RIETI Discussion Paper Series 25-E-097

Patent Information Disclosure and Market Reactions: Empirical investigation by using text-based novelty and impact indicators

DANG, Jianwei
Tongji University

MOTOHASHI, Kazuyuki
RIETI

ZHAO, Quihan
University of Tokyo

Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 25-E-095
October 2025

Patent Information Disclosure and Market Reactions: Empirical investigation by using text-based novelty and impact indicators[1]

Jianwei Dang (Tongji University)
Kazuyuki Motohashi (RIETI, University of Tokyo)
Quihan Zhao (University of Tokyo)

Abstract

We examine how financial markets value patented innovation across disclosure stages in the pharmaceutical sector. Using U.S. patents from publicly listed firms, we construct text-based measures of technological novelty and early diffusion impact. Event-study regressions based on KPSS returns reveal that market reactions are already sizable at the stage of scientific publication for patent–paper pairs (PPPs) and are nearly as large at pre-grant publication as at the patent granting stage—challenging the grant-centric view of patent valuation. Stage-specific regressions show a marked shift in valuation logic: novelty is priced early, especially when peer-reviewed science certifies the invention, while impact becomes salient only once technical content and downstream reuse become observable at the patent publication or granting stage. In PPPs, novelty is strongly rewarded in early stages, but impact is not, suggesting that science-linked inventions follow a distinct valuation channel. These patterns are robust to stricter PPP-matching thresholds and alternative impact metrics. Our findings highlight that both when and what gets disclosed jointly shape the financial value of innovation.

Keywords: Patent information disclosure, Market reactions, Patent novelty, Patent impact
JEL Classification: O32, O34, G14

[1] This study is conducted as a part of the project "Research on Digital Innovation Models" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). The draft of this paper was presented at the RIETI DP seminar for the paper. I would like to thank participants of the RIETI DP Seminar for their helpful comments. The authors also acknowledge financial support from MEXT/JSPS KAKENHI (Grant Number: 23K25540)

1. Introduction

Understanding how markets respond to the disclosure of patent information is a central question at the intersection of innovation studies and financial economics. Patents are not only legal instruments for securing intellectual property rights but also strategic signals to investors, competitors, and partners about a firm's technological capabilities and future market potential. Timely interpretation of patent-related information can influence investment decisions, firm valuation, and competitive positioning (Martens, 2023). Consequently, financial markets often react to patent disclosures, with these reactions serving as an indirect but powerful measure of how innovation is valued in real time.

Previous research has primarily examined market reactions at the time of patent grant, often relying on the KPSS model to quantify abnormal returns attributable to the release of patent-related information (Kogan et al., 2017). This approach implicitly assumes that the grant date is the pivotal moment at which value-relevant information is revealed to the market. However, the patenting process involves multiple stages of information dissemination, including application filing, pre-grant publication, examination, and eventual grant. In reality, these stages can vary substantially in the richness and market relevance of the information disclosed. Treating the grant date as the sole focal point risks overlooking earlier events that may shape market expectations.

The case of pharmaceutical patents is especially informative for studying these dynamics. First, pharmaceutical innovation is characterized by high R&D intensity, long development timelines, and significant regulatory hurdles. This means that market participants closely track each signal that could update expectations about the eventual commercial success of a drug. Second, patent rights in this sector often underpin the entire exclusivity period for blockbuster drugs, making them a primary determinant of firm valuation (Grabowski, 2002). Third, pharmaceutical technologies typically represent discrete innovations—a single drug is often protected by one or only a small number of key patents—unlike in sectors such as ICT, where products are the outcome of dozens or even hundreds of overlapping patents (Cohen et al., 2000). This discrete nature makes it easier for investors to attribute value directly to specific patents and to react strongly to their disclosure. Fourth, pre-grant publications in pharmaceuticals can reveal not only the chemical or biological structure of the compound but also its therapeutic target, claim scope, and early-stage efficacy data—information that is highly salient to investors, analysts, and competitors. As a result, markets may incorporate much of the expected value of a pharmaceutical patent well before the legal certainty of a grant is achieved.

In this paper, we extend the analytical scope beyond grant-day events to systematically examine market reactions to before-grant publication—when patent applications are first made publicly accessible through patent office databases. Using a comprehensive dataset of pharmaceutical patents, we document that KPSS values on the day of pre-grant publication are comparable in magnitude to those observed on grant dates. This challenges the prevailing assumption that market attention is concentrated solely at the grant stage and suggests that focusing only on grant-day KPSS may lead to systematic underestimation of patent value.

Beyond documenting this timing effect, we investigate the underlying drivers of these market responses. Drawing on recent advances in text-based patent analytics, we construct two distinct indicators: novelty, measured by the introduction of previously unseen technological terms, and impact, measured by the degree to which these terms are subsequently reused in later patents. Our analysis reveals that it is impact—rather than novelty per se—that most strongly predicts KPSS values, implying that investors prioritize the expected downstream adoption and commercialization of patented inventions over their originality alone. This distinction is particularly salient in the pharmaceutical sector, where the ultimate success of an invention depends not just on its scientific distinctiveness but also on clinical trial outcomes, regulatory approval, manufacturing feasibility, and market uptake.

To further disentangle the mechanisms linking information disclosure to market reactions, we examine a subset of Patent–Paper Pairs (PPPs)—patents associated with a peer-reviewed scientific publication. This subset allows us to observe how complementary channels of knowledge dissemination interact in shaping investor perceptions. We find that PPPs yield higher KPSS values than non-PPPs and that the publication of the corresponding scientific paper itself elicits a significant positive abnormal return. Interestingly, for PPPs, novelty appears to play a more prominent role at both the paper's publication and the patent's pre-grant publication stages, whereas impact shows no significant effect—suggesting that the value mechanism for science-based patents differs fundamentally from that of other patents.

By integrating the timing of disclosure, the content of disclosed information, and the interplay between scientific and patenting outputs, this study makes three contributions. First, it broadens the empirical focus of KPSS-based event studies to include pre-grant disclosures, revealing an earlier onset of market incorporation of patent information in pharmaceutical industry. Second, it introduces text-based novelty and impact measures into the analysis of market reactions, offering a finer-grained understanding of which technological attributes matter to investors. Third, it highlights the distinct valuation dynamics of science-linked patents in pharmaceuticals, underscoring the importance of considering the knowledge and industry context in which an invention emerges.

