological domains over time.

Together, these arguments motivate our *Potential* measure. Structural marginality captures why a patent may be overlooked at grant. Inventor-team composition captures whether the team brings together different experience that can connect established knowledge with exploratory search. Technological salience captures whether the surrounding domain later becomes more visible. Combining these components helps distinguish patents that remain marginal from "HiddenGems" whose long-term value is activated over time.

2.2 Realization of long-term patent value

Scientific Value (Forward Citations) - A primary channel through which a patent's latent potential is realized is scientific influence, captured by the accumulation of forward citations. When a patent introduces a valuable idea, other inventors build on it and cite it (Hall et al., 2005; Kogan et al., 2017). Late-blooming patents often show 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 also aligns with Weitzman (1998), where peripheral yet high-potential ideas can serve as building blocks for future developments. Consequently, for a structurally marginal patent to achieve this high-impact trajectory, its inventor team should embody experience-based foundations that support broader downstream applications. Its underlying technological domain should also move from low initial visibility into a growth phase that facilitates wider diffusion. We define such patents as having long-term potential and capture this potential using the *Potential* index. We hypothesize that this process requires the conjunction of a patent's structural marginality, inventor-team composition, and technology salience to unlock its scientific value.

H1: Patent-level long-term potential is positively associated with future citation counts.

Patent's Marketability - Patents are tradable assets whose value can be realized through market transactions (Arora et al., 2004; Serrano, 2010; Galasso et al., 2013). Serrano (2010) show that patents are more likely to be transferred when the expected gains from trade exceed the costs of technology adoption, while Akcigit et al. (2016) emphasize patent transfers as a mechanism through which firms monetize innovations that are misaligned with their core technological capabilities. In our setting, patents with higher long-term potential are more likely to contain latent applications that are not fully exploited by the original assignee. As these applications become more visible, the expected gains from reallocating the patent to a better-matched user increase (Cohen et al., 2024). Whether driven by the originator's desire to monetize an underutilized asset or an acquirer's ability to develop the technology more effectively, high *Potential* patents are therefore expected to be more likely to be traded as their latent value becomes recognized by the public.

H2: Patents with higher long-term potential are positively associated with future patent transactional activities.

Innovation Activities and Corporate Performance - Innovation can improve firm performance when it creates technologies that are difficult for competitors to replicate, support future product-market opportunities, or improve the firm's ability to adapt to technological change (Hall et al., 2005; Hirshleifer et al., 2018). A firm with more high-*Potential* patents has an innovation portfolio with greater scope for delayed value realization, whether through follow-on innovation, patent transactions, or later use in economically relevant applications (Serrano, 2010; Akcigit et al., 2016). As these channels raise the expected economic returns to the firm's technological assets, we expect the long-term potential embedded in a firm's patent portfolio to be reflected in future financial performance:

H3a: Aggregated long-term *Potential* of a firm's patent portfolio is positively associated with long-term profitability, measured by return on assets (ROA) and return on equity (ROE).

The firm-level consequences of high-*Potential* patent portfolios should also extend to firm-specific uncertainty. Idiosyncratic volatility reflects uncertainty about firm-specific future cash flows, as well as information risk faced by investors (Irvine and Pontiff, 2009; Yang et al., 2020). Patents and patent-protected R&D can be viewed as real options because they preserve flexibility over the timing and direction of future innovation investment (Bloom and Van Reenen, 2002; Schwartz, 2004). A portfolio rich in high-*Potential* patents gives the firm a stock of latent technological options. As related technological domains gain salience, these options become more viable, allowing firms to adjust the timing, scale, and direction of subsequent innovation investment as uncertainty resolves (Bloom and Van Reenen, 2002; Schwartz, 2004). This flexibility is expected to lower firm-specific cash-flow risk and idiosyncratic volatility.

H3b: Aggregated long-term *Potential* of firm level patent portfolio is negatively associated with idiosyncratic volatility (IVOL).

High-*Potential* patents are also more economically pledgeable forms of innovation capital. Suh (2023) show that stronger control rights over patents increase firms' debt capacity. Mann (2018) and Hochberg et al. (2018) also show that patents can serve as valuable collateral supporting external financing for innovative firms. Therefore, a firm whose patent portfolio contains more inventions with high long-run potential are expected to have lower financing constraints. Accordingly, firm-level patent portfolio potential is expected to be negatively associated not only with idiosyncratic volatility, but also with financing constraints as proxied by the SA index.

H3c: Aggregated long-term *Potential* of firm level patent portfolio is negatively associated with financing constraints, as proxied by the SA index.

3 Data and Variables

3.1 Data

Our sample comprises firms at the intersection of the Compustat 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. It also provides inventor identifiers and inventor-patent linkages. 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 measured as KPSS. 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) and we obtained the data from the authors' project website, `www.structureofnews.com`.

We obtain accounting data of firms from Compustat. Firm-level idiosyncratic volatility is obtained from the Open Source Asset Pricing Cross-Section database of Chen and Zimmermann (2022).[1] 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,

[1] The data is available at `www.openassetpricing.com/data`.

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

Composite index integrates ex-ante structural marginality, inventor-team composition, and ex-post trajectory of technology salience.

3.2.1 Structural marginality

Structural marginality measures is based on the framework of Cohen et al. (2024). We identify two key dimensions: (i) it falls outside the inventor's primary technology space, and (ii) it resides in a relatively novel technology space with few competitors.

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. 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 less crowded technology spaces (Cohen et al., 2024), which may reflect either emerging or less actively developed domains.

3.2.2 Inventor team composition

Inventor teams combine distinct knowledge and experience, expanding the set of ideas available for recombination and influencing invention impact (Singh and Fleming, 2010; Huo et al., 2019). We capture this aspect of inventor-team composition by measuring heterogeneity in inventors' prior experience within the focal team. It indicates whether team members accumulated their inventive experience in different periods, which help generate inventions with broader downstream applications whose relevance emerges over time.

Vintage diversity - We proxy inventor team composition using *vintage_ diversity*, a patent-level measure based on the dissimilarity of inventors' prior patenting experience within the focal team. We obtain inventor identifiers and inventor-patent linkages from PatentsView, which allow us to trace each inventor's prior patenting activity. For each inventor, we first build a yearly history vector that records the number of patents the inventor applied for in each year prior to the focal patent's filing year. We then compute the cosine distance for every inventor pair within the team based on these pre-filing historical vectors, and define *vintage_ diversity* as the average pairwise cosine distance across all team members. For single-inventor patents, we set *vintage_ diversity* to zero. In cases where an inventor has no observable patenting history before the focal filing year, the corresponding zero vector is treated as maximally distant from other inventors' histories, so the pairwise distance is set to one. A higher value of *vintage_ diversity* indicates that inventors on the focal patent have less similar prior patenting histories in terms of when and how intensively they patented before the focal filing year. Such teams combine inventors with different experience profiles, which may broaden the range of perspectives embedded in the patent.

3.2.3 Technological salience

Late-blooming patents are defined not only by their low initial visibility, but also by the later activation of their underlying technological domains. To capture this activation pro-

cess, we construct a news-based attention index using the business news topics from Bybee et al. (2024), who apply latent Dirichlet allocation (LDA) to approximately 800,000 Wall Street Journal articles from 1984 to 2017 and identify 180 topics. We use their topic labels, monthly topic-attention series, and topic-term probability distributions to measure changes in attention to technology-related domains.

The full set of 180 topics in Bybee et al. (2024) covers broad business news, including market conditions, corporate administration, and geopolitical events. We manually retain 40 topics closely related to technological fields, such as *Internet*, *Pharma*, *Microchips*, and *Aerospace/defense*. We exclude topics that capture abstract business descriptors (e.g., *Changes*, *Key role*), macro-financial conditions (e.g., *Bond yields*, *IPOs*), corporate governance or administration (e.g., *Executive pay*, *Management changes*), and geographic or political events (e.g., *Russia*, *US Senate*). This filtering step helps ensure that our attention measure reflects shifts in public attention toward technological domains rather than general business-cycle or news-reporting variation. The full list of selected topics is reported in Appendix B.

We then map the selected news topics to relevant Cooperative Patent Classification (CPC) classes. Specifically, we use Large Language Model(Gemini 2.5 Pro) to match each topic to 4-digit CPC codes, which provides a finer-grained link between business-news topics and technological subdomains. The resulting CPC-topic mapping, reported in Appendix B, is many-to-many. For instance, in each month we aggregate topic-attention scores from Bybee et al. (2024) across all topics mapped to a given CPC class. This procedure yields a monthly attention series for each 3-digit CPC class. For each patent, we measure cumulative attention to its technological field based on the CPC classes assigned to the patent. When a patent belongs to multiple CPC classes, we use the average attention across those classes, following the approach in Cohen et al. (2024) for constructing *tech-class-weight*.

- **Pre-grant attention** ($Attention_{L5}$): Cumulative public attention to the patent's

technological field during the five years before grant. A low value indicates that the patent originates from a relatively quiet or overlooked technology class.

- **Post-grant attention ($Attention_{W5}$)**: Cumulative public attention to the patent's technological field during the five years after grant. A high value indicates that the relevant technology class gains visibility after grant, providing an early signal of potential activation.

3.2.4 The composite potential index

We use the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [2] developed by Hwang and Yoon (2012) to construct our patent potential measure. The index is based on five components: *Tech-class-weight*, *Competition*, *Vintage_ diversity*, pre-grant attention ($Attention_{L5}$), and post-grant attention ($Attention_{W5}$). Consistent with the late-blooming framework, *Tech-class-weight*, *Competition*, and $Attention_{L5}$ are treated as cost-type attributes, for which lower values indicate stronger long-term potential, whereas $Attention_{W5}$ and *Vintage_ diversity* are treated as benefit-type attributes, for which higher values indicate stronger long-term potential.

We compute TOPSIS scores using an expanding-window procedure by patent grant cohort. For each cohort t, the component values are constructed over their predefined measurement windows, including post-grant windows where applicable. For example, $Attention_W5$ is measured over the five years following grant. We then use all patents granted up to and including year t to form the reference distribution for each component and compute the TOPSIS score for patents granted in year t. After normalization, the ideal best vector is defined by assigning 1 to benefit-type attributes and 0 to cost-type attributes. For each patent, we then compute the Euclidean distance to the ideal best solution, D^+, and the

[2] We also construct an alternative composite measure that combines the same dimensions in a simpler additive form. Specifically, $Potential = Attention_{W5} + Vintage_diversity - (Tech\text{-}class\text{-}weight + Competition + Attention_{L5})$. The patent-level and firm-level results obtained from this alternative measure are broadly consistent with those based on the baseline TOPSIS index. Results are available upon request.

Euclidean distance to the ideal worst solution, D^-. The composite potential score is defined as the standard TOPSIS relative closeness measure:

$$Potential_i = \frac{D_i^-}{D_i^+ + D_i^-}, \tag{1}$$

where a higher value indicates that the patent is closer to the profile of a high-potential late-blooming patent.

3.2.5 Control variables

We also incorporate a set of control variables at both the patent and firm levels, following established literature. Patent-level controls include the innovation originality (InnOrig), private valuation (KPSS), the patent significance score (KI index) from Kelly et al. (2021) (5-year forward window), the number of forward citations received in 5-year window after grant (Citation_5), 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). We winsorize all continuous variables at the 1st and 99th percentiles. Detailed definitions and construction methods for all variables are provided in Appendix A.

3.2.6 Prediction-window specification

Figure 1 summarizes the forecasting layout of inputs and outcomes used throughout the paper. At the patent level (Panel A), *Potential* combines four ex-ante components measured at grant—*Tech-class-weight*, *Competition*, *Vintage_ diversity*, and $Attention_{L5}$—with one early post-grant component, $Attention_{W5}$, and is used to predict forward citations measured over non-overlapping $[+5, +10]$, $[+10, +15]$, and $[+15, +20]$ windows as well as a binary $Trade$ indicator for the first assignment occurring more than five years after grant. At the firm level (Panel B), *Potential* is aggregated as a five-year rolling average over patents granted during $[t-4, t]$; because its post-grant attention component extends to $t+5$, we begin outcome measurement at $t+6$ to avoid overlap, and predict ROA, ROE, IVOL, and the SA index at every horizon $k \in [6, 20]$.

3.2.7 Summary statistics

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 TOPSIS-based *Potential* index has a mean of 0.49 and a standard deviation of 0.09. As the measure is bounded between zero and one, the distribution is centered near the middle with relatively modest dispersion. This suggests that most patents do not lie close to either the ideal late-blooming profile or its opposite benchmark, but instead exhibit intermediate levels of long-term potential, indicating that late-blooming patents represent an uncommon subset of the patent population. 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 cita-

tion measures generally decrease over time, from 1.23 to 1.12 in the 15-20 year window. The standard deviation of these log-transformed variables remains relatively stable, ranging from 1.13 to 1.18, indicating similar dispersion across citation windows.

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.

4 Results

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

Forward Citations - As preliminary evidence of the predictive content of our *Potential* index, we sort all sample patents into quintiles based on their *Potential* scores and report average forward citations across quintiles over three non-overlapping long-term horizons: 5–10, 10–15, and 15–20 years after grant. Table 2 Panel A shows a positive association between *Potential* and future citations in every window. The gap between the top and bottom quintiles widens with the horizon, rising from 0.64 (5–10 years, $t = 9.24$) to 1.67 (10–15 years, $t = 20.50$) and 2.22 (15–20 years, $t = 23.29$), consistent with *Potential* forecasting late-blooming citation trajectories.

Panel B of Table 2 reports regression-based predictive tests in which log-transformed forward citations over each window are regressed on *Potential* and a rich set of controls (including KPSS and InnOrig), with year and firm fixed effects and firm-clustered standard errors:

$$\ln(1 + \Delta\text{Citation}_{i,h}) = \alpha \; + \; \beta \, \textit{Potential}_{i,t} \; + \; \boldsymbol{\theta}' \mathbf{X}_{i,t} \; + \; \delta_t \; + \; \gamma_f \; + \; \epsilon_{i,h} \tag{2}$$

where i indexes patents, f indexes the assignee firm, t denotes the grant year, and h denotes the forward-citation window. $\text{Citation}_{i,h}$ is the citation outcome of patent i measured over horizon h. $\textit{Potential}_{i,t}$ is our measure of long-term patent potential. The vector $\mathbf{X}_{i,t}$ includes patent level and firm characteristics defined in section Appendix A. The regression includes grant-year fixed effects (δ_t) and firm fixed effects (γ_f). Standard errors are clustered at the firm level.

The coefficient on *Potential* is positive and highly significant across all three horizons, rising monotonically from 0.29 ($t = 5.77$) at 5–10 years to 0.53 ($t = 8.07$) at 10–15 years and 0.75 ($t = 10.03$) at 15–20 years. Importantly, these predictive effects are estimated after controlling for KPSS and InnOrig, so they reflect variation in future citations that grant-date market valuation and backward-citation-based originality does not capture. The monotone strengthening with horizon indicates *Potential* capturing an accrual component of scientific influence that materializes gradually, supporting Hypothesis 1.

Patent Trade - To test Hypothesis 2, we estimate a logit model for the probability that a patent's first assignment occurs more than five years after grant:

$$\Pr(\text{Trade}_i = 1) = \Lambda\Big(\alpha \; + \; \beta \, \textit{Potential}_{i,t} \; + \; \boldsymbol{\theta}' \mathbf{X}_{i,t} \; + \; \delta_t \; + \; \eta_k + \epsilon_i\Big), \tag{3}$$

where i indexes patents, t denotes the grant year, and k denotes the industry. Trade_i is an indicator equal to one if patent i is traded after 5 years of grant, and zero otherwise. The vector $\mathbf{X}_{i,t}$ includes the same controls as in Equation (2). The specification includes

grant-year fixed effects (δ_t) and industry fixed effects (η_k). Column (4) of Table 2 Panel B reports a coefficient of 0.95 ($t = 18.03$), confirming that high-*Potential* patents are markedly more likely to be traded later in life, consistent with a "market-for-ideas" channel (Gans and Stern, 2003; Arora et al., 2004).

4.2 Long-term patent potential and firm fundamentals

Table 3 reports predictive panel regression results for future firm-level outcomes using *Potential* measure. Following Hirshleifer et al. (2018), we aggregate patent-level *Potential* using a five-year rolling average over the firm's recent patents. Because the post-grant attention component of *Potential* extends to t+5 (see Figure 1), we begin the outcome window at t+6 to avoid mechanical overlap. The specification is

$$Y_{i,t+k} = \alpha + \beta\,\text{Potential}_{i,t} \; + \boldsymbol{\theta}'\mathbf{X}_{i,t} + \delta_t + \eta_j + \epsilon_{i,t+k}, \tag{4}$$

where $Y_{i,t+k}$ denotes one of the future firm-level outcomes: ROA, ROE, IVOL, or SA. In each regression, we include the corresponding lagged dependent variable as an additional control. The vector $\mathbf{X}_{i,t}$ denotes the firm-level patent-related controls and other financial controls.

Panels A and B show that *Potential* is a positive and statistically significant predictor of both ROA and ROE at every horizon from t+6 to t+20. The ROA coefficient peaks at 0.20 ($t = 8.83$) at t+8 and remains significant at t+20 ($\beta = 0.19$, $t = 5.41$); the ROE coefficient peaks at 0.39 ($t = 5.97$) at t+11 and is 0.36 ($t = 4.24$) at t+20. Because the regressions already control for InnOrig—the leading firm-level innovation-quality measure in Hirshleifer et al. (2018)—these coefficients capture long-horizon predictive content beyond that of backward-citation-based originality.

Panels C and D show results for firm-specific risk and financing capacity. The coefficient on

Potential is negative and significant for IVOL at every horizon (e.g., $\beta = -0.85$, $t = -8.85$ at t+10) and for the SA index at every horizon (e.g., $\beta = -1.01$, $t = -4.43$ at t+10; $\beta = -0.68$, $t = -2.34$ at t+20). Firms with higher-*Potential* patent portfolios forecast persistently lower idiosyncratic volatility and weaker financing constraints (SA index) in the long run, consistent with latent innovation assets becoming more economically relevant as their underlying technological domains gain salience.

5 Potential and KPSS: Long-Term Patent Value

This section evaluates whether our *Potential* index captures long-horizon innovative value that is not fully incorporated into market-based patent valuation at the time of grant. We benchmark *Potential* against KPSS, a widely used measure that infers patent value from stock market reactions. Because equity markets are inherently forward-looking, KPSS may already embed investors' expectations about future payoffs. However, such expectations can become less informative over very long horizons, particularly when the value of an invention depends on delayed adoption, complementary developments, and uncertain commercialization paths. If market-based valuation has limited ability to anticipate distant technological activation, patents with low KPSS but high *Potential* can be overlooked in the early post-issuance period (within 5 years), yet go on to deliver strong innovation outcomes ex post.

5.1 Long-term predictive power of Potential and KPSS

To illustrate the strength of the relationship over time, Figure 2 plots the estimated coefficients on *Potential* and KPSS across horizons, where the outcome is annual incremental forward citations measured from year $t+6$ to $t+20$. We estimate the following specification separately for each annual horizon:

$$\ln(1 + \Delta\text{Citation}_{i,\tau}) = \alpha + \beta\, \textit{Valuation}_{i,t} + \boldsymbol{\theta}'\mathbf{X}_{i,t} + \delta_t + \gamma_f + \epsilon_{i,\tau} \tag{5}$$

where i indexes patents, f indexes the assignee firm, t denotes the grant year, and $\tau \in t{+}6, \ldots, t{+}20$ denotes the annual post-grant citation year. $\Delta\text{Citation}_{i,\tau}$ is the number of new citations received by patent i in year τ. $\textit{Valuation}_{i,t}$ is either *Potential* or KPSS, and $\mathbf{X}_{i,t}$ includes same controls as in Equation (2).

The coefficient on *Potential* (dashed line) increases gradually over the horizon. Starting from a relatively modest level in the early years, the coefficient rises steadily and becomes more significant in later years, reaching its highest levels by the end of the sample window. This pattern is consistent with the interpretation that *Potential* captures a delayed and accumulating component of innovative impact, whose relevance becomes more visible as technological trajectories evolve. In contrast, the KPSS measure (solid line) exhibits a different dynamic. Its coefficients are relatively high during the earlier years, peaking in the first half of the sample window, but then decline over time. After approximately year $t + 12$, the KPSS coefficients fall persistently and converge toward zero by year $t + 20$, suggesting that the information embedded in grant-date market valuation becomes less informative about citation outcomes at very long horizons. As a result, the two trajectories intersect around year $t + 17$: *Potential* overtakes KPSS and remains higher thereafter. Overall, Figure 2 highlights a temporal shift in predictive content—KPSS is more closely associated with earlier-to-mid-term citation growth, whereas *Potential* becomes more informative for long-run scientific influence.

5.2 Patent-level validation: long-horizon citations

Table 4 presents the relationship between the *Potential* index and KPSS using conditional double sorts. We first partition patents into High- and Low-KPSS groups based on the sample median of KPSS, and then sort patents into High- and Low-*Potential* groups within each KPSS group using the median of *Potential*.[3] This procedure yields four mutually exclusive quadrants: named HiddenGems (Low KPSS, High Potential), Stars (High KPSS, High Potential), Overhyped (High KPSS, Low Potential), and Laggards (Low KPSS, Low Potential) respectively.

Panel A of Table 4 reports the raw citation differences implied by the double sorts. The first comparison holds KPSS at the low level. *HiddenGems* receive significantly more forward citations than *Laggards* in every horizon: the gap is unit of citation in years 5–10, increases to 1.30 in years 10–15, and remains large at 1.23 in years 15–20. This indicates that, among patents assigned low grant-time market value by KPSS, high-*Potential* patents deliver stronger long-run scientific impact. The second comparison holds KPSS at the high level. *Stars* also outperform *Overhyped* patents across all horizons. Together, these within-KPSS comparisons show that *Potential* complements KPSS by capturing a distinct long-horizon dimension of patent: conditional on either low or high grant-time market valuation, high-*Potential* patents exhibit stronger subsequent citation performance than low-*Potential* patents.

Figure 3 visualizes the same double-sort evidence by plotting average forward citations for the four patent groups across the three long-horizon windows. The figure shows a clear separation between high- and low-*Potential* patents within each KPSS group. Among low-KPSS patents, *HiddenGems* remain well above *Laggards* throughout the horizon. Among high-KPSS patents, *Stars* similarly stay above *Overhyped* patents. The trajectory of *HiddenGems* also illustrates the catch-up pattern: although they initially trail *Stars* in the

[3] Prior to this classification, all sorting variables are residualized with respect to firm and grant-year fixed effects.

5–10 year window, the gap nearly disappears by years 10–15 and remains small in years 15–20. This visual pattern reinforces the table evidence that *Potential* complements KPSS by identifying the long-horizon component of patent value.

Panel B of Table 4 formalizes the double-sort evidence in a regression framework. The omitted group is *Laggards*, so each coefficient measures the citation advantage of the corresponding group relative to Laggards group. The coefficient on *HiddenGems* is positive and highly significant across all horizons, ranging from 0.07 (t-stat=7.14) to 0.09 (t-stat=7.41). This means that, within the low-KPSS category, patents with high *Potential* receive significantly more long-horizon citations than otherwise low *Potential*. The effect also strengthens at the longest horizon. Panel C of Table 4 tests the role of *Potential* within the high-KPSS group by comparing the coefficients on *Stars* and *Overhyped* patents. The differences are positive and statistically significant in all three windows: 0.08, 0.09, and 0.10, with t-statistics of 6.50, 7.59, and 8.90, respectively. These estimates show that, among patents initially assigned high market value by KPSS, those with high *Potential* deliver substantially stronger long-horizon citation performance than those with low *Potential*. These results show that *Potential* identifies late-blooming value among low-KPSS patents and separates more durable scientific impact among high-KPSS patents.

5.3 Firm-level implications: high-Potential patents share and long-run fundamentals

To examine whether this undervaluation pattern has economically meaningful consequences at the firm level, we aggregate patent classifications into a portfolio-style measure of innovation composition. Specifically, we collect granted patents for each firm-year using a five year rolling window and compute the share of *HiddenGems* in the firm's recent patenting activity, *Ratio_ HiddenGems*, defined as the number of *HiddenGems* granted over the past five years scaled by the firm's total patent count in the same period. We also construct an analogous

measure for the *Stars* and *Overhyped* group. We then test whether these measures predict long-run firm fundamentals, risk, and financing capacity.

Table 5 reports forecasts for future ROA (Panel A), ROE (Panel B), IVOL (Panel C), and the SA index (Panel D). The coefficient on *Ratio_HiddenGems* is positive and statistically significant for both profitability measures at every horizon from $t+6$ to $t+20$ (e.g., $\beta = 0.05$, $t = 5.63$ for ROA_{t+20}; $\beta = 0.11$, $t = 5.18$ for ROE_{t+20}). Moreover, *Ratio_HiddenGems* is negative and highly significant for IVOL at every horizon (e.g., $\beta = -0.20$, $t = -4.99$ at $t+20$) and strongly negative for the SA index throughout the forecast horizon (e.g., $\beta = -0.44$, $t = -11.55$ at $t+20$). These results indicate that exposure to patents that appear low-valued under KPSS at grant but have high-*Potential* innovation contains persistent information about future operating performance, firm-specific uncertainty, and financial flexibility.

The comparison between *Stars* and *Overhyped* patents further highlights the role of *Potential* within the high-KPSS group. Firms with a larger share of *Stars* perform better than firms with a larger share of *Overhyped* patents across the main long-run outcomes. *Stars* is also associated with lower IVOL and a lower SA index over most horizons, whereas *Overhyped* is often associated with higher IVOL and becomes positively related to the SA index near the end of the forecast window. For profitability, *Overhyped* is negatively associated with ROA across most horizons, while the corresponding relation for *Stars* is weaker and less persistently negative. This contrast shows that, even among patents favored by KPSS at grant, *Potential* separates portfolios with stronger long-run firm outcomes from those with weaker ones.

6 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.

6.1 Alternative measures of long-term potential

In our baseline analysis, the construction of *Potential* relies on Wall Street Journal-based news attention, specific structural technology characteristics, and inventor team composition. To test the consistency of our results, we construct four alternative measures by redefining its core components:

Structural marginality measures - We first refine the *Tech-class-weight* by using a more granular classification. Specifically, we identify a patent's primary CPC as the 4-digit CPC category that appears most frequently among the CPC codes assigned to the patent; if there is a tie, we choose the category that is listed first. 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 flow-based measure, *Competition_roll*. For each CPC subclass, we calculate the number of patents granted within a three-year rolling window[4] preceding the focal patent's grant date. While the stock-based measure captures the cumulative density of a technological field, this flow-based indicator focuses on contemporaneous entry pressure. By filtering out historical patent counts that may no longer represent active threats, this alternative measure ensures that the competition is driven by the current lack of active entrants rather than merely the historical newness of the field, consistent with the dynamic competition framework discussed in Aghion et al. (2005).

Broadening Public Attention (Google Ngram) - To address concerns that business news captures only a limited range of public attention, we construct an alternative measure using Google Books Ngram data. The corpus reports word frequencies from millions of dig-

[4]We also constructed *Competition_roll* with alternative window lengths, including 5-year and 10-year rolling windows. The results remain consistent and are available upon request.

itized books and has been used to trace changes in intellectual salience and public discourse (Shiller, 2017; Magness and Makovi, 2023; Jha et al., 2025).

We retain the topics identified by Bybee et al. (2024) but replace the baseline topic-attention scores with a keyword-based measure. For each topic, we extract its 50 highest-frequency terms and use a Large Language Model (Gemini-2.5 Pro) to select up to 10 keywords that best represent the topic. The selected keywords are reported in Section Appendix E. We define topic attention as the mean frequency of these keywords in the Google Ngram corpus. Because Ngram frequencies are annual, we use a linear weighting formula to align the series with each patent's grant month. Let T denote the grant year, m the grant month, and $f = (m-1)/12$ the fraction of the grant year elapsed before the grant month. For topic q, let $Attention_{q,T}$ denote its annual Ngram-based attention score in year T. The pre-grant and post-grant cumulative attention measures, $L5$ and $W5$, are calculated as:

$$L5_q = (1-f) \cdot Attention_{q,T-5} + \sum_{k=1}^{4} Attention_{q,T-k} + f \cdot Attention_{q,T} \tag{6}$$

$$W5_q = (1-f) \cdot Attention_{q,T} + \sum_{k=1}^{4} Attention_{q,T+k} + f \cdot Attention_{q,T+5}. \tag{7}$$

Inventor team composition - To verify that the results are not driven by the specific proxy used for inventor-team vintage diversity, we construct an alternative measure based on Jaccard distance. Unlike the baseline cosine-distance measure, which uses the intensity of inventors' patenting activity across years, the Jaccard-distance measure focuses only on the overlap in inventors' active filing years. For each inventor, we first construct the set of active filing years prior to the focal patent's filing year. We then calculate the pairwise Jaccard distance for all inventor pairs within the team, defined as one minus the ratio of the intersection to the union of their prior active-year sets, and average these dyadic distances across all pairs. A larger value indicates that team members' experiences are less overlapped and therefore greater diversity. For single-inventor patents, the measure is set to zero.

Building on the variables defined above, we construct four alternative versions of the *Potential* index by changing one component at a time, holding all other elements of the baseline TOPSIS construction constant:

- *Potential_main*: replaces *Tech-class-weight* with its primary-CPC variant;
- *Potential_comp*: replaces the stock-based *Competition* with the flow-based *Competition_roll*;
- *Potential_book*: replaces WSJ-based $Attention_{L5}/Attention_{W5}$ with their Google Ngram counterparts;
- *Potential_vintage*: replaces cosine-based *Vintage_diversity* with the Jaccard-distance variant.

Panels A–D of Table 6 report the patent-level robustness results for the four alternative indices. The predictive coefficients remain positive and statistically significant for long-horizon forward citations and patent trading across every alternative, and the strengthening-with-horizon pattern is preserved.

6.2 Exclusion of pharmaceutical firms

A second concern is that our results could be disproportionately driven by the pharmaceutical sector, whose long development cycles naturally produce late-blooming patterns (Grabowski, 2002; Hall et al., 2001). We re-estimate the baseline models after excluding firms in Fama-French Industry 13 (Pharmaceutical Products; SIC 2830–2836).

Panel E of Table 6 shows that the coefficients on *Potential* remain positive and highly significant at every citation horizon and for *Trade*. The horizon-increasing pattern survives (citation coefficient rises from 0.31 at 5–10 years to 0.68 at 15–20 years, $t = 9.45$; $\beta = 0.78$, $t = 14.51$ for Trade), so the predictive content of *Potential* is not a pharmaceutical-specific artifact.

6.3 Subsample analysis: pre- and post-2000

We next split the sample at the year 2000 to address structural shifts in the innovation landscape. The Internet stock boom peaked around March 2000 and was followed by the dot-com collapse, making this period a natural breakpoint in the valuation and commercialization environment for Internet-related technologies (Ofek and Richardson, 2003). In our patent data, the combined share of G06 and H04 (Internet-related CPC classes) patents rises from about 6% in the early 1980s to roughly one quarter of all granted patents by 1999–2001, and continues to increase thereafter. Since information-technology patents represent a large and growing share of the patent universe around 2000, the pre- and post-2000 subsamples provide a useful test of whether the predictive content of *Potential* is robust across different technology-market regimes. Panel F of Table 6 shows that the citation coefficients are of comparable magnitude across sub-periods, while the $Trade$ coefficient is substantially larger post-2000 ($\beta = 1.00$, $t = 15.59$) than pre-2000 ($\beta = 0.49$, $t = 4.83$). The stronger trading response in the post-2000 period is consistent with the maturation of the market for technology.