2. Literature Review and Theories

2.1 Market Reactions to Patent Disclosures

Patenting serves as a signaling mechanism to investors, where the informational value of patents plays a crucial role in revealing firms' innovative activities and, consequently, influencing firm value. Two main streams of literature have emerged in this domain. The first, more traditional line of research examines the relationship between firms' market value or investment decisions and various indicators of patent quality or value—such as forward citations and oppositions (Hall et al., 2005; Harhoff et al., 2003). The second stream, pioneered by Kogan et al. (2017), measures patent value using the abnormal stock market returns on the day of patent issuance. This approach has inspired a substantial body of follow-up research (Kelly et al., 2021).

In their original model, Kogan et al. (2017) focus on the patent grant date, arguing that this marks the resolution of uncertainty surrounding the application. They report no significant price reaction around the filing date and only weak reactions around the pre-grant publication date. However, subsequent studies—particularly in pharmaceutical and biotechnology contexts—suggest that investors closely monitor patent disclosures (Beyhaghi et al., 2023; Mohammadi & Khashabi, 2021), and that grant decisions in these industries can often be predicted with reasonable accuracy (Zhu et al., 2022). This makes the pre-grant publication date potentially highly informative for the market. Evidence from biotech and pharma indicates that investors track pipeline developments well before the grant, hinting at the pre-grant stage's importance (Ahn et al., 2015).

Despite these insights, there remains a clear literature gap: few studies have conducted systematic, cross-event comparisons of market reactions—measured by KPSS magnitudes—between the grant date and the pre-grant publication date. More importantly, there is value in examining which patent attributes shape market responses at each stage and in conducting a comparative analysis to uncover potential differences in these effects.

2.2 Patent Content as a Driver of Market Response

The economic value of patents is highly skewed, with a small proportion accounting for the majority of total returns (Harhoff et al., 1999). This skewness creates substantial uncertainty for investors and market participants, who must assess not only whether a patent has potential value, but also the likely magnitude of that value. Reliable evaluation of patent quality is therefore essential for guiding investment decisions, informing licensing strategies, and shaping corporate innovation portfolios. Over the years, researchers have developed a range of indicators to approximate patent quality, each capturing different facets of technological significance, commercial potential, and strategic importance.

Traditional approaches to measuring patent quality have primarily relied on citation-based indicators, with forward citations widely used as proxies for a patent's technological importance (Harhoff et al., 1999; Jaffe & Trajtenberg, 2002). A higher number of forward citations is generally interpreted as reflecting greater influence on subsequent technological developments, and, by extension, higher perceived value. In addition to citations, other structural characteristics of patents, such as claim breadth and family size, have been employed as indicators of the potential market reach and scope of protection (Lanjouw & Schankerman, 2004; Lerner, 1994). Broad claims may signal a wide range of applications, while large family size suggests strategic efforts to secure protection in multiple jurisdictions, both of which can be associated with higher commercial significance (Putnam, 1996).

In recent years, the increasing availability of digitized patent documents and advances in computational linguistics have facilitated the emergence of text-based analytics as an alternative and complementary approach to assessing patent quality. One notable approach is to measure the novelty of patent by text mining, for instance, detect the appearance of new technical terms or rare semantic combinations in patent texts, thereby capturing the originality of disclosed inventions (Jeon et al., 2022; Arts et al., 2021). Another approach is to get new indicator to measure the impact of invention, as replacements of the traditionally used citation indicator (Packalen & Bhattacharya, 2015; Zhao et al., 2025).

From a theoretical perspective, novelty reflects technological differentiation—how distinct an invention is from existing knowledge—while impact captures both commercial and technological diffusion—how far and widely the disclosed knowledge spreads and is utilized. While these dimensions are conceptually distinct, their relative importance in shaping external stakeholders' evaluations, particularly those of investors, remains underexplored. Specifically, there is limited empirical evidence on which dimension—novelty or impact—investors weight more heavily, and whether such weighting differs across disclosure stages, such as pre-grant publication and patent grant.

2.3 Science–Technology Linkages and Patent Valuation

The intersection between scientific research and technological invention has long been recognized as a critical driver of innovation outcomes (Mansfield, 1991; Narin et al., 1997). One prominent way to operationalize these linkages is through Patent–Paper Pairs (PPPs)—instances where a patent is directly associated with a peer-reviewed scientific publication. PPPs have been shown to represent inventions that are both more novel and more influential than patents without such linkages (Fleming & Sorenson, 2004; Ahmadpoor & Jones, 2017). However, they also tend to carry greater uncertainty and longer time-to-market, as they often emerge from early-stage scientific research rather than incremental or market-driven development (Ardito & Svensson, 2024).

Scientific publications associated with PPPs can serve as complementary disclosure channels to patents (Agrawal & Henderson, 2002). While patents disclose technical and legal claims, peer-reviewed articles communicate underlying scientific principles, methods, and results to a broader research audience. The peer-review process itself acts as a credibility signal, assuring investors and other stakeholders of the originality and rigor of the disclosed knowledge, often before the patent reaches the grant stage. This early-stage signaling may influence how the market evaluates the prospective value of science-based inventions, as firms that publicly disclose research through scientific publications tend to enjoy higher market valuations (Simeth & Cincera, 2016).

In the context of valuation differences for science-based inventions, PPPs may attract distinctive patterns of investor attention. The reputational weight of scientific peer review can increase the salience of novelty—since the scientific community typically prioritizes originality as a core evaluative criterion. Consequently, for PPP-related patents, novelty measures may be more strongly associated with market reactions than for non-PPP patents, particularly in early disclosure stages such as pre-grant publication (Kelly et al., 2021). In contrast, impact measures—which capture the reuse and diffusion of disclosed terms—may be less relevant before commercialization pathways are clear (Hall et al., 2005).

Despite the conceptual plausibility of these differences, there is a notable gap in the empirical literature. Few studies have examined how patent and paper disclosures jointly influence market valuation within the same innovation, particularly by comparing KPSS magnitudes across disclosure events. Integrating patent-based text metrics (novelty and impact) with the presence or absence of a scientific publication could yield a more nuanced understanding of how science–technology linkages shape investor perceptions at different stages of the innovation disclosure process.

3. Methodology

3.1. Data Sources and Initial Sample Construction

We construct a comprehensive dataset by integrating multiple sources of patent, firm, and financial information. The core data sources include:

1) Patent data is retrieved from the United States Patent and Trademark Office (USPTO), focusing on utility patents granted between December 2000 and December 2023.[1] The restriction to post-AIPA (American Inventors Protection Act) patents ensures the availability of key lifecycle dates, including application date (*adate*), pre-grant publication date (*pdate*), and grant date (*gdate*), and enables the identification of publication-stage market response.

2) Firm-level identifiers for publicly listed U.S. pharmaceutical companies are obtained from the Orbis-Moody's database, filtered by four-digit Nomenclature of Economic Activities (NACE) industry codes 2110 and 2120.[2] We link patents to firms via standardized fuzzy matching of patent assignee names with company names in Orbis-Moody's. The matching process accounts for name variants and subsidiaries to ensure that each patent is associated with the correct parent firm. By focusing on publicly traded firms, we can observe stock market reactions to patent-related events.

3) Financial market data is sourced from the CRSP (Center for Research in Security Prices) database. This allows us to compute cumulative abnormal returns (CARs) around key patent event dates as a measure of changes in firm market value attributable to patent disclosures (Kogan et al., 2017). The CRSP data provide a high-frequency view of market valuation and are merged with patent data via the entity's name and stock ticker (PERMCO and PERMNO), after matching the patent assignee to the name. All stock price and market value data are adjusted for stock splits and dividends. We express monetary values (such as changes in market

---

[1] Patent Data: https://patentsview.org/download/data-download-tables.

[2] NECE Rev. 2: https://www.openriskmanual.org/wiki/NACE_Classification.

capitalization) in constant 1982 U.S. dollars by deflating with the Consumer Price Index (CPI), ensuring that all monetary figures are CPI-adjusted to 1982 USD for consistency over time.

4) Scientific linkage data is derived from the Patent-Paper Pairs (PPP) database, which links USPTO patents with peer-reviewed scientific publications (Marx and Fuegi, 2022). A patent-paper pair is defined as a patent whose underlying invention was also disclosed through a scientific publication (typically a journal article by the inventors with very similar title or abstract). This yields a subset of patents (henceforth *PPP* patents) that have an external publication, providing an early public disclosure proxy for the invention's details. Only patents with an identifiable scientific publication are considered PPPs, consistent with prior large-scale studies linking millions of patents to publications (Poege et al., 2019; Kown, 2024). For each *PPP* patent, we record the publication date of the corresponding scientific article as an approximate public disclosure date of the invention (denoted *pub_adate*), since the U.S. patent application itself is not publicly revealed until 18 months after filing. This proxy publication date will serve as an application-stage event date for *PPP* patents in our analysis.

After assembling these sources, we construct a panel of patents linked to firm-day observations around critical event dates. These initial sample (a total of 38,902 granted patents linked to 283 pharmaceutical firms) consists of USPTO-granted patents held by publicly traded pharmaceutical firms (as identified via Orbis-Moody's), along with each patent's key dates (*adate*, *pdate*, *gdate*) and linkages to scientific publications (*PPP*, *pub_adate*) if any. This integrated dataset enables us to trace how information about a patented innovation enters the public domain (through scientific publication, pre-grant publication, and grant) and to analyze investors' reactions at each stage.

3.2. Variable Construction and Descriptive Statistics

3.2.1. Market Value and Financial Variables

The primary outcome variable (ξ) in our analysis is the change in firm market value in response to patent-related disclosure. We measure this via CAR around each patent event date (Kogan et al., 2017), which is calculated from the CRSP daily stock returns using a standard market model. For each event (e.g., a patent's grant date), we compute the firm's stock return over a short window (such as the day of the event and the following day, or a [-1, +1] day window) and subtract the expected return based on an overall market index. The resulting CAR represents the excess return attributable to the information released by that event. We multiply

CAR by the firm's market capitalization to translate it into dollar change in market value for interpretability.

Specifically, for each focal patent, we compute the idiosyncratic return of the associated firm's stock over a 3-day window [-1, +1] surrounding each of the following disclosure events:

(1) Scientific publication date (*pub_adate*) – PPP patents only

(2) Pre-grant publication date (*pdate*) – all patents

(3) Grant date (*gdate*) – all patents

Abnormal returns are obtained by regressing daily firm-level returns on market returns to isolate firm-specific innovations. These are then translated into monetary value changes using the firm's market capitalization at the time. All monetary values are deflated using the CPI to constant 1982 USD and in logarithm form.

3.2.2. Scientific Disclosure and Science-Technology Linkage

To examine whether the market reacts differently to patents that are closely connected to frontier science, we incorporate information from the PPP database,[3] which links USPTO granted patents to scientific publications they are directly (or high semantically) based on. A PPP match indicates that the core idea underlying the invention has been previously documented in the scientific literature (Kwon et al., 2024). This allows us to distinguish science-based patents from other inventions (*PPP* = 1 or 0).

Ahmadpoor & Jones (2017) suggests that science-based patents differ systematically from others in terms of novelty, applicability, and market value. In cases where a single patent corresponds to multiple scientific articles (which is not uncommon for major inventions with long development histories), we use the earliest associated publication date as the benchmark for information disclosure (*pub_adate*). This choice reflects the idea that the initial scientific disclosure, even if followed by multiple extensions or related papers, often marks the first point at which the core technological idea becomes visible to the scientific and commercial communities.[4] Early scientific disclosure may serve as a leading signal to investors,

---

[3] Patent-Paper Pairs: https://relianceonscience.org/patent-paper-pairs.

[4] This convention (see e.g., Marx & Fuegi, 2020) ensures that (i) the scientific disclosure chronologically precedes the patented invention, preserving causal interpretation; (ii) each patent enters the sample once,

competitors, and collaborators, thereby affecting market expectations about the patent's future value before formal patent filings are made public.

3.2.3. Patent Novelty and Window Impact Indicators

We construct novelty and impact measures for each patent to quantify the originality of its technological content and the influence of that content on subsequent inventions. In line with recent research using natural language processing on patent text (Arts et al., 2021), we analyze the text of patent documents (titles, abstracts, and claims) to identify novel technical terms introduced by each patent. [5]

Specifically, a patent's novelty is measured by the introduction of new keywords, ngrams and combinations of words (regardless the sequence) that have not appeared in any prior patent in the USPTO corpus (pre-grant & grant) up to that time (application date of the focal patent):

$$Novelty_j = \log\left(1 + \sum_{j \in new} \mathcal{N}_j^{(1)} + \mathcal{N}_j^{(2)} + \mathcal{N}_j^{(3)} + \mathcal{N}_j^{(4)}\right),$$

where $\mathcal{N}_j^{()}$ refer to the number of new words, 2-gram, 3-gram and combinations for each focal patent j respectively. Importantly, $Novelty_j$ is fixed once the document was filed into USPTO and remains unchanged across lifecycle stages.

These new technical words signal previously untapped ideas or concepts. We then track the diffusion of these novel terms in future patents to gauge impact. The impact indicator is defined as the frequency with which other subsequent patents adopt the focal patent's novel terms within a certain time frame after the focal patent's disclosure. For example, one metric counts the number of times the patent's new words are reused by other patents within one year of the pre-grant publication or grant. A higher count means the patented innovation's language is quickly propagating, indicating that other inventors find the idea useful and are building on it.