6.4 Firm-level robustness: alternative measures, pharmaceutical exclusion, and subsamples

We repeat the firm-level analysis under the same three robustness exercises and report the ROA results in Table 7. The results for ROE, IVOL, and the SA index are consistent with ROA and are reported in the Appendix B.

Panel A of Table 7 shows that all four alternative constructions of *Potential* deliver predictive coefficients of the expected sign and comparable magnitude. Panel B, which excludes pharmaceutical firms, retains sign and significance on every outcome at essentially every horizon. Panel C (pre- vs. post-2000) shows that the negative predictive effects on IVOL and SA remain highly significant in both sub-periods and across the full horizon; the positive effects

on ROA and ROE are robust in the pre-2000 period and attenuate at very long horizons in the post-2000 sub-sample.

7 Conclusion

This paper proposes a new approach to measuring long-horizon innovation value. We construct a *Potential* index that integrates structural marginality, inventor-team heterogeneity, and technology salience. This combination reflects the idea that late-blooming patents are not defined by marginality alone: they are initially overlooked, contain experience-based foundations that can support broader downstream applications, and become valuable as their technological domains gain salience. We validate the predictive power of this index using forward citations, patent trading, and long-run firm outcomes. Across horizons up to two decades, patents with higher *Potential* generate more future citations and are more likely to be traded. At the firm level, a higher average *Potential* of a firm's recent patents predicts higher future profitability (ROA and ROE), lower idiosyncratic volatility, and reduced financing constraints. In addition, we benchmark *Potential* against KPSS. While KPSS may incorporate forward-looking information, its predictive content weakens at very long horizons. In contrast, the statistical significance of *Potential* strengthens with the horizon, producing a scissor-shaped pattern in which *Potential* overtakes KPSS in forecasting distant citation growth. Consistent with this pattern, conditional double-sorts identify *HiddenGems*: patents with low KPSS but high *Potential*. These patents eventually catch up to *Stars* patents in long-run scientific impact. Furthermore, the share of high-*Potential* patents in a firm's innovation portfolio predicts superior long-run firm fundamentals across the dimensions we examined. These findings highlight a value-discovery role of *Potential*: it flags late-blooming innovations that appear undervalued at issuance yet deliver substantial long-run impact.

This evidence implies that innovation value can be realized gradually and unevenly over

time. For investors and managers, the results suggest that long-horizon opportunities can be found not only in inventions that receive high market valuation at grant, but also in initially overlooked patents whose relevance rises as technological trends evolve.

References

Aghion, P., Bloom, N., Blundell, R., Griffith, R., and Howitt, P. (2005). Competition and innovation: An inverted-u relationship. *The Quarterly Journal of Economics*, 120(2):701–728.

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., Fosfuri, A., and Gambardella, A. (2004). *Markets for Technology: The Economics of Innovation and Corporate Strategy*. MIT Press, Cambridge, MA.

Baruffaldi, S. H. and Simeth, M. (2020). Patents and knowledge diffusion: The effect of early disclosure. *Research Policy*, 49(4):103927.

Bloom, N. and Van Reenen, J. (2002). Patents, real options and firm performance. *The Economic Journal*, 112(478):C97–C116.

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.

Chen, A. Y. and Zimmermann, T. (2022). Open source cross-sectional asset pricing. *Critical Finance Review*, pages 207–264.

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.

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.

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

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.

Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study. *Research Policy*, 31(8-9):1257–1274.

Glode, V. and Ordonez, G. (2025). Technological progress and rent seeking. *The Review of Financial Studies*, 38(4):1259–1289.

Grabowski, H. (2002). Patents, innovation and access to new pharmaceuticals. *Journal of International Economic Law*, 5(4):849–860.

Hadlock, C. J. and Pierce, J. R. (2010). New evidence on measuring financial constraints: Moving beyond the kz index. *The Review of Financial Studies*, 23(5):1909–1940.

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.

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.

Hochberg, Y. V., Serrano, C. J., and Ziedonis, R. H. (2018). Patent collateral, investor commitment, and the market for venture lending. *Journal of Financial Economics*, 130(1):74–94.

Hou, J. and Yang, X. (2019). Patent sleeping beauties: evolutionary trajectories and identification methods. *Scientometrics*, 120(1):187–215.

Huo, D., Motohashi, K., and Gong, H. (2019). Team diversity as dissimilarity and variety in organizational innovation. *Research Policy*, 48(6):1564–1572.

Hwang, C.-L. and Yoon, K. (2012). *Multiple attribute decision making: methods and applications a state-of-the-art survey*. Springer Science & Business Media.

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.

Jones, B. F. (2009). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *The Review of Economic Studies*, 76(1):283–317.

Jones, B. F. (2010). Age and great invention. *The Review of Economics and Statistics*, 92(1):1–14.

Kaltenberg, M., Jaffe, A. B., and Lachman, M. E. (2023). Invention and the life course: Age differences in patenting. *Research Policy*, 52(1):104629.

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.

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.

Mann, W. (2018). Creditor rights and innovation: Evidence from patent collateral. *Journal of Financial Economics*, 130(1):25–47.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization science*, 2(1):71–87.

Ofek, E. and Richardson, M. (2003). Dotcom mania: The rise and fall of internet stock prices. *the Journal of Finance*, 58(3):1113–1137.

Schwartz, E. S. (2004). Patents and r&d as real options. *Economic Notes*, 33(1):23–54.

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.

Suh, P. (2023). Intellectual property rights and debt financing. *The Review of Financial Studies*, 36(5):1970–2003.

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.

Wuchty, S., Jones, B. F., and Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827):1036–1039.

Yang, S. (2023). Predictive patentomics: Forecasting innovation success and valuation with chatgpt. *arXiv preprint arXiv:2307.01202*.

Yang, Y. C., Zhang, B., and Zhang, C. (2020). Is information risk priced? evidence from abnormal idiosyncratic volatility. *Journal of Financial Economics*, 135(2):528–554.

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

Panel A: Patent-level

	N	Mean	SD	Min	Max
Potential	987,572	0.49	0.09	0.01	0.75
Citation_5	987,572	1.02	0.96	0.00	6.65
Citation5_10	920,396	1.23	1.13	0.00	7.85
Citation10_15	636,866	1.19	1.18	0.00	7.51
Citation15_20	437,694	1.12	1.18	0.00	7.62
Trade(binary)	773,333	0.15	0.36	0.00	1.00
InnOrig	987,572	0.52	0.29	0.00	2.94
KPSS	987,572	2.13	1.02	0.01	4.94
KI5	897,731	0.70	0.26	0.00	2.93
Num_claims	987,572	2.82	0.61	0.69	6.77
Num_citations	987,572	2.39	1.00	0.69	7.82
Num_nonpat_citations	987,572	1.05	1.23	0.00	7.82
Num_inventors	987,572	1.23	0.43	0.69	4.34
Num_ipc	987,572	0.97	0.42	0.00	5.36
Grant_lag	987,572	6.83	0.54	5.50	8.00

Panel B: Firm-level

	N	Mean	SD	Min	Max
Potential	27,504	0.78	1.52	-10.37	5.19
InnOrig	27,504	0.73	0.34	0	2.00
CIE	27,504	0.61	1.21	0	9.30
PIE	27,504	0.19	0.36	0	3.02
ADV	27,504	0.01	0.03	0	0.25
CAPEX	27,504	0.05	0.05	0	0.43
R&D	27,504	0.08	0.13	0	1.08
BTM	27,504	0.56	0.48	0.02	7.81
ROE	25,533	-0.01	0.38	-2.84	0.87
ROA	27,504	0.06	0.23	-1.63	0.48
IVOL	27,137	0.03	0.02	0	2.50
SA_index	11,783	-2.89	0.75	-4.64	4.27
Asset	25,850	5.96	2.00	3.56	13.26
Sales growth	25,850	0.15	0.31	-0.40	0.89
PP&E	25,850	0.43	0.31	0.05	1.87
LEV	25,850	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. 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). KI are constructed following Kelly et al. (2021) using forward-looking windows of 5. Num_claims, Num_citations, Num_nonpat_citations, Num_inventors, and Num_ipc denote the number of claims, citations to other patents, non-patent references, inventors, and IPC classifications (in logs). 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 are citations-based innovative efficiency and patents-based innovative efficiency following (Hirshleifer et al., 2013). Financial controls include: ADV (advertising to assets), CAPEX (capital expenditures to assets), R&D (R&D expenditures to assets), BTM (book-to-market ratio), ROE (return on equity), ROA (return on asset), 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, Forward Citations, and Patent Transactions

Panel A: Univariate Sorts of Forward Citations by Patent Potential

	High_Potential	Low_Potential	Difference	T-statistic
Citation5_10	6.70	6.06	0.64***	9.24
Citation10_15	7.51	5.84	1.67***	20.50
Citation15_20	7.47	5.25	2.22***	23.29

Panel B: Regression Results of Forward Citations and Trades

	(1) Citation5_10	(2) Citation10_15	(3) Citation15_20	(4) Trade
Potential	0.2879***	0.5287***	0.7510***	0.9507***
	(5.77)	(8.07)	(10.03)	(18.03)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	No
Industry FE	No	No	No	Yes
Observations	919,936	636,418	437,264	768,910
Adj/Pseudo R^2	0.5024	0.3682	0.3200	0.1783

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. All specifications in Panel B include the same set of control variables, including both patent-level characteristics (KPSS, InnOrig, number of claims, number of backward citations, number of non-patent citations, number of inventors, number of IPC classes, and grant lag days) and firm-level financial controls (Assets, Sales growth, PP&E, R&D, Capex, Book-to-market ratio, Leverage ratio, and ROA). Definitions are in Section Appendix A. Standard errors are clustered at the firm level for citation specifications, and robust t-statistics are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 3: Long-term Patent Potential and Future Firm Performance

	t+6	t+7	t+8	t+9	t+10	t+11	t+12	t+13	t+14	t+15	t+16	t+17	t+18	t+19	t+20
Panel A: Return on Assets (ROA_{t+k})															
Potential	0.160***	0.174***	0.196***	0.194***	0.178***	0.168***	0.154***	0.146***	0.136***	0.120***	0.130***	0.141***	0.142***	0.167***	0.192***
	(8.08)	(8.20)	(8.83)	(8.25)	(7.33)	(6.54)	(5.70)	(5.20)	(4.68)	(3.94)	(4.16)	(4.44)	(4.34)	(4.87)	(5.41)
Adj. R^2	0.4286	0.4003	0.3780	0.3665	0.3538	0.3403	0.3259	0.3158	0.3128	0.3063	0.3017	0.3048	0.3049	0.2952	0.2920
Total Obs.	26,515	24,184	21,974	19,965	18,135	16,535	15,102	13,769	12,519	11,364	10,287	9,266	8,356	7,511	6,693
Panel B: Return on Equity (ROE_{t+k})															
Potential	0.254***	0.311***	0.343***	0.303***	0.356***	0.386***	0.330***	0.299***	0.253***	0.192***	0.174**	0.203***	0.285***	0.287***	0.356***
	(4.72)	(5.85)	(6.13)	(5.31)	(5.86)	(5.97)	(5.55)	(4.76)	(3.80)	(2.72)	(2.31)	(2.71)	(3.63)	(3.56)	(4.24)
Adj. R^2	0.1832	0.1843	0.1658	0.1698	0.1569	0.1478	0.1595	0.1513	0.1582	0.1619	0.1479	0.1552	0.1540	0.1443	0.1405
Total Obs.	24,886	22,697	20,621	18,748	17,047	15,544	14,174	12,911	11,741	10,654	9,643	8,698	7,846	7,057	6,298
Panel C: Idiosyncratic Volatility ($IVOL_{t+k}$)															
Potential	-0.711***	-0.792***	-0.777***	-0.843***	-0.850***	-0.812***	-0.736***	-0.722***	-0.737***	-0.734***	-0.610***	-0.721***	-0.528***	-0.541***	-0.636***
	(-9.43)	(-10.12)	(-9.25)	(-9.39)	(-8.85)	(-7.98)	(-6.84)	(-6.42)	(-6.20)	(-5.86)	(-4.65)	(-5.30)	(-3.72)	(-3.57)	(-4.00)
Adj. R^2	0.4245	0.4173	0.4045	0.3964	0.3827	0.3795	0.3738	0.3739	0.3688	0.3539	0.3514	0.3528	0.3473	0.3364	0.3314
Total Obs.	27,137	25,564	23,447	21,376	19,508	17,809	16,316	14,944	13,644	12,443	11,311	10,273	9,291	8,410	7,562
Panel D: Financial Constraints (SA_{t+k})															
Potential	-0.981***	-0.980***	-1.023***	-1.027***	-1.015***	-1.010***	-0.985***	-0.945***	-0.908***	-0.804***	-0.759***	-0.745***	-0.793***	-0.769**	-0.679**
	(-6.07)	(-5.53)	(-5.01)	(-4.77)	(-4.43)	(-4.12)	(-3.90)	(-3.83)	(-3.59)	(-3.18)	(-3.05)	(-2.78)	(-2.86)	(-2.64)	(-2.34)
Adj. R^2	0.4486	0.4235	0.4058	0.3861	0.3676	0.3488	0.3332	0.3092	0.2921	0.2886	0.2829	0.2791	0.2792	0.2748	0.2751
Total Obs.	11,783	10,587	9,482	8,501	7,631	6,841	6,123	5,463	4,843	4,272	3,748	3,266	2,831	2,435	2,069
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ind. & Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports regression results predicting future firm performance using *Potential*. The dependent variables are the firm-level Return on Assets (ROA) in **Panel A**, Return on Equity (ROE) in **Panel B**, Idiosyncratic Volatility (IVOL) in **Panel C**, and the SA index in **Panel D**, all measured in year $t + k$ where k ranges from 6 to 20. All specifications include the same set of control variables, including firm-level KPSS, InnOrig, CIE, PIE, Advertising expense, R&D expense, Book-to-Market ratio, Capital expenditure, ROEt, ROAt,and IVOLt. Definition of variables are available in Section Appendix A. t-statistics based on standard errors clustered at the firm level are reported in parentheses. All independent variables are winsorized at the 1% and 99% levels.

Table 4: Supplementing KPSS: The Long-Term Value of Potential

Panel A: Long-term Citations by Patent Potential and KPSS Groups							
	Full Sample	HiddenGems	Stars	Overhyped	Laggards	HiddenGems vs. Laggards	Stars vs. Overhyped
Citation5_10	6.68	6.94	7.53	6.34	5.94	1.01***	1.19***
Citation10_15	7.07	7.74	7.77	6.33	6.43	1.30***	1.44***
Citation15_20	6.09	6.70	6.82	5.38	5.47	1.23***	1.44***

Panel B: Regression Results of Patent Potential and KPSS Groups			
	(1) Citation5-10	(2) Citation10-15	(3) Citation15-20
HiddenGems	0.0737***	0.0735***	0.0872***
	(7.18)	(7.14)	(7.42)
Star	0.0399*	0.1001***	0.1170***
	(1.74)	(6.25)	(7.00)
Overhyped	-0.0360	0.0144	0.0193
	(-1.32)	(1.07)	(1.50)
Controls	Yes	Yes	Yes
Year & Firm FE	Yes	Yes	Yes
Observations	919,936	636,418	437,264
Adj R^2	0.2149	0.2254	0.2244
Panel C: Difference of Coefficients			
Diff of Star and Overhyped	0.076***	0.086***	0.098***
T-statistics	(6.50)	(7.59)	(8.90)

Notes: This table compares the long-term performance of patents classified by the *Potential* index and KPSS valuation. We employ a conditional double-sorting strategy using residualized measures of *Potential* and KPSS (net of firm and grant-year fixed effects). Patents are first divided into High/Low groups based on the sample median of KPSS. Within each KPSS group, patents are further sorted into High/Low *Potential* subgroups using the group-specific median. This procedure yields four mutually exclusive quadrants: HiddenGems (Low KPSS, High Potential), Stars (High KPSS, High Potential), Overhyped (High KPSS, Low Potential), and Laggards (Low KPSS, Low Potential). **Panel A** reports the mean forward citations for the full sample and each quadrant. The last two columns report the mean differences between HiddenGems and Laggards, and between Stars and Overhyped. **Panel B** presents regression results where the independent variables are dummy indicators for the HiddenGems, Stars, and Overhyped groups. The omitted reference group is Laggards. Dependent variables are log-transformed forward citations over 5–10, 10–15, and 15–20 years. **Panel C** reports the differences in estimated coefficients between the Stars and Overhyped groups from Panel B, along with their corresponding t-statistics. Statistical significance at the 1%, 5%, and 10% levels is indicated by ***, , and *, respectively.

Table 5: High-Potential Patent Share and Future Firm Performance

	t+6	t+7	t+8	t+9	t+10	t+11	t+12	t+13	t+14	t+15	t+16	t+17	t+18	t+19	t+20
Panel A: Return on Assets (ROA_{t+k})															
Ratio_HiddenGem	0.0354***	0.0381***	0.0396***	0.0351***	0.0329***	0.0290***	0.0244***	0.0233***	0.0282***	0.0342***	0.0403***	0.0430***	0.0449***	0.0450***	0.0501***
	(6.44)	(6.63)	(6.62)	(5.67)	(5.24)	(4.45)	(3.63)	(3.38)	(3.97)	(4.58)	(5.17)	(5.28)	(5.10)	(4.92)	(5.63)
Ratio_Stars	-0.0072	-0.0136**	-0.0201***	-0.0229***	-0.0198***	-0.0216***	-0.0149**	-0.0055	-0.0068	-0.0080	-0.0089	-0.0149	-0.0124	-0.0129	-0.0081
	(-1.41)	(-2.51)	(-3.49)	(-3.76)	(-3.11)	(-3.21)	(-2.09)	(-0.73)	(-0.86)	(-0.96)	(-1.01)	(-1.59)	(-1.22)	(-1.23)	(-0.80)
Ratio_Overhyped	-0.0160***	-0.0220***	-0.0228***	-0.0295***	-0.0359***	-0.0407***	-0.0436***	-0.0499***	-0.0470***	-0.0422***	-0.0347***	-0.0219**	-0.0192*	-0.0164	-0.0219**
	(-2.99)	(-3.89)	(-3.82)	(-4.70)	(-5.50)	(-5.91)	(-6.00)	(-6.57)	(-5.93)	(-4.99)	(-3.89)	(-2.35)	(-1.91)	(-1.59)	(-2.22)
Adj. R^2	0.4193	0.3953	0.3704	0.3533	0.3404	0.3293	0.3153	0.3057	0.2993	0.2991	0.2911	0.2887	0.2833	0.2811	0.2906
Total Obs.	31,682	28,913	26,313	23,900	21,747	19,826	18,095	16,529	15,048	13,686	12,424	11,221	10,133	9,127	8,185
Panel B: Return on Equity (ROE_{t+k})															
Ratio_HiddenGem	0.1265***	0.1249***	0.1206***	0.0999***	0.1031***	0.1039***	0.1031***	0.1040***	0.0873***	0.0636***	0.0696***	0.0632***	0.0765***	0.0740***	0.1071***
	(7.78)	(8.39)	(8.15)	(6.65)	(6.72)	(6.51)	(6.89)	(6.80)	(5.42)	(3.74)	(3.85)	(3.44)	(3.85)	(3.65)	(5.18)
Ratio_Stars	0.0088	-0.0046	-0.0080	-0.0134	-0.0023	0.0106	0.0281*	0.0424***	0.0427**	0.0268	0.0301	0.0173	0.0178	0.0362	0.0353
	(0.58)	(-0.33)	(-0.56)	(-0.91)	(-0.15)	(0.65)	(1.79)	(2.58)	(2.42)	(1.42)	(1.49)	(0.84)	(0.79)	(1.56)	(1.51)
Ratio_Overhyped	0.0111	0.0003	-0.0177	-0.0293*	-0.0342**	-0.0381**	-0.0294*	-0.0257	-0.0327*	-0.0273	-0.0037	0.0308	0.0414*	0.0371	0.0178
	(0.70)	(0.02)	(-1.19)	(-1.92)	(-2.15)	(-2.26)	(-1.82)	(-1.53)	(-1.82)	(-1.42)	(-0.18)	(1.47)	(1.84)	(1.63)	(0.78)
Adj. R^2	0.1614	0.1745	0.1644	0.1590	0.1519	0.1419	0.1495	0.1461	0.1502	0.1519	0.1370	0.1412	0.1428	0.1400	0.1453
Total Obs.	29,684	27,095	24,662	22,436	20,437	18,644	17,008	15,525	14,137	12,850	11,658	10,535	9,515	8,582	7,705
Panel C: Idiosyncratic Volatility ($IVOL_{t+k}$)															
Ratio_HiddenGem	-0.1855***	-0.1927***	-0.1886***	-0.1942***	-0.1840***	-0.1775***	-0.2015***	-0.1869***	-0.1793***	-0.1616***	-0.1792***	-0.1768***	-0.1776***	-0.2119***	-0.2047***
	(-8.51)	(-8.50)	(-7.94)	(-7.86)	(-7.12)	(-6.62)	(-7.15)	(-6.40)	(-5.86)	(-5.08)	(-5.34)	(-5.06)	(-4.89)	(-5.48)	(-4.99)
Ratio_Stars	-0.0812***	-0.0852***	-0.0546**	-0.0489**	-0.0502**	-0.0404	-0.0477*	-0.0408	-0.0524	-0.0392	-0.0429	-0.0450	0.0287	0.0800*	0.0835*
	(-4.03)	(-4.07)	(-2.45)	(-2.07)	(-1.99)	(-1.50)	(-1.65)	(-1.33)	(-1.59)	(-1.12)	(-1.14)	(-1.14)	(0.69)	(1.80)	(1.78)
Ratio_Overhyped	-0.0277	0.0081	0.0187	0.0314	0.0788***	0.1004***	0.0797***	0.0749**	0.1120***	0.1147***	0.1377***	0.1331***	0.0876**	0.0237	0.0651
	(-1.32)	(0.37)	(0.80)	(1.27)	(3.00)	(3.62)	(2.68)	(2.38)	(3.33)	(3.22)	(3.63)	(3.34)	(2.10)	(0.53)	(1.40)
Adj. R^2	0.3995	0.3906	0.3788	0.3721	0.3604	0.3570	0.3482	0.3470	0.3364	0.3300	0.3251	0.3239	0.3224	0.3178	0.3092
Total Obs.	31,845	29,966	27,502	25,107	22,937	20,963	19,196	17,590	16,102	14,683	13,400	12,185	11,065	10,026	9,057
Panel D: Financial Constraints (SA_{t+k})															
Ratio_HiddenGem	-0.4897***	-0.4739***	-0.4592***	-0.4364***	-0.4161***	-0.4064***	-0.4043***	-0.4021***	-0.3950***	-0.4129***	-0.4232***	-0.4322***	-0.4352***	-0.4422***	-0.4406***
	(-25.18)	(-23.43)	(-21.84)	(-20.09)	(-18.44)	(-17.30)	(-16.46)	(-15.49)	(-14.53)	(-14.58)	(-14.11)	(-13.69)	(-12.98)	(-12.36)	(-11.55)
Ratio_Stars	-0.1166***	-0.0948***	-0.0860***	-0.0849***	-0.0846***	-0.0832***	-0.0972***	-0.1191***	-0.1197***	-0.1066***	-0.1026***	-0.1032**	-0.1155***	-0.1174**	-0.1290***
	(-6.24)	(-4.78)	(-4.08)	(-3.81)	(-3.55)	(-3.27)	(-3.57)	(-4.02)	(-3.76)	(-3.15)	(-2.77)	(-2.57)	(-2.66)	(-2.57)	(-2.70)
Ratio_Overhyped	-0.0754***	-0.0685***	-0.0526**	-0.0323	-0.0061	0.0079	0.0105	0.0186	0.0209	0.0161	0.0190	0.0294	0.0711	0.1143**	0.1700***
	(-3.77)	(-3.22)	(-2.34)	(-1.36)	(-0.24)	(0.29)	(0.36)	(0.58)	(0.61)	(0.44)	(0.48)	(0.69)	(1.56)	(2.40)	(3.43)
Adj. R^2	0.4612	0.4350	0.4152	0.3943	0.3736	0.3556	0.3367	0.3143	0.2951	0.2888	0.2810	0.2771	0.2765	0.2798	0.2854
Total Obs.	13,890	12,488	11,186	10,037	9,036	8,120	7,274	6,505	5,771	5,100	4,484	3,924	3,413	2,947	2,527
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ind. & Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Firm-level patent metrics are aggregated using a 5-year rolling window. First, all granted patents are classified into four mutually exclusive groups based on the independent double-sorting of KPSS and *Potential* (as detailed in Table 5). We then calculate the share of each patent type for a firm in year t (eg. *Ratio_ HiddenGems* is defined as the number of *HiddenGems* granted to the firm over the past five years scaled by the firm's total patent count in the same period). The dependent variables are the firm-level Return on Assets (ROA) in **Panel A**, Return on Equity (ROE) in **Panel B**, Idiosyncratic Volatility (IVOL) in **Panel C**, and the SA index in **Panel D**, in year $t + k$. The horizon ranges k from 6 to 20 years. All controls are the same as main results. Significance levels are denoted by *, **, and *** for 10%, 5%, and 1% levels, respectively.

Table 6: Robustness: Alternative Measures of Potential and Long-term Patent Outcomes

	(1) Citation5-10	(2) Citation10-15	(3) Citation15-20	(4) Trade
Panel A: Alternative Measure - Potential_main				
Potential_main	0.2702***	0.5008***	0.7079***	0.7411***
	(5.03)	(7.36)	(9.08)	(13.92)
Observations	919,936	636,418	437,264	768,910
Adj. R-squared	0.5024	0.3681	0.3198	0.1781
Panel B: Alternative Measure - Potential_comp				
Potential_comp	0.1798***	0.3531***	0.5623***	0.9844***
	(3.49)	(5.79)	(8.07)	(18.60)
Observations	919,936	636,418	437,264	768,910
Adj. R-squared	0.5022	0.3678	0.3193	0.1783
Panel C: Alternative Measure - Potential_book				
Potential_book	0.3010***	0.5231***	0.7065***	0.8749***
	(5.98)	(8.48)	(9.77)	(18.10)
Observations	1,032,680	804,906	599,382	804,336
Adj. R-squared	0.4801	0.3560	0.3013	0.1463
Panel D: Alternative Measure - Potential_vintage				
Potential_vintage	0.2380***	0.3968***	0.5494***	1.0390***
	(5.35)	(6.56)	(6.57)	(18.82)
Observations	919,936	636,418	437,264	768,910
Adj. R-squared	0.5023	0.3679	0.3192	0.1781
Panel E: Exclude Pharmaceutical Firms				
Potential	0.3149***	0.5063***	0.6766***	0.7843***
	(6.29)	(7.85)	(9.45)	(14.51)
Observations	866,580	595,010	406,491	725,154
Adj. R-squared	0.5003	0.3677	0.3176	0.1818
Panel F: Subsample Analysis				
Year < 2000				
Potential	0.3539***	0.5546***	0.7538***	0.4885***
	(5.18)	(6.11)	(7.42)	(4.83)
Adj. R-squared	0.4650	0.3603	0.3195	0.1363
Year ≥ 2000				
Potential	0.2774***	0.4815***	0.6281***	0.9951***
	(6.15)	(8.10)	(10.25)	(15.59)
Adj. R-squared	0.5120	0.3786	0.3252	0.2168
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	No
Industry FE	No	No	No	Yes

Notes: This table reports results using *Potential_book*, *Potential_main*, *Potential_comp*, and *Potential_vintage* respectively. As detailed in Section 6, each alternative measure is a composite index constructed by substituting specific components of the original *Potential*. *Potential_book* replaces the Wall Street Journal-based pre- and post-grant attention measures with Google Ngram-based attention measures. *Potential_main* replaces *Tech-class-weight* with a primary-CPC-based variant. *Potential_comp* replaces the stock-based *Competition* measure with the rolling flow-based *Competition_roll*. *Potential_vintage* replaces the cosine-distance-based *Vintage_diversity* measure with a Jaccard-distance-based variant. **Panel E** excludes pharmaceutical firms' patents, defined as Industry 13 in the Fama-French 49-industry classification. **Panel F** conducts subsample analysis by splitting the sample into two periods: patents granted before 2000 (Sub-panel B1) and patents granted in or after 2000 (Sub-panel B2). The dependent variables and the set of control variables are the same as Table 2. Standard errors are clustered at the firm level for citation specifications, and t-statistics are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 7: Robustness: Alternative Long-term Patent Potential and Future Firm Performance (ROA)