To trace how the impact of an invention evolves across the patent lifecycle, we focus on three distinct stages (*pub_adate*, *pdate*, *gdate*). For each stage, we define a symmetric 180-day

---

preventing overweighting of inventions that cite many papers; and (iii) the analysis focuses on the initial scientific trigger rather than later, potentially follow-up publications.

[5] We adopt preprocessing steps based on a USPTO-specific stopword list developed by Arts et al. (2021), available at https://github.com/sam-arts/respol_patents_code.

window starting from the corresponding date to capture the short-term response of the innovation ecosystem. This setting offers several advantages:

(1) It reflects a realistic timeframe during which market participants (competitors, collaborators, investors) can process disclosed information and act upon it (e.g., by initiating related filings or technological follow-ups).

(2) It avoids contamination across stages by ensuring that adate + 180 days ⩽ pdate, and pdate + 180 days ⩽ gdate for most patents.

3) For the *pub_adate* stage, since reuse data is not available, we proxy scientific influence of this stage using reuse counts from the 180 days following the *adate*, assuming that *pub_adate* typically follows *adate* closely in *PPP* patents (we restrict this order as well). [6]

We define the window impact as:

$$Impact_j = \log\left(1 + \sum_{i \in \mathcal{N}_j} R_i\right),$$

where R_i represents how many subsequent patents in 180-day time window after the pre-grant publication or grant date of the focal patent reuse the new term i. $\mathcal{N}_j$ contains all the new

[6] Because many Paper-Patent Pairs have no patent on file at the moment the article is published, we cannot observe standard impact metrics (e.g., forward-citation counts) exactly at the paper publication date. To avoid conflating "a technology's intrinsic worth" with "the moment markets discover that worth," we approximate ex-ante impact accrued in 180 days after the patent's application date:

(1) Earliest observable engineering signal: The application is the first formal disclosure of the invention's engineered embodiment to the patent office and technical diligence community. Impact metrics (e.g., citations) received shortly thereafter are widely accepted as a technology-pull indicator (Jaffe & Trajtenberg, 2002; Hall, Jaffe & Trajtenberg, 2005).

(2) Causal ordering preserved: Although measured post-application, this impact reflects how attractive the underlying technology is to subsequent inventors, not how financial markets react at paper publication, patent publication, or grant. Thus, they serve as a conservative stand-in for the invention's latent quality at the article's release.

(3) Bias direction is conservative: If a breakthrough's influence materializes only after patent application, our proxy will understate—never overstate—its true ex-ante attractiveness, making our stock-market tests stricter.

(4) Comparability with non-PPP patents: Patents unlinked to papers can only be evaluated from their own filing date forward. Applying the same early-impact window across PPP and non-PPP groups preserves methodological symmetry for cross-sectional comparison.

expressions of the focal patent j. As a result, we restrict the initial sample by each duration equal or more than 180-day window (e.g., pdate + 180 days ≤ gdate).

We further restrict the sample so that the 180-day window around a given disclosure ends before the next disclosure date; this eliminates any overlap between adjacent windows and prevents "peeking" at information that belongs to a later event.[7] The content cluster and statistics of the filtered samples is shown in Appendix Figure 2 and Table 1.

3.2.4. Control Variables and Fixed Effects

To ensure that the estimated effects of patent novelty and technological impact are not confounded by systematic differences in patent characteristics, we control for the length of the patent text, measured by the total number of words across the title, abstract, and claims. Since our novelty and impact indicators are constructed from textual analysis (e.g., the emergence and reuse of new words or word combinations), longer texts are mechanically more likely to contain novel concepts and to be reused by others.

Second, we control for technology heterogeneity using fixed effects at the level of the CPC subclass. Patents in different technology domains may receive systematically different levels of market attention (e.g., breakthrough drugs vs. incremental packaging innovations), and novelty may be more or less expected in some domains than others. The use of CPC subclass fixed effects ensures that our coefficients are identified from within-field comparisons, preventing bias from cross-domain variation. The distribution of patents across CPC subclasses is summarized in Appendix Figure 3.

Finally, we include firm fixed effects and corresponding year fixed effects to absorb unobserved heterogeneity across firms and over time. Firm fixed effects account for differences in firm size, visibility, and R&D strategy that may jointly affect patent characteristics and market response. Year fixed effects absorb time-varying macroeconomic conditions and market-wide fluctuations.

Together, these controls ensure that the estimated effects of patent novelty and post-filing technological influence are identified based on within-firm, within-field, and within-period variation, thereby supporting a more credible interpretation of our coefficients as reflecting

[7] We provide the duration distribution of the filtered sample in Appendix Figure 1.

economic relevance rather than spurious correlation. Appendix TAble 1 concludes variables and their explanations.

3.3. Empirical Analysis

3.3.1. Baseline Model

We estimate a series of baseline regression models to examine how patent novelty and impact relate to firm market value at three key stages of the patent lifecycle: (1) the scientific paper disclosure stage (*pub_adate*), (2) the pre-grant publication stage (*pdate*), and (3) the grant stage (*gdate*). Each stage corresponds to an event in which information about the patent becomes available to researchers or investors, and we expect the market to update the firm's value based on the perceived importance of the invention. We use an event-study framework, regressing the firm's abnormal stock return around the event on measures of the patent's novelty and impact (and control variables). This approach is consistent with prior research that uses stock market reactions to infer the private value of patent innovations (Kogan et al., 2017; Zheng, 2025). By analyzing each stage separately, we allow the effect of novelty to potentially differ depending on the timing and context of the disclosure. Below we outline the baseline model for each stage:

$$\xi_{j,\ stage\_year}^{stage} =$$

$$\alpha + \beta_1 * Novelty_j + \beta_2 * Impact_{stage,\ j} + \gamma X_j$$

$$+ \delta_{firm} + \delta_{stage\_year} + \delta_{cpc\_sub} + \varepsilon_j\ ,$$

where $stage \in \{pub\_adate,\ pdate,\ gdate\}$ representing three different stages of information disclosure in patent life circle. We control the content length and make the firm, corresponding year and technology field as the fixed effect.

For patents that are part of a Patent-Paper Pair, the associated scientific publication provides a public signal of the invention around the time of application. Thus, for PPP patents we define the event as the publication date of the scientific article (*pub_adate*). This serves as a proxy for the disclosure of the invention's novelty to the market at the earliest stage. We estimate a baseline model for the PPP subset at this stage, regressing the ξ around the publication of the scientific article on the patent's novelty and impact indicators. This baseline paper publication-stage model tests whether a highly novel invention (as signaled by the scientific publication content) yields a positive immediate revaluation of the firm.