ROA_{t+k}	t+6	t+7	t+8	t+9	t+10	t+11	t+12	t+13	t+14	t+15	t+16	t+17	t+18	t+19	t+20
Panel A: Alternative Measures of Potential															
Potential_main	0.1227***	0.1269***	0.1541***	0.1556***	0.1488***	0.1416***	0.1290***	0.1366***	0.1157***	0.1038***	0.1109***	0.1253***	0.1160***	0.1395***	0.1759***
	(6.20)	(5.95)	(6.90)	(6.58)	(6.07)	(5.45)	(4.73)	(4.80)	(3.94)	(3.35)	(3.49)	(3.90)	(3.49)	(4.00)	(4.88)
Adj. R^2	0.4280	0.3995	0.3771	0.3657	0.3532	0.3398	0.3254	0.3156	0.3124	0.3061	0.3014	0.3045	0.3043	0.2945	0.2914
Obs.	26,515	24,184	21,974	19,965	18,135	16,535	15,102	13,769	12,519	11,364	10,287	9,266	8,356	7,511	6,693
Potential_comp	0.1478***	0.1649***	0.1868***	0.1821***	0.1676***	0.1564***	0.1397***	0.1329***	0.1254***	0.1116***	0.1245***	0.1400***	0.1440***	0.1702***	0.1991***
	(7.61)	(7.90)	(8.55)	(7.87)	(7.01)	(6.16)	(5.24)	(4.79)	(4.38)	(3.70)	(4.02)	(4.47)	(4.44)	(5.00)	(5.64)
Adj. R^2	0.4284	0.4002	0.3779	0.3663	0.3537	0.3401	0.3257	0.3156	0.3126	0.3062	0.3016	0.3048	0.3050	0.2953	0.2922
Obs.	26,515	24,184	21,974	19,965	18,135	16,535	15,102	13,769	12,519	11,364	10,287	9,266	8,356	7,511	6,693
Potential_book	0.1581***	0.1690***	0.1902***	0.1915***	0.1855***	0.1732***	0.1599***	0.1540***	0.1485***	0.1442***	0.1658***	0.1883***	0.1818***	0.1799***	0.1938***
	(8.42)	(8.41)	(9.11)	(8.74)	(8.19)	(7.24)	(6.36)	(5.87)	(5.52)	(5.11)	(5.64)	(6.12)	(5.54)	(5.25)	(5.83)
Adj. R^2	0.4193	0.3921	0.3730	0.3644	0.3510	0.3382	0.3237	0.3110	0.3096	0.3070	0.3011	0.2988	0.2908	0.2852	0.2891
Obs.	28,150	25,888	23,714	21,647	19,694	17,951	16,394	14,957	13,620	12,391	11,252	10,160	9,184	8,282	7,420
Potential_vintage	0.1358***	0.1550***	0.1811***	0.1905***	0.1673***	0.1481***	0.1297***	0.1164***	0.1062***	0.1004***	0.1279***	0.1548***	0.1593***	0.1983***	0.2325***
	(6.53)	(6.93)	(7.75)	(7.71)	(6.56)	(5.47)	(4.56)	(3.93)	(3.47)	(3.11)	(3.85)	(4.61)	(4.58)	(5.42)	(6.12)
Adj. R^2	0.4281	0.3998	0.3775	0.3662	0.3535	0.3398	0.3254	0.3152	0.3122	0.3060	0.3015	0.3049	0.3051	0.2957	0.2928
Obs.	26,515	24,184	21,974	19,965	18,135	16,535	15,102	13,769	12,519	11,364	10,287	9,266	8,356	7,511	6,693
Panel B: Panel Regression (Exclude Pharmaceutical Firms)															
Potential	0.1440***	0.1555***	0.1693***	0.1627***	0.1307***	0.1242***	0.1107***	0.1069***	0.1142***	0.1177***	0.1162***	0.1117***	0.1142***	0.1264***	0.1461***
	(8.44)	(8.55)	(8.84)	(8.20)	(6.60)	(5.97)	(5.10)	(4.72)	(4.83)	(4.75)	(4.48)	(4.14)	(4.03)	(4.26)	(4.77)
Adj. R^2	0.4011	0.3595	0.3238	0.2957	0.2801	0.2729	0.2633	0.2530	0.2438	0.2376	0.2266	0.2297	0.2298	0.2215	0.2215
Obs.	23,958	21,885	19,912	18,122	16,488	15,056	13,772	12,576	11,454	10,414	9,447	8,528	7,705	6,940	6,198
Panel C: Subsample Analysis by Time Period															
Year < 2000															
Potential	0.1875***	0.2441***	0.2427***	0.2550***	0.2270***	0.2141***	0.1819***	0.1531***	0.1457***	0.1104**	0.1299***	0.1717***	0.1877***	0.2017***	0.2375***
	(6.54)	(7.46)	(6.97)	(6.78)	(5.99)	(5.44)	(4.53)	(3.79)	(3.53)	(2.57)	(3.08)	(4.17)	(4.70)	(5.04)	(5.73)
Adj. R^2	0.3850	0.3441	0.3168	0.2925	0.2816	0.2567	0.2342	0.2275	0.2381	0.2348	0.2367	0.2519	0.2596	0.2575	0.2570
Obs.	9,786	9,138	8,517	7,941	7,432	7,010	6,655	6,323	6,007	5,722	5,431	5,165	4,935	4,730	4,532
Year ≥ 2000															
Potential	0.1471***	0.1328***	0.1677***	0.1581***	0.1479***	0.1343***	0.1314***	0.1423***	0.1414***	0.1458***	0.1433***	0.1135**	0.0798	0.1070*	0.0872
	(5.46)	(4.71)	(5.72)	(5.18)	(4.61)	(3.89)	(3.56)	(3.61)	(3.43)	(3.32)	(3.04)	(2.28)	(1.43)	(1.70)	(1.28)
Adj. R^2	0.4473	0.4302	0.4137	0.4142	0.4023	0.3975	0.3898	0.3803	0.3721	0.3710	0.3614	0.3594	0.3567	0.3429	0.3502
Obs.	16,728	15,046	13,457	12,023	10,703	9,525	8,447	7,446	6,512	5,642	4,856	4,101	3,421	2,781	2,161
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ind. & Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports robustness tests for the impact of *Potential* on future firm performance (ROA). The dependent variable is the firm-level return on assets (ROA) in year $t+k$, where k ranges from 6 to 20. **Panel A** uses an alternative measures of *Potential* as detailed in Section 6, each alternative measure is a composite index constructed by substituting specific components of the original *Potential*. Note that these coefficients are estimated from separate regressions, not a single specification. For presentation brevity, the results are stacked. **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** 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.

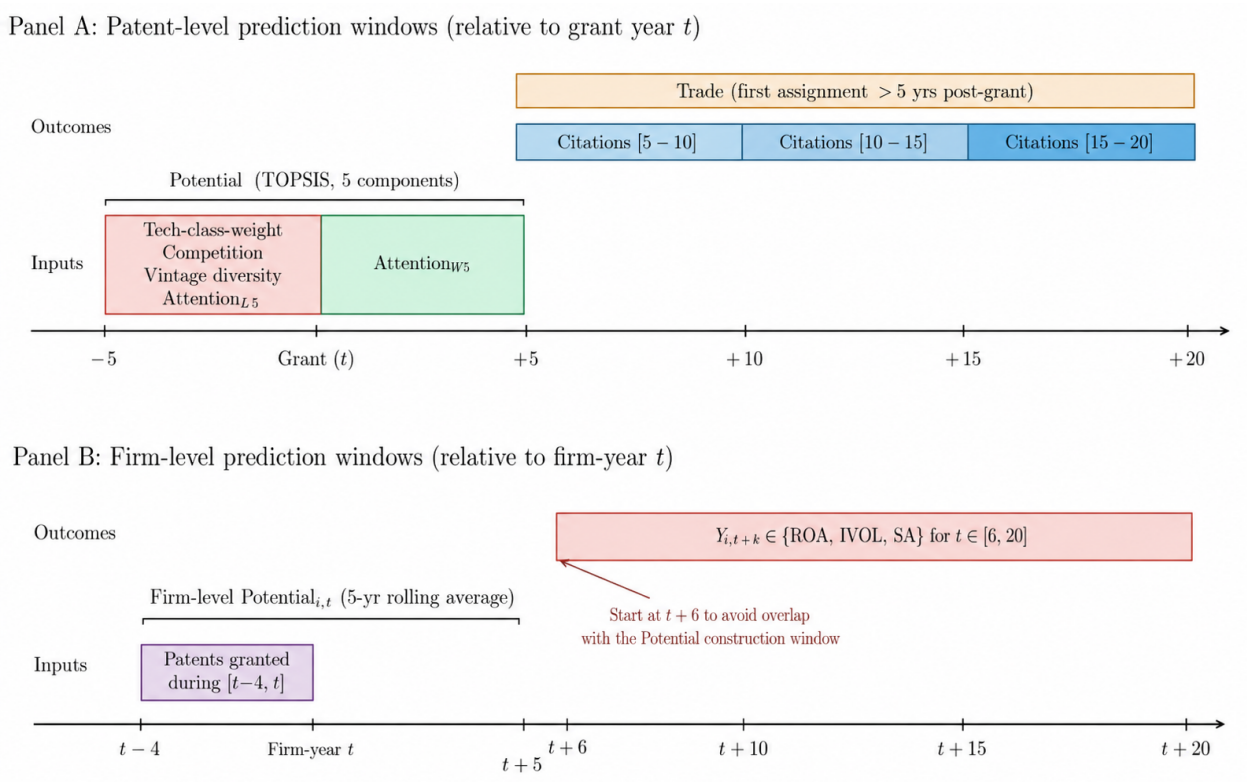


Figure 1: Prediction windows

Notes: Panel A shows the patent-level timing of *Potential* inputs and outcomes relative to grant year t. Panel B shows the firm-level timing relative to firm-year t: firm-level $Potential_{i,t}$ is constructed from patents granted during $[t-4, t]$ with a post-grant attention window extending to $t+5$, and outcomes are measured from $t+6$ to $t+20$.

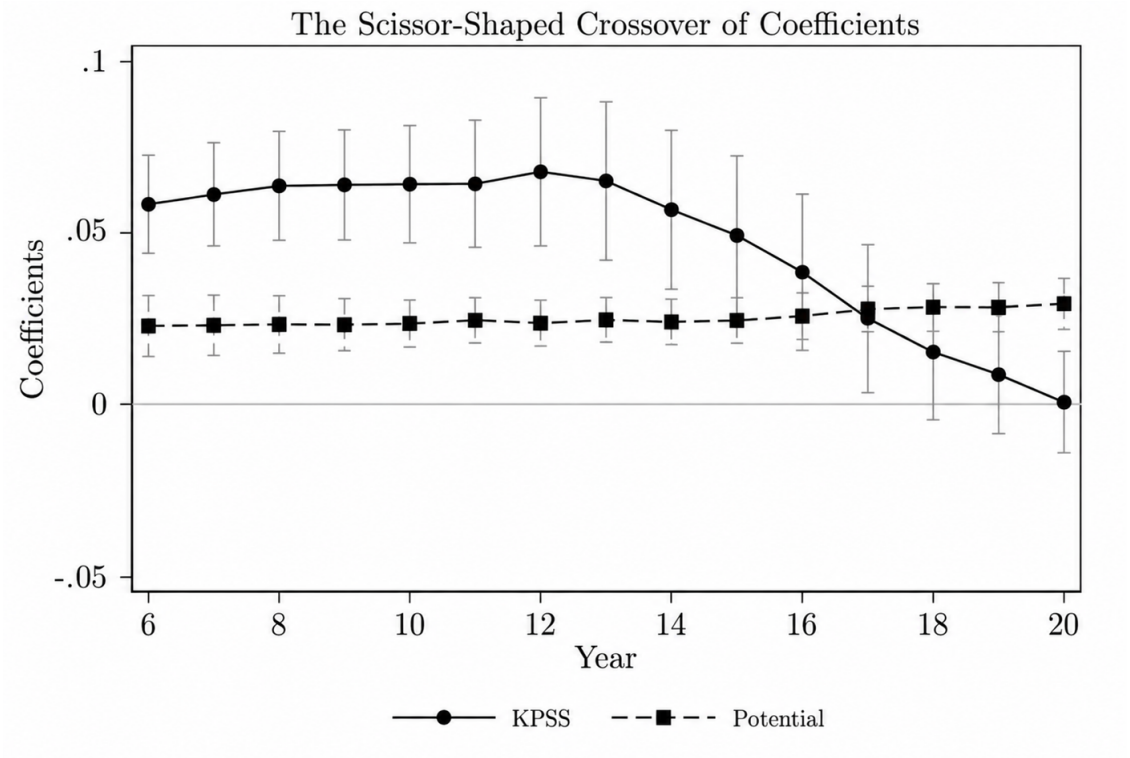


Figure 2: The Scissor-Shaped Crossover: Coefficient Patterns Over Horizons

Notes: This figure illustrates the estimated coefficients 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*; and (2) KPSS (Kogan et al., 2017). 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.

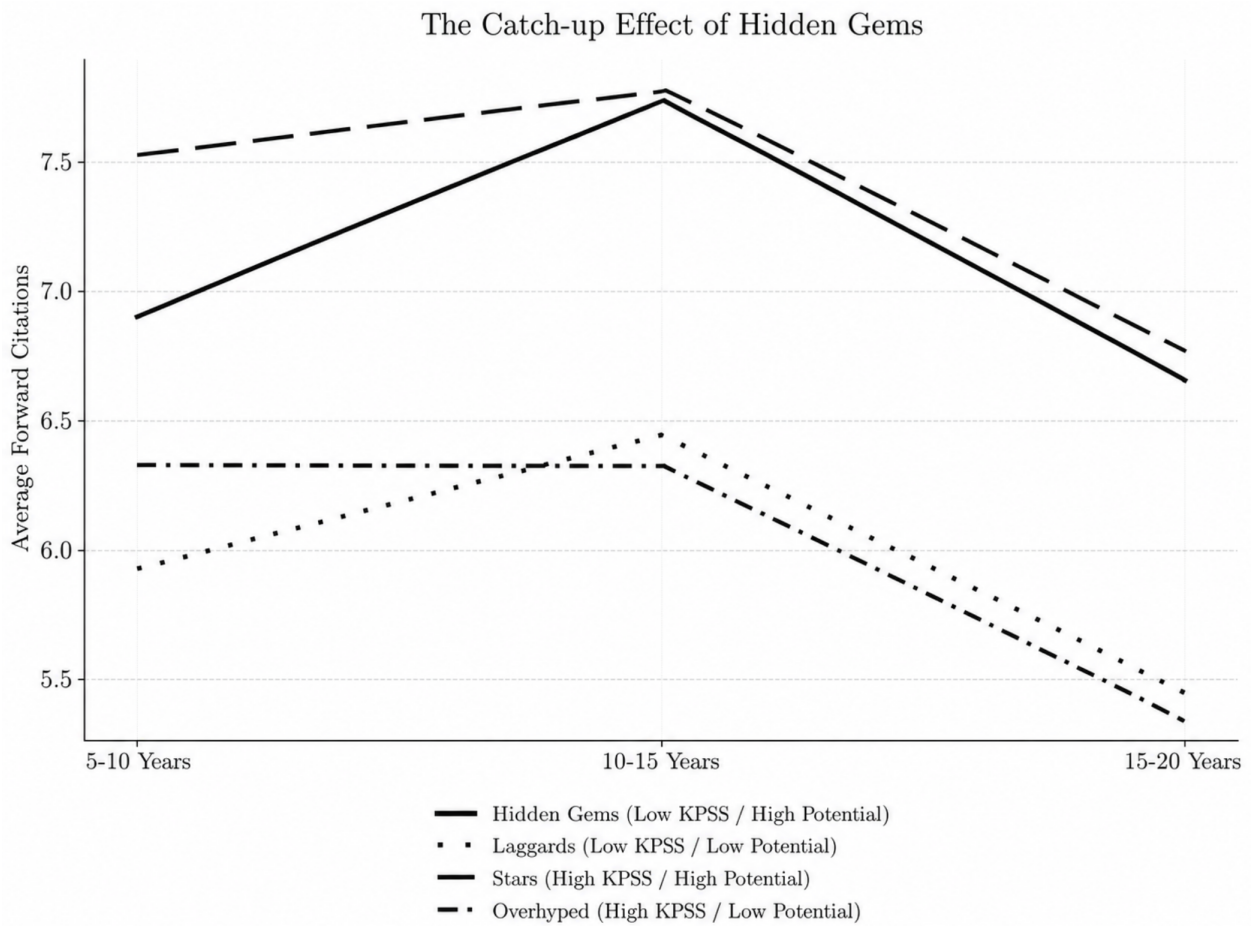


Figure 3: Long-term Citation Trajectory by Double-sorted Patent Groups

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 market value (KPSS). Within each KPSS group, patents are further sorted into High and Low groups based on the *Potential* index. This yields four mutually exclusive quadrants: HiddenGems (Low KPSS, High Potential), Stars (High KPSS, High Potential), Overhyped (High KPSS, Low Potential), and Laggards (Low KPSS, Low Potential). All sorting variables are residualized with respect to firm and grant-year fixed effects.

Appendix A Variable Definitions

Table A1: Variable Definitions

Variable	Definition
A. Core Patent-Level Variables	
Potential	The primary TOPSIS-based composite index of a patent's long-term potential. It is constructed from five components: *Tech-class-weight*, *Competition*, $Attention_{L5}$, $Attention_{W5}$, and *Vintage_diversity*. *Tech-class-weight*, *Competition*, and $Attention_{L5}$ are treated as cost-type attributes, while $Attention_{W5}$ and *Vintage_diversity* are treated as benefit-type attributes. A higher value indicates that the patent is closer to the profile of a high-potential late-blooming patent.
Attention_W5	(Component of *Potential*) 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 *Potential*) 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 *Potential*) 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 *Potential*) 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.
Vintage_diversity	A patent-level measure of inventor-team composition based on inventors' pre-filing patenting histories. It is calculated as the average pairwise cosine distance between yearly history vectors that record each inventor's prior patent applications by year.
Potential_main	Alternative TOPSIS-based measure of potential. It replaces *Tech-class-weight* with a primary-CPC-based variant, where the patent's main technology class is defined using the most frequently assigned 4-digit CPC category. All other components are the same as in the baseline *Potential* index.
Potential_comp	Alternative TOPSIS-based measure of potential. It replaces the stock-based *Competition* measure with *Competition_roll*, a flow-based measure defined as the logged number of patents granted in the same CPC subclass during the three-year rolling window prior to the focal patent's grant date. All other components are the same as in the baseline *Potential* index.

Variable	Definition
Potential_book	Alternative TOPSIS-based measure of potential. It replaces the Wall Street Journal-based pre- and post-grant attention measures, $Attention_{L5}$ and $Attention_{W5}$, with Google Ngram-based attention measures, $AttentionBook_{L5}$ and $AttentionBook_{W5}$. All other components are the same as in the baseline *Potential* index.
Potential_vintage	Alternative TOPSIS-based measure of potential. It replaces the cosine-distance-based *Vintage_diversity* measure with a Jaccard-distance-based variant constructed from inventors' prior active filing-year sets. All other components are the same as in the baseline *Potential* index.
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_main	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.
Competition_roll	The natural logarithm of one plus the number of patents granted in the focal patent's CPC subclass within a three-year rolling window preceding the grant year.
Vintage_diversity_alt	Jaccard-distance-based alternative to *Vintage_diversity*. It is constructed from inventors' pre-filing active-year sets and equals the average pairwise Jaccard distance across inventor pairs within the focal team.
B. Other Patent-Level Variables	
KPSS	Natural logarithm of one plus private value of a patent, calculated based on the KPSS method (Kogan et al., 2017), representing the market value of the patent.
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.

Variable	Definition
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.
KI-Index (KI5)	Text-based breakthrough index constructed over a forward-looking window of 5. Calculation is defined in Kelly et al. (2021). We use $\ln(1 + \text{KI})$ in regressions.
InnOrig	Innovation Originality. Natural logarithm of one plus the number technological fields associated with the patents cited by the focal patent. Following Hirshleifer et al. (2018).
Num_claims	Natural logarithm of one plus number of patent claims.
Num_citations	Natural logarithm of one plus number of other patents cited by the patent (backward citations).
Num_nonpat_citations	Natural logarithm of one plus number of non-patent citations, such as scientific literature, cited by the patent.
Num_inventors	Natural logarithm of one plus number of inventors listed on the patent.
Num_IPC	Natural logarithm of one plus number of International Patent Classification (IPC) codes assigned to the patent.
Grant_lag	Natural logarithm of one plus lag in days from the patent application date to the grant date.
C. Firm-Level Variables	
Potential (firm-level)	Average *Potential* index across all patents granted to a firm over the preceding five-year rolling window.
InnOrig (firm-level)	Innovation Originality. Average patent-level InnOrig across all patents granted to a firm over the preceding five-year rolling window. Following Hirshleifer et al. (2018).
CIE	Citations-based innovative efficiency measure. Natural logarithm of one plus 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. Natural logarithm of one plus number of 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. $BTM = \frac{\text{Book Equity}}{\text{Market Capitalization}}$.
ROE	Return on equity, defined as $\frac{IB+XINT}{BE_{t-1}}$, where IB is income before extraordinary items, $XINT$ is interest expense, and BE_{t-1} is lagged book equity.
ROA	Return on assets, defined as $\frac{IB+XINT}{AT_{t-1}}$, where AT_{t-1} is lagged total assets.
Assets	Natural logarithm of one plus total assets.

Variable	**Definition**
ADV	Advertising expenses scaled by total assets.
Capex	Capital expenditure scaled by total assets.
R&D	R&D expenditure scaled by total assets.
Salesgrowth	Calculated as the year-over-year growth rate of revenue.
PPE	Property, plant, and equipment scaled by total assets.
LEV	Debt to assets ratio. $LEV = \frac{\text{Total Debt}}{\text{Total Assets}}$.
IVOL	Idiosyncratic volatility measured as the standard deviation of the residuals from regressing daily excess stock returns on the Fama-French three-factor returns over the previous 12 months with a minimum of 31 trading days.
SA index	Financial constraints measured by the SA index of Hadlock and Pierce (2010), defined as $$SA = -0.737 \times Size + 0.043 \times Size^2 - 0.040 \times Age$$

Appendix B Event Responsiveness and Internal Consistency of the CPC Attention Measure

This appendix examines whether externally dated topic-level news events are reflected in the CPC attention series constructed to include the corresponding topics. For each 3-digit CPC class and calendar month, CPC attention aggregates the attention assigned to all news topics linked to that CPC through the topic–CPC mapping. The analysis therefore provides an event-responsiveness and internal-consistency test of the CPC attention measure.

The regression sample consists of 147 topic–event entries covering all 40 news topics. These entries correspond to 120 distinct historical shocks because the same shock may be relevant to more than one topic. Event months are defined by when a shock occurred or first became publicly salient, rather than by peaks in the CPC attention series. Table C2 reports the events used in the regressions and their mapped 3-digit CPC classes.

Event-Specific Standardization

For each event, we standardize CPC attention using a fixed 36-month pre-event baseline:

$$z_{c,m}^{(e)} = \frac{\text{Attention}_{c,m} - \overline{\text{Attention}}_{c,M_e-39:M_e-4}}{\sigma(\text{Attention}_{c,M_e-39:M_e-4})} \tag{C1}$$

In this expression, c indexes 3-digit CPC classes, e indexes topic–event entries, and m indexes calendar months. M_e denotes the calendar month in which event e occurs. $\text{Attention}_{c,m}$ is the raw attention assigned to CPC class c in month m. The term $\overline{\text{Attention}}_{c,M_e-39:M_e-4}$ is the arithmetic mean of CPC attention over the 36 months from $M_e - 39$ through $M_e - 4$, inclusive. The denominator $\sigma(\text{Attention}_{c,M_e-39:M_e-4})$ is the corresponding time-series standard deviation over the same 36-month period. Finally, $z_{c,m}^{(e)}$ is the event-specific standardized attention measure.

The superscript e emphasizes that the standardization is performed separately for each event. Thus, the same raw CPC-month observation can receive different standardized values in different event stacks because the relevant pre-event baseline depends on M_e. Using a fixed baseline ensures that attention observed during or after the event does not affect the mean or standard deviation used to scale the outcome.

Dynamic Event-Time Specification

Define event time as

$$s = m - M_e, \tag{C2}$$

where $s = 0$ denotes the event month, negative values denote months before the event, and positive values denote months after the event. We estimate the following stacked event-time specification over $s \in [-24, +24]$:

$$z_{c,m}^{(e)} = \alpha_{c,e} + \lambda_{e,m} + \sum_{s=-24}^{24} \beta_s \left[D_{c,e} \times \mathbf{1}\{m - M_e = s\} \right] + \varepsilon_{c,e,m}, \qquad \frac{1}{9} \sum_{s=-12}^{-4} \beta_s = 0 \tag{C3}$$

where $D_{c,e}$ is an indicator equal to one if CPC class c is mapped to the news topic associated with event e, and zero otherwise. The function $\mathbf{1}\{m - M_e = s\}$ equals one when calendar month m is exactly s months from event month M_e. Their interaction therefore identifies mapped CPC classes at each event-time month. The term $\alpha_{c,e}$ denotes CPC-by-event fixed effects. The term $\lambda_{e,m}$ denotes event-by-calendar-month fixed effects. The $\varepsilon_{c,e,m}$ is residual term. The coefficient β_s measures the mapped-minus-unmapped difference in standardized CPC attention at event time s, after accounting for both sets of fixed effects. Because all event-time interactions cannot be identified without a normalization, we impose

$$\frac{1}{9} \sum_{s=-12}^{-4} \beta_s = 0. \tag{C4}$$

Thus, the average mapped-minus-unmapped differential during months -12 through -4 is normalized to zero. Each β_s is interpreted relative to this pre-event comparison period. Months -3 through -1 are not included in the normalization and can therefore reveal anticipatory attention immediately before the event.

Each news topic receives equal total weight in the pooled estimation. Operationally, events belonging to topics with more selected events receive proportionally less weight, preventing topics with more events from contributing disproportionately to the estimated event-time pattern. Figure C1 reports the resulting event-time coefficients and their 95% wild-cluster bootstrap confidence intervals.

Event-Window Specification

We complement the dynamic estimates with regressions that summarize the response over economically meaningful event windows. For each event window $[a, b]$, we retain observations from the comparison period $s \in [-12, -4]$ and the indicated event window $s \in [a, b]$, and estimate

$$z_{c,m}^{(e)} = \alpha_{c,e} + \lambda_{e,m} + \beta^{[a,b]} \left[D_{c,e} \times \mathbf{1}\{m - M_e \in [a, b]\} \right] + \varepsilon_{c,e,m}. \tag{C5}$$

In this specification, a and b denote the first and last event-time months included in the event window. The indicator $\mathbf{1}\{m - M_e \in [a, b]\}$ equals one when month m falls within that window. All remaining symbols have the same definitions as in the dynamic specification.

The coefficient $\beta^{[a,b]}$ is a difference-in-differences estimate. It measures the change in the mapped-minus-unmapped attention differential from the pre-event comparison period $s \in [-12, -4]$ to the event window $s \in [a, b]$. Because the dependent variable is standardized separately for each CPC–event pair, the coefficient is expressed in event-specific pre-event

standard-deviation units. Standard errors are two-way clustered by CPC class and news topic. Statistical significance is evaluated using a topic-level wild-cluster bootstrap with 9,999 Rademacher draws, following **?**.

Results

Table C1 reports the event-window estimates. In the primary window $s \in [-1, +3]$, the mapped-minus-unmapped attention differential increases by 0.57 (t=4.23) event-specific pre-event standard deviations relative to the comparison period $s \in [-12, -4]$. The estimated response is also positive in each alternative window. The coefficient is 0.90 (t=4.38) for $s \in [0, +1]$, 0.46 (t=3.94) for $s \in [0, +6]$, and 0.34 (t=3.98) for $s \in [-3, +6]$. The results therefore do not depend on a single definition of the event window. These findings show that externally dated topic-level news events are accompanied by disproportionately larger increases in attention among the CPC classes linked to the affected topics. The evidence supports the event responsiveness and internal consistency of the CPC attention measure.

Table C1: Event Responsiveness of CPC Attention

	Event window			
	$[-1,+3]$	$[0,+1]$	$[0,+6]$	$[-3,+6]$
Mapped CPC × event window	0.573***	0.896***	0.458***	0.341***
	(4.23)	(4.38)	(3.94)	(3.98)
CPC × event fixed effects	Yes	Yes	Yes	Yes
Event × calendar-month fixed effects	Yes	Yes	Yes	Yes
Observations	133,770	105,105	152,880	181,545

Notes: The dependent variable is event-specific standardized CPC attention, $z^{(e)}_{c,m}$, constructed using the fixed 36-month baseline $s \in [-39,-4]$. Each column retains observations from the comparison period $s \in [-12,-4]$ and the event window indicated at the top of the column. Events are weighted so that each of the 40 news topics receives equal total weight. t-statistics based on standard errors two-way clustered by CPC class and topic are reported in parentheses. Significance stars are based on 9,999 topic-level wild-cluster bootstrap draws. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

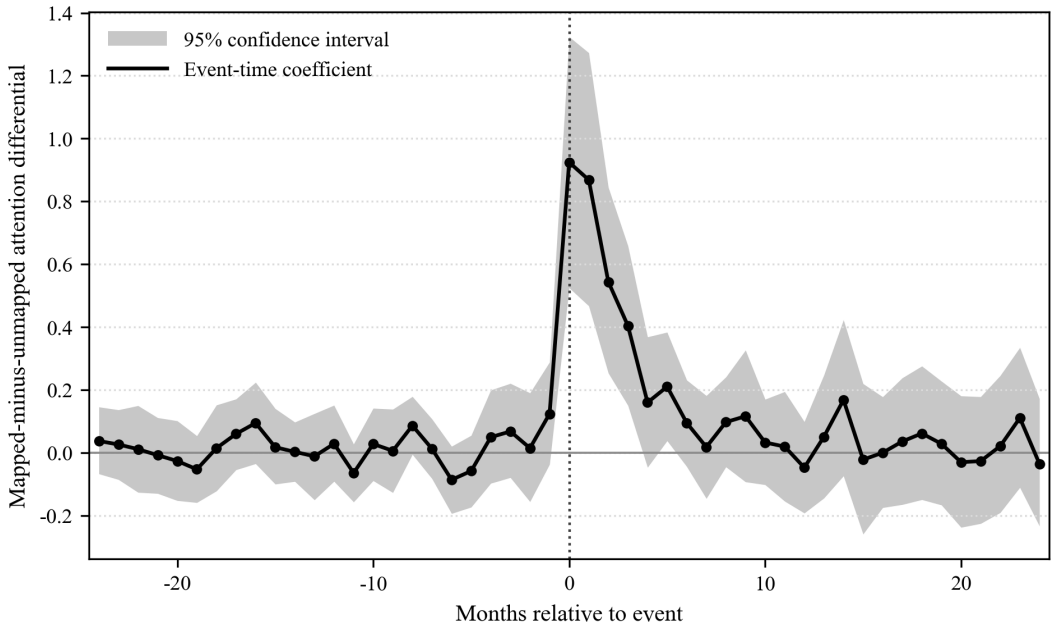


Figure C1: Event-Time Response of CPC Attention

Notes: The figure reports equal-topic-weighted estimates of the event-time coefficients β_s from the stacked CPC–event–month specification. The estimation uses 147 topic–event entries covering all 40 news topics. Event month zero corresponds to $m = M_e$. The shaded region reports 95% topic-level wild-cluster bootstrap confidence intervals. The event-time coefficients are normalized to average zero over $s \in [-12, -4]$. The specification includes CPC-by-event fixed effects, $\alpha_{c,e}$, and event-by-calendar-month fixed effects, $\lambda_{e,m}$.