For the first stage, we expect that, if markets pay attention to scientific disclosures, patents whose papers introduce new scientific ideas (high novelty) will generate a positive ξ. This stage is particularly important for science-linked inventions, as it reflects the market's initial evaluation of novel ideas before any formal patent disclosure. Because only PPP patents have a *pub_adate* event, this regression is necessarily limited to that subset. We emphasize that this model captures the ex-ante commercial potential of an invention as soon as it becomes publicly known in any form.

The second stage corresponds to the pre-grant publication date (*pdate*) in the patent system. For U.S. patents, this typically occurs 18 months after the application filing, when the USPTO publishes the application (unless the applicant opted for non-publication, which is uncommon in this sample). At this *pdate* event, the detailed contents of the patent (specifications, claims, etc.) become publicly available to all, which may reveal technical information beyond what was in a scientific paper. The key regressors are the patent's novelty and impact measures, which are now fully observable to the market (even for non-PPP patents that had no prior scientific disclosure).

For the second stage, this model captures the market's reaction to the first patent-system disclosure of the invention. We anticipate a patent that quickly gains influence (high impact within 180 days) may be viewed as more commercial promising. However, it is also possible that much information is already anticipated by investors (especially for PPP patents whose science was known), so the incremental surprise at *pdate* can vary.

The third and typically most salient stage is the patent grant date, when the patent is officially issued by the USPTO and the property right is secured. The grant event removes any residual uncertainty about the patent's validity and scope. This model assesses whether, controlling for the information already revealed earlier, the patent's novelty and impact still explain additional market value upon grant.

We still expect β_2 to be positive if investors reward the company for securing rights to an influential innovation. It is at the grant stage that the patent's commercial exclusivity is confirmed, so even if an invention's technical details were known earlier, the grant can be a value-creating event by cementing the firm's competitive advantage (Kogan et al., 2017).

3.3.2. Interaction Model

While the baseline models establish average effects of patent novelty and impact, our analysis also explores how these effects might differ conditional on the patent's scientific context. In particular, we estimate an interaction model that introduce moderation by the patent's PPP status. These interaction models allow us to test whether the stock market values a patent's novelty differently when the invention is scientifically enriched or academically disclosed. Prior literature suggests that science-linked inventions can carry extra signal value for markets (Poege et al., 2019; Masclans et al., 2024). Therefore, we augment our regression framework with interaction terms at the pre-grant publication and grant stages (*pdate* and *gdate*) to capture these nuances. Both interaction models are specified for the pre-grant publication and grant events (we do not use the interaction model at the first stage, since by definition PPP status would be uniform in the very limited PPP-only sample). Specifically, we estimate it as follows:

$$\xi_{j,stage\_year}^{stage} =$$

$$\alpha + \beta_1 * Novelty_j + \beta_2 * Impact_{stage,\,j}$$

$$+ \beta_3 * PPP + \beta_4 * PPP * Novelty_j + \beta_5 * PPP * Impact_{stage,\,j}$$

$$+ \gamma X_j + \delta_{firm} + \delta_{stage\_year} + \delta_{cpc\_sub} + \varepsilon_j,$$

here, $stage \in \{pdate,\ gdate\}$.

β_4 tests whether being a PPP patent amplifies or dampens the effect of novelty on the market value. There are two plausible, non-mutually-exclusive hypotheses: (a) PPP patents might receive stronger market rewards for novelty because an academic publication can validate the invention's significance (investors see the peer-reviewed disclosure as affirmation of the invention's importance) (Masclans et al., 2024). Alternatively, (b) PPP patents might experience weaker incremental reactions at later patent stages because much of the information was already revealed to the public through the paper (the novelty was partially "priced in" at the time of the scientific publication). By including the interaction, we allow the data to inform which effect dominates. We also include β_3 (main effect of PPP) to capture baseline differences between PPP and non-PPP patents. Notably, $PPP_i = 1$ patents may generally be more fundamental and could have higher average value (Poege et al., 2019), so β_3 is expected to be positive. In practice, which mechanism dominates—and how the interaction terms manifest—remains an empirical question that we aim to test in the following sections.

4. Empirical Results

4.1. Summary Statistics

Table 1 presents descriptive statistics for the key variables used in our analysis, grouped into three distinct stages: scientific publication (paper-publication), pre-grant publication, and grant. Specifically, these measures reflect the monetary value derived from stock-market reactions, measured in log-transformed 1982 U.S. dollars.

Table 1. Descriptive Statistics by Disclosure Stage

Vars	Obs	Mean	Std. dev.	Min	Max
Scientific Publication Stage					
ξ_{pub_adate}	2,082	3.798	3.047	0.0000225	9.958
Novelty	2,082	3.229	1.843	0	9.927
Impact	2,082	0.646	0.984	0	6.324
Length	2,082	5.567	0.825	2.708	8.907
Pre-grant publication Stage					
ξ_{pdate}	22,733	0.299	0.419	-0.00283	3.322
Novelty	22,733	3.558	1.556	0.693	10.216
Impact	22,733	0.962	1.247	0	6.762
Length	22,733	5.708	0.827	2.197	9.565
Grant Stage					
ξ_{gdate}	23,322	0.329	0.466	-0.00175	4.013
Novelty	23,322	3.339	1.878	0	11.735
Impact	23,322	1.265	1.289	0	6.767
Length	23,322	5.660	0.911	1.386	9.989

The average market value is highest at the scientific publication stage (mean = 3.798), followed by much lower values at the pre-grant publication (0.299) and grant (0.329) stages. This pattern reflects a strong early reaction to foundational scientific information, but it also partly stems from sample composition: patents matched to scientific publications (PPP group) tend to be of higher intrinsic value than the general population of patents (Fleming & Sorenson, 2001; Azoulay et al., 2019). In other words, patents grounded in peer-reviewed science are not only disclosed earlier but also more likely to be high-impact innovations, which is reflected in higher observed market value even before formal patent recognition.

This pattern is consistent with prior studies suggesting that scientific disclosures often precede patent filings by several months or even years, allowing the market to respond to signals embedded in the scientific literature before any patent document becomes available (Finardi, 2011; Zhang et al., 2017). Although our analysis does not explicitly quantify these science-to-patent lags, the observed peak at the science stage supports the interpretation that early visibility of rigorous scientific content enhances investor attention and anticipation of valuable innovation (Ke, 2020).

Novelty scores are moderate across all stages, averaging 3.23 for scientific-stage patents and around 3.55 and 3.34 for pre-grant publication and grant stages, respectively. Impact—measured by how often a patent's novel content is reused within 180 days—increases over time: from 0.65 at the scientific publication stage to 1.26 by the grant stage, consistent with cumulative knowledge diffusion patterns (Griliches, 1990; Lanjouw & Schankerman, 2004).

Taken together, this pattern reinforces those early scientific disclosures are treated as meaningful innovation signals by investors, even before patent content (e.g., claims) are published. Meanwhile, low early-stage impact values reflect limited opportunity for reuse before pre-grant publication.

4.2. Baseline Regression Results

Table 2 (a-c) report the baseline regression results for distinctive stages, all regressions control for firm, year, and technology (CPC subclass) fixed effects, and exhibit high explanatory power with adjusted R-squared values ranging from 0.70 to 0.88.

Effect of novelty. We observe that at the scientific publication stage (Table 2 (a)), the coefficient on novelty is positive and statistically significant (0.0926, $p<0.05$), indicating that markets place higher value on inventions emerging from novel scientific research when first disclosed. This stage likely represents the point of greatest information asymmetry, so novelty serves as a credible early signal of breakthrough potential. This finding aligns with prior literature indicating that markets assign premium valuations to groundbreaking scientific discoveries due to their potential for future disruptive innovations (Fleming & Sorenson, 2001). In contrast, at the pre-grant and grant stages (Table 2 (b) & 2 (c)), the novelty coefficients become negligible and statistically insignificant, suggesting that once patent details are formally disclosed, the incremental informational value derived solely from novelty diminishes, potentially due to prior absorption of the novelty-related information at earlier disclosure stages or reduced market uncertainty.

Effect of Impact. In contrast to novelty, technological impact—measured as the frequency at which patent-specific new content is reused within 180 days after disclosure—a short-window proxy for early diffusion—shows a different valuation pattern. At the scientific publication stage (Table 2 (a)), the impact coefficient is negative (–0.0511) but statistically insignificant, suggesting that early-stage scientific disclosures alone do not guarantee immediate market appreciation of technological impact. However, at the pre-grant publication and grant stages (Table 2 (b) & 2 (c)), the impact coefficients become positive and significant (0.00346, $p<0.05$, at pre-grant publication; 0.00279, $p<0.05$, at grant), indicating that patents with higher short-term technological impact achieve significantly higher market valuations. Economically, these coefficients imply that patents whose contents are promptly reused by subsequent inventions within the 180-day window are perceived by investors as more valuable, reflecting their foundational importance and higher expected economic return. This is consistent with the literature emphasizing that market valuations of patents often increase as technological significance and citation-based influence accumulate over time (Trajtenberg, 1990; Griliches, 1990).

Overall, these baseline results provide robust evidence of distinct mechanisms shaping market valuations across the innovation disclosure lifecycle. Early-stage valuations primarily reflect initial investor relies on signals of scientific novelty when information asymmetry is high (Hegde & Sampat, 2015), whereas later-stage valuations increasingly depend on tangible evidence of technological impact and diffusion capabilities. This differentiation supports existing innovation theories emphasizing distinct valuation criteria (Kogan et al., 2017) at different stages of technological development.

Together, these results support our theoretical framework: novelty matters most when uncertainty is high and signaling via scientific channels is salient; impact becomes increasingly important as the market gains access to concrete technical content and can observe evidence of diffusion. The shift in valuation drivers across stages reflects an investor learning process shaped by disclosure timing and content.

Table 2 (a). Baseline Regressions at Scientific Publication Stage

	ξ_{pub_adate}	ξ_{pub_adate}	ξ_{pub_adate}
Novelty	0.0791**		0.0926**
	(0.0364)		(0.0438)
Impact		-0.0731	-0.0511
		(0.0536)	(0.0403)
Length	-0.0756***	-0.0760***	-0.0725***
	(0.0210)	(0.0202)	(0.0218)

_cons	4.000***	3.741***	3.972***
	(0.198)	(0.295)	(0.211)
FEs			
Firm	Y	Y	Y
Year	Y	Y	Y
CPC_subclass	Y	Y	Y
Obs	2,082	2,082	2,082
Adj R^2	0.866	0.875	0.887

Table 2 (b). Baseline Regressions at Pre-grant Publication Stage

	ξ_{pdate}	ξ_{pdate}	ξ_{pdate}
Novelty	0.000953		0.000702
	(0.000903)		(0.00116)
Impact		0.00365***	0.00346**
		(0.00114)	(0.00138)
Length	-0.00108	-0.000947	-0.00157
	(0.0216)	(0.0261)	(0.0215)
_cons	0.292***	0.290***	0.281***
	(0.0128)	(0.0136)	(0.0109)
FEs			
Firm	Y	Y	Y
Year	Y	Y	Y
CPC_subclass	Y	Y	Y
Obs	22,733	22,733	22,733
Adj R^2	0.680	0.678	0.703

Table 2 (c). Baseline Regressions at Grant Stage

	ξ_{gdate}	ξ_{gdate}	ξ_{gdate}
Novelty	0.000496		0.000480
	(0.000875)		(0.000915)
Impact		0.00304**	0.00279**
		(0.00157)	(0.00120)
Length	-0.00177	-0.00122	-0.00178
	(0.0179)	(0.00164)	(0.0187)
_cons	0.338***	0.336***	0.337***
	(0.00943)	(0.00989)	(0.00932)
FEs			
Firm	Y	Y	Y
Year	Y	Y	Y
CPC_subclass	Y	Y	Y
Obs	23,322	23,322	23,322
Adj R^2	0.754	0.761	0.763

Note: *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The numbers in parentheses represent robust standard errors clustered at the firm level.

4.3. Interaction Regression Results

Next, we examine how the linkage between scientific publications and patents—captured by whether an invention originates from a PPP—affects the market valuation of patents at different disclosure stages. Columns 2 and 4 of Table 3 report regressions at the pre-grant publication and grant stages, including PPP and its interactions with novelty and impact. The PPP group includes inventions explicitly matched to scientific publications (2,082 pairs). A positive PPP coefficient would imply that, all else equal, patents closely linked to scientific research receive systematically higher market valuations than patents without direct scientific ties.

We expect that the certification effect of peer-reviewed science may differentially influence how investors interpret novelty versus impact. If science serves primarily as an early signal of idea quality, we should observe enhanced returns to novelty in PPPs, but not necessarily stronger returns to impact—since the latter depends more on market-facing diffusion.