Table C2: Historical Event Library for CPC Attention Validation

Topic	Event month	Event	Mapped CPC3
Natural disasters	1989-10	Loma Prieta earthquake	A62, E02, E04, G01, G08
Natural disasters	1992-08	Hurricane Andrew	A62, E02, E04, G01, G08
Natural disasters	2004-12	Indian Ocean earthquake and tsunami	A62, E02, E04, G01, G08
Natural disasters	2005-08	Hurricane Katrina	A62, E02, E04, G01, G08
Natural disasters	2011-03	Tohoku earthquake and tsunami	A62, E02, E04, G01, G08
Internet	1995-08	Netscape IPO	G06, H04
Internet	1998-09	Google founded	G06, H04
Internet	2000-03	Dot-com market peak	G06, H04
Internet	2004-02	Facebook launched	G06, H04
Internet	2007-06	iPhone launch	G06, H04
Soft drinks	1999-06	Coca-Cola Belgium contamination scare	A23, B01, B65, B67, C02
Soft drinks	2005-03	Coca-Cola Zero launch	A23, B01, B65, B67, C02
Mobile devices	1992-12	First SMS sent	G06, G09, H01, H02, H04
Mobile devices	2002-11	BlackBerry smartphone launch	G06, G09, H01, H02, H04
Mobile devices	2007-06	iPhone launch	G06, G09, H01, H02, H04
Mobile devices	2008-10	First Android phone launch	G06, G09, H01, H02, H04
Mobile devices	2010-04	iPad launch	G06, G09, H01, H02, H04
Police/crime	1992-04	Los Angeles riots	A61, E05, F41, G01, G06, G08
Police/crime	1995-04	Oklahoma City bombing	A61, E05, F41, G01, G06, G08
Police/crime	2014-08	Ferguson unrest	A61, E05, F41, G01, G06, G08
Electronics	1995-08	Windows 95 release	G06, H01, H02, H03, H05
Electronics	2004-11	Nintendo DS launch	G06, H01, H02, H03, H05
Electronics	2006-11	PlayStation 3 launch	G06, H01, H02, H03, H05
Electronics	2007-06	iPhone launch	G06, H01, H02, H03, H05
Steel	1997-07	Asian financial crisis begins	B21, C21, C22, C23
Steel	2002-03	US steel safeguard tariffs announced	B21, C21, C22, C23
Steel	2008-09	Global financial crisis	B21, C21, C22, C23
Cable	1996-02	Telecommunications Act of 1996	B65, D07, G02, H01, H02
Cable	1999-11	AT&T MediaOne merger announced	B65, D07, G02, H01, H02
Cable	2002-11	Comcast AT&T Broadband merger completed	B65, D07, G02, H01, H02
Cable	2015-02	FCC Open Internet Order	B65, D07, G02, H01, H02
Fast food	1993-01	Jack in the Box E coli outbreak	A21, A23, A47, B65, B67, G06
Fast food	2001-01	McDonalds India expansion milestone	A21, A23, A47, B65, B67, G06
Fast food	2006-12	Taco Bell E coli outbreak	A21, A23, A47, B65, B67, G06
Disease	2003-03	SARS global alert	A61, C12, G01, G16
Disease	2009-04	H1N1 influenza outbreak	A61, C12, G01, G16
Disease	2012-09	MERS coronavirus identified	A61, C12, G01, G16
Disease	2014-03	West Africa Ebola outbreak	A61, C12, G01, G16
Music industry	1999-06	Napster launched	G06, G10, G11, H04
Music industry	2001-10	iPod launch	G06, G10, G11, H04
Music industry	2003-04	iTunes Music Store launch	G06, G10, G11, H04
Music industry	2008-10	Spotify launched	G06, G10, G11, H04
Chemicals/paper	1989-03	Exxon Valdez oil spill	B31, C08, C09, D21
Chemicals/paper	2006-07	REACH regulation adopted	B31, C08, C09, D21
Chemicals/paper	2010-04	Deepwater Horizon oil spill	B31, C08, C09, D21
Environment	1989-03	Exxon Valdez oil spill	B01, B09, C02, E03, G01, Y02
Environment	1997-12	Kyoto Protocol adopted	B01, B09, C02, E03, G01, Y02
Environment	2005-02	Kyoto Protocol entered into force	B01, B09, C02, E03, G01, Y02
Environment	2010-04	Deepwater Horizon oil spill	B01, B09, C02, E03, G01, Y02
Mining	1989-03	Exxon Valdez oil spill	B03, E21, G01
Mining	1995-05	Bre-X gold discovery announced	B03, E21, G01
Mining	2006-01	Sago Mine disaster	B03, E21, G01
Mining	2010-08	Chile mine collapse	B03, E21, G01
Pharma	1998-03	Viagra approved	A61, C07, C12, G01
Pharma	2001-11	Anthrax antibiotics procurement	A61, C07, C12, G01
Pharma	2013-12	Sovaldi approved	A61, C07, C12, G01
Aerospace/defense	1991-01	Gulf War air campaign begins	B64, F02, F41, F42, G01
Aerospace/defense	2001-09	September 11 attacks	B64, F02, F41, F42, G01
Aerospace/defense	2003-03	Iraq War begins	B64, F02, F41, F42, G01
Phone companies	1996-02	Telecommunications Act of 1996	G06, H04
Phone companies	2000-07	Verizon formed	G06, H04
Phone companies	2007-06	iPhone launch	G06, H04
Computers	1995-08	Windows 95 release	G06, G11, H01, H03, H04, H05

Topic	Event month	Event	Mapped CPC3
Computers	2001-10	Windows XP release	G06, G11, H01, H03, H04, H05
Computers	2007-01	Windows Vista release	G06, G11, H01, H03, H04, H05
Nuclear/North Korea	1994-10	Agreed Framework signed	B01, F41, F42, G21
Nuclear/North Korea	2006-10	North Korea first nuclear test	B01, F41, F42, G21
Nuclear/North Korea	2011-03	Fukushima nuclear accident	B01, F41, F42, G21
Tobacco	1998-11	Tobacco Master Settlement Agreement	A24
Tobacco	2006-08	US tobacco racketeering ruling	A24
Tobacco	2009-06	Family Smoking Prevention and Tobacco Control Act	A24
Bio/Chem/Physics	1996-07	Dolly sheep birth announced	A61, C07, C08, C12, G01, G21, H01
Bio/Chem/Physics	2000-06	Human Genome Project draft announced	A61, C07, C08, C12, G01, G21, H01
Bio/Chem/Physics	2008-09	LHC first beam	A61, C07, C08, C12, G01, G21, H01
Bio/Chem/Physics	2012-07	Higgs boson discovery announced	A61, C07, C08, C12, G01, G21, H01
Movie industry	1995-11	Toy Story release	G02, G03, G06, G11, H04
Movie industry	1997-12	Titanic release	G02, G03, G06, G11, H04
Movie industry	2009-12	Avatar release	G02, G03, G06, G11, H04
Movie industry	2010-03	Netflix streaming international expansion begins	G02, G03, G06, G11, H04
Automotive	1997-12	Toyota Prius launched	B60, B62, F02, H01
Automotive	2008-09	Global auto crisis	B60, B62, F02, H01
Automotive	2010-01	Toyota unintended acceleration recall	B60, B62, F02, H01
Automotive	2012-06	Tesla Model S deliveries begin	B60, B62, F02, H01
Machinery	1995-01	WTO Agreement on Technical Barriers begins	B23, B25, F01, F04, F15, F16, G05
Machinery	2001-12	China joins WTO	B23, B25, F01, F04, F15, F16, G05
Machinery	2008-09	Global financial crisis	B23, B25, F01, F04, F15, F16, G05
Machinery	2011-03	Japan earthquake supply chain shock	B23, B25, F01, F04, F15, F16, G05
Space program	1998-11	International Space Station first module launched	A62, B64, F02, G01, G05, H02, H04
Space program	2003-02	Space Shuttle Columbia disaster	A62, B64, F02, G01, G05, H02, H04
Space program	2012-05	SpaceX Dragon ISS mission	A62, B64, F02, G01, G05, H02, H04
Oil drilling	1989-03	Exxon Valdez oil spill	B63, C09, E21, F16, G01
Oil drilling	2005-08	Hurricane Katrina Gulf energy disruption	B63, C09, E21, F16, G01
Oil drilling	2010-04	Deepwater Horizon oil spill	B63, C09, E21, F16, G01
Oil drilling	2014-11	OPEC maintains output amid oil price fall	B63, C09, E21, F16, G01
Rail/trucking/shipping	1994-12	NAFTA implementation early period	B60, B61, B62, B63, B65
Rail/trucking/shipping	2002-01	West Coast port slowdown begins	B60, B61, B62, B63, B65
Rail/trucking/shipping	2008-09	Global financial crisis	B60, B61, B62, B63, B65
Airlines	1988-12	Pan Am Flight 103 bombing	B64, G06, G07, G08
Airlines	2001-09	September 11 attacks	B64, G06, G07, G08
Airlines	2003-03	SARS aviation demand shock	B64, G06, G07, G08
Airlines	2010-04	Eyjafjallajokull air travel disruption	B64, G06, G07, G08
Software	1995-08	Windows 95 release	G06, H04
Software	1998-05	US v Microsoft antitrust case filed	G06, H04
Software	2001-10	Windows XP release	G06, H04
Software	2008-09	Google Chrome released	G06, H04
Software	2014-04	Heartbleed vulnerability disclosed	G06, H04
Agriculture	1996-03	Mad cow disease UK crisis	A01, C12
Agriculture	2001-02	UK foot-and-mouth outbreak	A01, C12
Agriculture	2008-06	Global food price crisis peaks	A01, C12
Agriculture	2012-07	US drought intensifies	A01, C12
Real estate	1990-08	Savings and loan crisis peak	E04, G01, G06, G08
Real estate	2007-08	Subprime mortgage crisis begins	E04, G01, G06, G08
Real estate	2008-09	Lehman Brothers bankruptcy	E04, G01, G06, G08
Currencies/metals	1997-07	Asian financial crisis begins	B42, C22, G01, G06, G07
Currencies/metals	1999-01	Euro introduced	B42, C22, G01, G06, G07
Currencies/metals	2011-08	Gold price surge during debt crisis	B42, C22, G01, G06, G07
Credit cards	1996-10	EMV standard first published	B42, G06, G07, H04
Credit cards	2006-12	PCI DSS version 1.1 released	B42, G06, G07, H04
Credit cards	2010-07	Dodd-Frank Act signed	B42, G06, G07, H04
Credit cards	2013-12	Target payment card breach disclosed	B42, G06, G07, H04
Terrorism	1993-02	World Trade Center bombing	A62, F41, G01, G06, G08
Terrorism	1995-04	Oklahoma City bombing	A62, F41, G01, G06, G08
Terrorism	2001-09	September 11 attacks	A62, F41, G01, G06, G08
Terrorism	2005-07	London bombings	A62, F41, G01, G06, G08
Oil market	1990-08	Iraq invades Kuwait	B63, B65, C10, E21, F17, G06

Topic	Event month	Event	Mapped CPC3
Oil market	1998-12	Oil price collapse period	B63, B65, C10, E21, F17, G06
Oil market	2008-07	Oil price peak	B63, B65, C10, E21, F17, G06
Oil market	2014-11	OPEC maintains output amid oil price fall	B63, B65, C10, E21, F17, G06
Couriers	1997-08	UPS strike	B60, B65, G06, G08
Couriers	2001-09	September 11 logistics disruption	B60, B65, G06, G08
Couriers	2002-01	West Coast port slowdown begins	B60, B65, G06, G08
Couriers	2013-12	UPS holiday delivery delays	B60, B65, G06, G08
Safety administrations	1989-03	Exxon Valdez oil spill	A23, A62, B60, F16, G01, G08
Safety administrations	1993-01	Jack in the Box E coli outbreak	A23, A62, B60, F16, G01, G08
Safety administrations	2001-09	September 11 attacks	A23, A62, B60, F16, G01, G08
Safety administrations	2010-01	Toyota unintended acceleration recall	A23, A62, B60, F16, G01, G08
Microchips	1993-03	Intel Pentium introduced	B82, G03, G06, G11, H01, H03, H05
Microchips	2000-11	PlayStation 2 US launch	B82, G03, G06, G11, H01, H03, H05
Microchips	2008-09	Global financial crisis	B82, G03, G06, G11, H01, H03, H05
Microchips	2011-03	Japan earthquake electronics supply shock	B82, G03, G06, G11, H01, H03, H05
Luxury/beverages	1999-01	LVMH Gucci takeover battle begins	A47, B42, B65, B67, C12, G06
Luxury/beverages	2001-09	September 11 luxury travel demand shock	A47, B42, B65, B67, C12, G06
Luxury/beverages	2008-09	Global financial crisis	A47, B42, B65, B67, C12, G06
Luxury/beverages	2012-11	China anti-corruption campaign begins	A47, B42, B65, B67, C12, G06
Retail	1995-07	Amazon launched	A47, B65, G06, G07
Retail	2005-09	Amazon Prime launched	A47, B65, G06, G07
Retail	2008-09	Global financial crisis	A47, B65, G06, G07

Notes: The table lists the 147 historical events used in the regressions and the 3-digit CPC classes (CPC3) implied by the topic–CPC mapping.

Appendix C
Robustness Tests

Table C1: Robustness: Long-term Patent Potential and Future Firm Performance (ROE)

ROE_{t+k}	t+6	t+7	t+8	t+9	t+10	t+11	t+12	t+13	t+14	t+15	t+16	t+17	t+18	t+19	t+20
Panel A: Alternative Measures of Potential															
Potential_main	0.1684***	0.2004***	0.2508***	0.2633***	0.3317***	0.3523***	0.2895***	0.2458***	0.2154***	0.1356*	0.1282*	0.1872**	0.2517***	0.2666***	0.3410***
	(3.12)	(3.75)	(4.44)	(4.58)	(5.40)	(5.39)	(4.81)	(3.86)	(3.19)	(1.89)	(1.69)	(2.46)	(3.15)	(3.26)	(4.00)
Adj. R^2	0.1828	0.1836	0.1651	0.1695	0.1566	0.1475	0.1591	0.1508	0.1579	0.1616	0.1476	0.1550	0.1536	0.1440	0.1402
Obs.	24,886	22,697	20,621	18,748	17,047	15,544	14,174	12,911	11,741	10,654	9,643	8,698	7,846	7,057	6,298
Potential_comp	0.2199***	0.2878***	0.3147***	0.2626***	0.3225***	0.3438***	0.2969***	0.2688***	0.2148***	0.1526**	0.1452*	0.1779**	0.2643***	0.2522***	0.3318***
	(4.16)	(5.51)	(5.72)	(4.68)	(5.39)	(5.40)	(5.07)	(4.34)	(3.27)	(2.19)	(1.96)	(2.40)	(3.39)	(3.16)	(3.97)
Adj. R^2	0.1830	0.1842	0.1656	0.1696	0.1566	0.1475	0.1592	0.1511	0.1580	0.1617	0.1477	0.1550	0.1538	0.1440	0.1402
Obs.	24,886	22,697	20,621	18,748	17,047	15,544	14,174	12,911	11,741	10,654	9,643	8,698	7,846	7,057	6,298
Potential_book	0.2777***	0.3002***	0.3195***	0.3054***	0.3638***	0.3932***	0.3228***	0.3097***	0.2717***	0.2489***	0.2373***	0.2534***	0.3215***	0.2946***	0.3360***
	(5.42)	(5.94)	(6.04)	(5.70)	(6.39)	(6.54)	(5.70)	(5.29)	(4.39)	(3.79)	(3.41)	(3.61)	(4.34)	(3.87)	(4.33)
Adj. R^2	0.1785	0.1794	0.1602	0.1656	0.1525	0.1447	0.1522	0.1471	0.1558	0.1576	0.1420	0.1448	0.1502	0.1460	0.1406
Obs.	26,513	24,372	22,316	20,379	18,567	16,926	15,432	14,067	12,812	11,653	10,578	9,560	8,643	7,802	6,994
Potential_vintage	0.2624***	0.3202***	0.3418***	0.3005***	0.3322***	0.3458***	0.2943***	0.2437***	0.2174***	0.1810**	0.1797**	0.2323***	0.3192***	0.3079***	0.3487***
	(4.63)	(5.72)	(5.80)	(5.02)	(5.21)	(5.10)	(4.72)	(3.69)	(3.10)	(2.43)	(2.26)	(2.91)	(3.81)	(3.58)	(3.88)
Adj. R^2	0.1832	0.1843	0.1657	0.1697	0.1565	0.1473	0.1590	0.1507	0.1579	0.1618	0.1478	0.1553	0.1541	0.1443	0.1401
Obs.	24,886	22,697	20,621	18,748	17,047	15,544	14,174	12,911	11,741	10,654	9,643	8,698	7,846	7,057	6,298
Panel B: Exclude Pharmaceutical Firms															
Potential	0.2341***	0.2594***	0.3112***	0.2908***	0.2981***	0.2934***	0.2556***	0.2498***	0.2213***	0.1523**	0.1289*	0.1732**	0.2207***	0.2002***	0.2617***
	(5.27)	(5.60)	(6.52)	(5.87)	(5.68)	(5.27)	(4.52)	(4.14)	(3.40)	(2.21)	(1.77)	(2.39)	(2.88)	(2.60)	(3.23)
Adj. R^2	0.1723	0.1596	0.1363	0.1237	0.1175	0.1155	0.1134	0.1068	0.1068	0.1109	0.1058	0.1170	0.1163	0.1139	0.1114
Obs.	22,712	20,748	18,875	17,184	15,648	14,294	13,063	11,917	10,851	9,860	8,941	8,082	7,296	6,573	5,878
Panel C: Subsample Analysis															
Year < 2000															
Potential	0.2469***	0.4262***	0.5187***	0.4516***	0.4283***	0.4076***	0.2879***	0.3136***	0.2815***	0.2506***	0.3194***	0.2989***	0.3286***	0.3614***	0.3903***
	(3.70)	(6.53)	(5.97)	(6.64)	(5.83)	(5.37)	(4.04)	(4.20)	(3.75)	(3.10)	(3.78)	(3.53)	(3.75)	(4.06)	(4.28)
Adj. R^2	0.1776	0.1739	0.1092	0.1559	0.1436	0.1366	0.1300	0.1198	0.1302	0.1296	0.1254	0.1359	0.1408	0.1279	0.1362
Obs.	9,305	8,683	8,093	7,542	7,060	6,665	6,326	6,011	5,712	5,446	5,170	4,910	4,691	4,489	4,297
Year ≥ 2000															
Potential	0.2554***	0.2183***	0.2036***	0.1818**	0.2786***	0.3200***	0.3041***	0.2243**	0.1759	0.0893	-0.0237	0.0549	0.1890	0.1163	0.2213
	(3.27)	(2.80)	(2.74)	(2.13)	(3.04)	(3.22)	(3.32)	(2.28)	(1.62)	(0.76)	(-0.18)	(0.41)	(1.29)	(0.74)	(1.24)
Adj. R^2	0.1916	0.1957	0.2074	0.1851	0.1716	0.1610	0.1857	0.1821	0.1873	0.1937	0.1727	0.1801	0.1762	0.1752	0.1636
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
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ind. & Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