Table 3. Interaction Effects of Scientific Linkage (PPP)

	ξ_{pdate}	ξ_{pdate}	ξ_{gdate}	ξ_{gdate}
Novelty	0.000424	0.000419	0.000478	0.000482
	(0.000928)	(0.00103)	(0.000893)	(0.00101)
Impact	0.00468**	0.00494**	0.00413***	0.00396**
	(0.00213)	(0.00226)	(0.00158)	(0.00169)
PPP	0.00610**	0.00510*	0.00527	0.00504
	(0.00239)	(0.00290)	(0.00328)	(0.00355)
PPP × Novelty		0.00398*		0.00178
		(0.00214)		(0.00252)
PPP × Impact		0.00239		0.00318
		(0.00592)		(0.00636)
Length	-0.00101	-0.00105	-0.00178	-0.00179
	(0.00221)	(0.00222)	(0.00178)	(0.00183)
_cons	0.283***	0.289***	0.338***	0.3337***
	(0.0124)	(0.0126)	(0.00986)	(0.00979)
FEs				
Firm	Y	Y	Y	Y
Year	Y	Y	Y	Y
CPC_subclass	Y	Y	Y	Y
Obs	22,733	22,733	22,733	23,322
Adj R^2	0.694	0.697	0.761	0.763

Note: *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The numbers in parentheses represent robust standard errors clustered at the firm level.

The results indicate that the PPP indicator is positively associated with patent market value at the pre-grant publication stage (0.00510, $p < 0.10$), although at the patent grant stage, the coefficient remains positive but not statistically significant (0.00504). Economically, this suggests that inventions explicitly grounded in scientific research receive modestly higher market valuations when their patent details are first disclosed, controlling for novelty, impact, and fixed effects. This finding aligns with prior studies that emphasize how scientific foundations signal higher quality or more fundamental innovations to investors (Narin et al., 1997; Masclans et al., 2024).[8]

Turning to the interaction terms, the effect of novelty on market valuation is significantly amplified when a patent is linked to scientific research. At the pre-grant publication stage, novelty is already positively and significantly associated with market returns, suggesting that investors respond to novel technical disclosures upon public availability. However, the interaction coefficient for *PPP × Novelty* is also positive and significant (0.00398, $p < 0.10$), indicating an additional market premium for novelty when it is grounded in scientific research.

This result suggests that although novelty itself is already recognized by the market at this stage, investors attribute greater credibility and commercial potential to novel inventions when these are explicitly tied to prior academic publications. In other words, scientific grounding appears to reinforce investors' confidence in the feasibility and significance of novel inventions. This aligns with recent findings that scientific linkages enhance perceived patent value by reducing information asymmetry and providing clearer signals of technical quality and originality (e.g., Azoulay et al., 2019; Kwon et al., 2024).

By contrast, the *PPP × Impact* interaction terms are positive but not statistically significant at both the pre-grant publication and grant stages. This suggests that the valuation of technological impact—measured by early reuse of novel content—is relatively independent of scientific linkage. Investors appear to value impact consistently based on observable diffusion potential, rather than the scientific background of the patent.

In summary, while technological impact commands a uniform market premium across science-based and non-science-based patents, the valuation of novelty is not uniform: patents rooted in science receive a significantly higher novelty premium. These results extend earlier qualitative insights (e.g., Murray, 2002) by quantitatively confirming the economic advantage of science-based novelty.

[8] See Appendix Table A2 for PPP-only regressions at three stages. The coefficients mirror the baseline pattern—novelty fades while impact remains economically meaningful—indicating that results are not driven by mixing PPP and non-PPP observations.

Appendix Figure 4 illustrates the distribution of the PPP matching score used to identify paper–patent pairs. In our analysis, we use a broad threshold (PPP score ≥ 1); robustness checks with stricter thresholds will be discussed later.

4.4. Robustness Checks

4.4.1. Controlling for Time Lags

One concern is that the timing of various stages could themselves influence the adoption lag. To address this, we add controls for the duration between stages in the regressions. Specifically, for the PPP subset we control for the gap (in log years) between the paper's publication and the pre-grant publication; for the full patent sample, we control for the time between pre-grant publication and grant (the patent's prosecution duration). The idea is to account for any effect of prolonged development or examination times on eventual adoption.

The results (Table 4) show that adding duration controls does not alter our main coefficients qualitatively. The gap between a paper's publication and its associated patent's pre-grant publication enters with a negative and weakly significant coefficient (–0.00787, $p<0.10$ in col. 1). This indicates that, holding novelty, impact, and fixed effects constant, longer science-to-patent intervals are associated with smaller event-day market valuations at pre-grant publication. A natural interpretation is an anticipation/price-discovery channel: when more time elapses before formal patent disclosure, investors have more opportunities to learn and partially capitalize information (through scientific dissemination, conferences, or complementary firm signals), thereby reducing the incremental surprise at *pdate*. Similarly, the time from pre-grant publication to grant is negative (–0.00446, $p<0.10$ in col. 2), consistent with prolonged prosecution being read by markets as added uncertainty or already partially priced information, which dampens the incremental grant-day reaction. In both cases, the negative signs suggest less event-day value uplift when disclosure is temporally distant from its preceding stage, rather than differences in ex-post diffusion speed (which we do not observe). Importantly, the coefficients on novelty and impact remain consistent with our baseline—novelty remains insignificant at patent stages, while impact is positive and significant—indicating that our core results are not driven by timing structure but by the content revealed at each stage.

Table 4. Regressions Controlling for Disclosure and Grant Timing

	ξ_{pdate}	ξ_{gdate}
Novelty	0.000273	0.000173
	(0.00248)	(0.000938)
Impact	0.00589**	0.00427***
	(0.00274)	(0.00124)
Length	-0.00471	-0.00180
	(0.00604)	(0.00187)
Gap	-0.00787*	-0.00446*
	(0.00448)	(0.00251)
_cons	0.170***	0.319***
	(0.0518)	(0.0192)
FEs		
Firm	Y	Y
Year	Y	Y
CPC_subclass	Y	Y
Obs	2,082	23,322
Adj R^2	0.775	0.763

Note: *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The numbers in parentheses represent robust standard errors clustered at the firm level.

4.4.2. Alternative Impact Measure: Window Forward Citations

First, we substitute our original measure of technological impact—window impact (short-term reuse within 180 days)—with the number of forward citations each patent receives from subsequent USPTO patents in the same time window. Results (Table 5) remain highly consistent with baseline findings.

At the scientific publication stage, the coefficient on novelty remains positively significant (0.0779, $p<0.05$), confirming robust evidence that higher scientific novelty significantly increases patent market value at initial scientific disclosure. At later pre-grant publication and grant stages, novelty remains statistically insignificant. In contrast, the forward-citation impact metric shows significantly positive associations with patent market value at both patent publication (0.00573, $p<0.10$) and grant stages (0.00943, $p<0.01$). Patents receiving more forward citations—indicating higher technological impact and diffusion—are valued more highly by the market upon patent disclosure.