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 measures of *Potential* as detailed in Section 6, each alternative measure is a composite index constructed by substituting specific components of the original *Potential*. It is important to note that these coefficients are estimated from separate regressions, not a single specification. For presentation brevity, the results are stacked; each sub-section within Panel A represents a distinct regression model where the specified alternative measure is used as the key independent variable. **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** 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 C2: Robustness: Long-term Patent Potential and Future Firm Idiosyncratic Volatility

$IVOL_{t+k}$	t+6	t+7	t+8	t+9	t+10	t+11	t+12	t+13	t+14	t+15	t+16	t+17	t+18	t+19	t+20
Panel A: Alternative Measures of Potential															
Potential_main	-0.6050***	-0.6883***	-0.6728***	-0.7466***	-0.7420***	-0.7241***	-0.6477***	-0.6327***	-0.6605***	-0.6736***	-0.5940***	-0.7105***	-0.5497***	-0.5546***	-0.5678***
	(-8.02)	(-8.79)	(-7.99)	(-8.28)	(-7.67)	(-7.07)	(-5.97)	(-5.57)	(-5.50)	(-5.32)	(-4.47)	(-5.16)	(-3.82)	(-3.61)	(-3.51)
Adj. R^2	0.4240	0.4167	0.4040	0.3959	0.3821	0.3790	0.3734	0.3735	0.3685	0.3536	0.3513	0.3527	0.3473	0.3364	0.3311
Obs.	27,137	25,564	23,447	21,376	19,508	17,809	16,316	14,944	13,644	12,443	11,311	10,273	9,291	8,410	7,562
Potential_comp	-0.7072***	-0.7758***	-0.7622***	-0.8280***	-0.8326***	-0.7847***	-0.7050***	-0.6899***	-0.7053***	-0.7105***	-0.5770***	-0.6848***	-0.5043***	-0.5218***	-0.6032***
	(-9.54)	(-10.09)	(-9.23)	(-9.38)	(-8.81)	(-7.84)	(-6.65)	(-6.22)	(-6.01)	(-5.75)	(-4.45)	(-5.08)	(-3.58)	(-3.47)	(-3.82)
Adj. R^2	0.4245	0.4173	0.4045	0.3964	0.3827	0.3794	0.3737	0.3738	0.3687	0.3538	0.3513	0.3526	0.3472	0.3363	0.3313
Obs.	27,137	25,564	23,447	21,376	19,508	17,809	16,316	14,944	13,644	12,443	11,311	10,273	9,291	8,410	7,562
Potential_book	-0.7392***	-0.8284***	-0.7947***	-0.8264***	-0.8452***	-0.8367***	-0.7996***	-0.7752***	-0.8278***	-0.8670***	-0.8356***	-0.9365***	-0.8049***	-0.9017***	-0.9588***
	(-10.03)	(-10.81)	(-9.77)	(-9.61)	(-9.21)	(-8.61)	(-7.75)	(-7.15)	(-7.25)	(-7.30)	(-6.72)	(-7.25)	(-5.93)	(-6.23)	(-6.31)
Adj. R^2	0.4220	0.4118	0.4004	0.3950	0.3800	0.3749	0.3683	0.3647	0.3592	0.3511	0.3487	0.3499	0.3389	0.3298	0.3231
Obs.	28,821	27,162	25,099	23,052	21,131	19,316	17,683	16,202	14,810	13,509	12,322	11,211	10,170	9,225	8,330
Potential_vintage	-0.7905***	-0.8318***	-0.7805***	-0.8067***	-0.8176***	-0.7487***	-0.6771***	-0.6866***	-0.6949***	-0.7487***	-0.6591***	-0.8386***	-0.6785***	-0.7829***	-0.9434***
	(-9.97)	(-10.11)	(-8.84)	(-8.55)	(-8.10)	(-7.01)	(-6.00)	(-5.81)	(-5.55)	(-5.67)	(-4.75)	(-5.82)	(-4.51)	(-4.87)	(-5.57)
Adj. R^2	0.4247	0.4173	0.4043	0.3960	0.3823	0.3790	0.3734	0.3736	0.3685	0.3538	0.3514	0.3531	0.3477	0.3372	0.3327
Obs.	27,137	25,564	23,447	21,376	19,508	17,809	16,316	14,944	13,644	12,443	11,311	10,273	9,291	8,410	7,562
Panel B: Panel Regression (Exclude Pharmaceutical Firms)															
Potential	-0.6057***	-0.6629***	-0.6581***	-0.6861***	-0.6473***	-0.5710***	-0.5095***	-0.4570***	-0.4861***	-0.4718***	-0.3538***	-0.4181***	-0.2006	-0.2412	-0.2640*
	(-8.05)	(-8.45)	(-7.86)	(-7.69)	(-6.78)	(-5.63)	(-4.78)	(-4.10)	(-4.11)	(-3.78)	(-2.70)	(-3.08)	(-1.41)	(-1.60)	(-1.67)
Adj. R^2	0.4087	0.3964	0.3821	0.3738	0.3578	0.3519	0.3481	0.3495	0.3351	0.3218	0.3166	0.3201	0.3091	0.2983	0.2979
Obs.	24,529	23,143	21,263	19,410	17,734	16,214	14,862	13,634	12,470	11,396	10,381	9,448	8,563	7,764	6,996
Panel C: Subsample Analysis by Time Period															
Year < 2000															
Potential	-0.5839***	-0.7402***	-0.7151***	-0.9911***	-0.9991***	-0.8838***	-0.6928***	-0.5348***	-0.4603***	-0.5470***	-0.4873***	-0.6944***	-0.4883***	-0.3791*	-0.6047***
	(-5.08)	(-6.13)	(-5.52)	(-7.15)	(-6.61)	(-5.56)	(-4.24)	(-3.19)	(-2.66)	(-3.06)	(-2.66)	(-3.74)	(-2.62)	(-1.95)	(-3.07)
Adj. R^2	0.4387	0.4346	0.4145	0.3926	0.3517	0.3496	0.3501	0.3579	0.3636	0.3536	0.3561	0.3515	0.3641	0.3548	0.3515
Obs.	9,985	9,339	8,719	8,147	7,644	7,183	6,808	6,471	6,155	5,883	5,613	5,357	5,119	4,910	4,712
Year ≥ 2000															
Potential	-0.9128***	-0.9398***	-0.9015***	-0.8042***	-0.7706***	-0.7580***	-0.7491***	-0.8269***	-0.9557***	-0.9013***	-0.7197***	-0.7195***	-0.5464**	-0.7163***	-0.5854**
	(-9.09)	(-9.08)	(-8.10)	(-6.74)	(-6.10)	(-5.64)	(-5.18)	(-5.40)	(-5.80)	(-5.10)	(-3.80)	(-3.57)	(-2.47)	(-2.94)	(-2.17)
Adj. R^2	0.4259	0.4169	0.4069	0.4040	0.4071	0.4044	0.3968	0.3951	0.3850	0.3675	0.3599	0.3654	0.3416	0.3278	0.3226
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
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ind. & Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports robustness tests for the impact of Potential on future firm's idiosyncratic volatility (IVOL). The dependent variable is the firm-level IVOL in year $t+k$, where k ranges from 6 to 20. **Panel A** uses an alternative measures of *Potential* as detailed in Section 6, each alternative measure is a composite index constructed by substituting specific components of the original *Potential*. It is important to note that these coefficients are estimated from separate regressions, not a single specification. For presentation brevity, the results are stacked; each sub-section within Panel A represents a distinct regression model where the specified alternative measure is used as the key independent variable. **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** 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 C3: Robustness: Long-term Patent Potential and Future Firm Financial Constraints

SA_{t+k}	t+6	t+7	t+8	t+9	t+10	t+11	t+12	t+13	t+14	t+15	t+16	t+17	t+18	t+19	t+20
Panel A: Alternative Measures of Potential															
Potential_main	-0.7341***	-0.7430***	-0.7829***	-0.8009***	-0.7860***	-0.7999***	-0.7796***	-0.7216***	-0.6643**	-0.5905**	-0.5302*	-0.4556	-0.4221	-0.3753	-0.2671
	(-5.03)	(-4.53)	(-4.05)	(-3.92)	(-3.75)	(-3.63)	(-3.38)	(-3.15)	(-2.68)	(-2.21)	(-1.95)	(-1.63)	(-1.57)	(-1.45)	(-1.15)
Adj. R^2	0.4447	0.4196	0.4016	0.3820	0.3634	0.3447	0.3292	0.3048	0.2873	0.2847	0.2790	0.2743	0.2728	0.2681	0.2691
Obs.	11,783	10,587	9,482	8,501	7,631	6,841	6,123	5,463	4,843	4,272	3,748	3,266	2,831	2,435	2,069
Potential_comp	-0.9956***	-1.0034***	-1.0489***	-1.0578***	-1.0505***	-1.0507***	-1.0357***	-0.9944***	-0.9657***	-0.8616***	-0.8219***	-0.7980***	-0.8304***	-0.8074***	-0.7007**
	(-6.18)	(-5.61)	(-5.02)	(-4.81)	(-4.54)	(-4.24)	(-4.09)	(-4.09)	(-3.93)	(-3.51)	(-3.35)	(-2.98)	(-2.98)	(-2.74)	(-2.37)
Adj. R^2	0.4492	0.4243	0.4067	0.3871	0.3688	0.3501	0.3348	0.3107	0.2937	0.2901	0.2844	0.2804	0.2801	0.2757	0.2755
Obs.	11,783	10,587	9,482	8,501	7,631	6,841	6,123	5,463	4,843	4,272	3,748	3,266	2,831	2,435	2,069
Potential_book	-0.9911***	-0.9792***	-0.9938***	-0.9737***	-0.9697***	-0.9645***	-0.9407***	-0.9113***	-0.8889***	-0.7948***	-0.7652***	-0.7647***	-0.7983***	-0.7458***	-0.6699***
	(-5.80)	(-5.31)	(-4.80)	(-4.48)	(-4.21)	(-3.88)	(-3.65)	(-3.71)	(-3.62)	(-3.29)	(-3.27)	(-3.15)	(-3.25)	(-2.92)	(-2.80)
Adj. R^2	0.4473	0.4220	0.4042	0.3855	0.3670	0.3481	0.3323	0.3088	0.2920	0.2884	0.2829	0.2782	0.2787	0.2761	0.2770
Obs.	11,890	10,781	9,735	8,775	7,885	7,076	6,334	5,648	5,003	4,413	3,874	3,381	2,939	2,535	2,164
Potential_vintage	-1.0650***	-1.0726***	-1.1132***	-1.0899***	-1.0438***	-1.0084***	-0.9726***	-0.9360***	-0.9106***	-0.8441***	-0.8230***	-0.8052**	-0.8997***	-0.8719**	-0.7896**
	(-5.46)	(-5.18)	(-4.73)	(-4.23)	(-3.78)	(-3.31)	(-3.12)	(-3.03)	(-2.95)	(-2.80)	(-2.77)	(-2.52)	(-2.81)	(-2.63)	(-2.20)
Adj. R^2	0.4493	0.4243	0.4066	0.3863	0.3673	0.3478	0.3321	0.3081	0.2912	0.2886	0.2834	0.2795	0.2804	0.2759	0.2764
Obs.	11,783	10,587	9,482	8,501	7,631	6,841	6,123	5,463	4,843	4,272	3,748	3,266	2,831	2,435	2,069
Panel B: Panel Regression (Exclude Pharmaceutical Firms)															
Potential	-0.8054***	-0.7800***	-0.8027***	-0.8143***	-0.7581***	-0.7326***	-0.7100***	-0.6932***	-0.6421***	-0.5537***	-0.4630***	-0.4179***	-0.4937***	-0.4578***	-0.4024**
	(-11.42)	(-10.33)	(-9.94)	(-9.46)	(-8.36)	(-7.66)	(-7.04)	(-6.46)	(-5.68)	(-4.70)	(-3.73)	(-3.16)	(-3.48)	(-3.04)	(-2.54)
Adj. R^2	0.4901	0.4643	0.4409	0.4151	0.3945	0.3746	0.3574	0.3325	0.3110	0.3052	0.2993	0.2961	0.2993	0.3004	0.3056
Obs.	10,340	9,294	8,316	7,448	6,676	5,980	5,354	4,784	4,246	3,756	3,306	2,882	2,505	2,157	1,834
Panel C: Subsample Analysis by Time Period															
Year $<$ *2000*															
Potential	-0.6869***	-0.7239***	-0.7806***	-0.8152***	-0.8722***	-0.9213***	-0.9174***	-0.8843***	-0.8062***	-0.6529***	-0.5282***	-0.5722***	-0.6132***	-0.6254***	-0.6927***
	(-4.83)	(-4.90)	(-5.14)	(-5.27)	(-5.55)	(-5.67)	(-5.40)	(-5.10)	(-4.49)	(-3.59)	(-2.80)	(-2.99)	(-3.14)	(-3.15)	(-3.42)
Adj. R^2	0.3829	0.3666	0.3519	0.3404	0.3417	0.3322	0.3242	0.3109	0.3123	0.3128	0.3196	0.3175	0.3264	0.3380	0.3527
Obs.	2,807	2,605	2,413	2,238	2,088	1,953	1,840	1,741	1,643	1,554	1,466	1,383	1,303	1,232	1,162
Year $\geq$ *2000*															
Potential	-0.9298***	-0.8994***	-0.9413***	-0.9525***	-0.8959***	-0.8500***	-0.8094***	-0.7791***	-0.7522***	-0.7051***	-0.6563***	-0.5966***	-0.6550***	-0.5692***	-0.3999*
	(-12.21)	(-10.96)	(-10.61)	(-9.98)	(-8.77)	(-7.80)	(-7.02)	(-6.27)	(-5.70)	(-5.05)	(-4.40)	(-3.64)	(-3.63)	(-2.86)	(-1.86)
Adj. R^2	0.4613	0.4394	0.4251	0.4022	0.3798	0.3635	0.3481	0.3256	0.3043	0.3048	0.2935	0.2853	0.2827	0.2774	0.2715
Obs.	9,303	8,281	7,329	6,483	5,728	5,051	4,431	3,859	3,325	2,832	2,386	1,968	1,606	1,269	960
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ind. & Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports robustness tests for the impact of Potential on future firm's financial constraints (SA index). The dependent variable is the firm-level SA index in year $t+k$, where k ranges from 6 to 20. **Panel A** uses an alternative measures of *Potential* as detailed in Section 6, each alternative measure is a composite index constructed by substituting specific components of the original *Potential*. It is important to note that these coefficients are estimated from separate regressions, not a single specification. For presentation brevity, the results are stacked; each sub-section within Panel A represents a distinct regression model where the specified alternative measure is used as the key independent variable. **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** 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.

Appendix D Mapping of CPC Classifications and News Topics

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 E 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 Hennessy
Retail	Wal-Mart, department store, Home Depot, Costco, Sears, Kmart, Macy's, retail industry, Nordstrom, discount store