Interestingly, at the scientific publication stage, forward citations exhibit a significant negative relationship (–0.254, p<0.01). This suggests that highly cited scientific papers may initially receive lower market valuation, potentially due to market uncertainty about translating fundamental science quickly into commercial innovation. Once patents formally disclose specific details, the market clearly recognizes the technological significance, leading to higher valuations for highly cited inventions. These findings reinforce our earlier conclusions that technological impact is strongly valued by markets, particularly after detailed patent information becomes available, confirming that our baseline results are not driven by metric choice.

Table 5. Alternative Impact Measure: Window Forward Citations

	ξ_{pub_adate}	ξ_{pdate}	ξ_{gdate}
Novelty	0.0779**	0.000878	0.000313
	(0.0362)	(0.00116)	(0.000912)
Citation	-0.254***	0.00573*	0.00943***
	(0.0636)	(0.00311)	(0.00293)
Length	-0.0704***	-0.00165	-0.00180
	(0.0202)	(0.00215)	(0.00187)
_cons	3.994***	0.283***	0.336***
	(0.196)	(0.0109)	(0.00930)
FEs			
Firm	Y	Y	Y
Year	Y	Y	Y
CPC_subclass	Y	Y	Y
Obs	2,082	22,733	23,322
Adj R^2	0.887	0.703	0.772

Note: *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The numbers in parentheses represent robust standard errors clustered at the firm level.

4.4.3. Higher Matching Threshold for PPP Linkages

Next, we apply a stricter threshold for identifying patents closely tied to scientific publications, focusing only on PPP with a matching score ≥ 3 (high confidence matches) suggested by previous research (Marx & Scharfmann, 2024; Kwon, 2024). This reduces the PPP subsample from 2,082 to 908 patents, capturing cases with clear scientific linkage.

Under this stricter definition, Table 6 shows that the positive effect of novelty on market value at the scientific publication stage becomes even stronger and remains significant (0.0892, $p<0.01$). This confirms that among clearly science-based patents, the market consistently assigns higher value to novelty.

At pre-grant publication stage, the PPP dummy remains positively significant (0.00486, $p<0.10$), suggesting a valuation premium associated explicitly with clear scientific linkage. The interaction *PPP × Novelty* also remains positive and significant (0.00479, $p<0.10$), reaffirming that the market's higher valuation of novelty strongly depends on the explicit scientific underpinning. At the patent grant stage, PPP interactions remain positive but statistically insignificant, consistent with baseline results.

Overall, restricting the PPP definition to higher-confidence matches confirms that the market consistently attributes higher valuations to scientifically grounded novelty, validating the robustness of our earlier findings.

Table 6. Higher Matching Threshold for PPP Linkages

	ξ_{pub_adate}	ξ_{pdate}	ξ_{gdate}
Novelty	0.0892***	0.000292	0.000482
	(0.0258)	(0.00102)	(0.00101)
Impact	-0.0460	0.00392**	0.00427***
	(0.0347)	(0.00132)	(0.00158)
PPP		0.00486*	0.0136
		(0.00291)	(0.0102)
PPP × Novelty		0.00479*	0.00398
		(0.00273)	(0.00246)
PPP × Impact		0.00533	0.00279
		(0.00570)	(0.00481)
Length	-0.0537	-0.00108	-0.00356
	(0.0558)	(0.00204)	(0.00196)
_cons	3.95***	0.289***	0.329***
	(0.350)	(0.0102)	(0.00966)

FEs			
Firm	Y	Y	Y
Year	Y	Y	Y
CPC_subclass	Y	Y	Y
Obs	908	22,733	23,322
Adj R^2	0.888	0.696	0.763

Note: *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The numbers in parentheses represent robust standard errors clustered at the firm level.

4.4.4. Patents' Non-Patent References (NPL) as Science Linkage

Finally, we introduce an alternative measure of scientific linkage based on the number of non-patent literature (NPL) references cited in each patent (in logarithm, *NPL_count*), as reported in Table 7. Instead of using a binary measure, we utilize the continuous measure *NPL_count* because more than 80% of our pharmaceutical patent samples cite at least one scientific paper, rendering a binary measure insufficient in capturing meaningful variation.[9] Employing a continuous NPL measure thus allows us to better quantify the degree to which a patent is embedded within scientific knowledge.

The main effect of *NPL_count* on patent valuation is positive but statistically insignificant, indicating that merely citing more scientific articles does not automatically yield a valuation premium when controlling for novelty, impact, and fixed effects. However, at the pre-grant publication stage, the interaction term (*NPL_count* × *Novelty*) is positive and significant (0.00153, $p < 0.05$). This result suggests that investors assign greater market value to highly novel patents that extensively reference scientific literature. Thus, extensive citation of scientific sources enhances the credibility and reduces uncertainty around novel technologies, reinforcing their perceived economic potential.

By contrast, the interaction term between *NPL_count* and technological impact remains positive but statistically insignificant. This indicates that the market's valuation of

[9] We use a confidence score (confscore) threshold of ≥8 to identify NPL citations. According to Marx & Fuegi (2020), over 80% of all NPL citations in their benchmark dataset have a perfect confidence score of 10. Moreover, when applying a threshold of 8 or higher, the precision of NPL citation identification exceeds 99.8%, with a recall of approximately 89%. Based on this performance, a cutoff of 8 is recommended as a high-quality filter for reliable NPL matches. Accordingly, we adopt this threshold as our primary specification.

technological impact—captured through forward citations and early diffusion—does not depend significantly on the extent of scientific referencing. Instead, technological impact is consistently valued based on observable evidence of diffusion and market influence, independent of explicit scientific citation intensity.

Overall, these findings align closely with established insights from innovation studies. Patents extensively citing scientific literature often represent frontier knowledge, commanding higher market valuations particularly when highly novel (Narin et al., 1997; Azoulay et al., 2019). Importantly, our analysis highlights a critical distinction between PPP and NPL measures: PPP provides a precise, explicit linkage between scientific discoveries and subsequent patented inventions, capturing direct paths of science-to-innovation transfer; NPL referencing reflects a more general scientific embeddedness and knowledge integration, capturing broader but less direct science-to-technology pathways. Consequently, the market consistently places a valuation premium on explicitly science-based novel inventions (captured by PPP), while the general intensity of scientific referencing (captured by NPL) primarily amplifies valuation when combined with high novelty. These results reinforce our earlier claim that only novelty benefits from scientific anchoring, while impact is priced via observable reuse—regardless of its scientific depth.

Table 7. Scientific Embeddedness via NPL Citation Intensity

	ξ_{pdate}	ξ_{gdate}
Novelty	0.000277	0.000213
	(0.00165)	(0.00231)
Impact	0.00434***	0.00422**
	(0.00165)	(0.00180)
NPL_count	0.00229	0.00302
	(0.00241)	(0.00193)
NPL_count × *Novelty*	0.00153**	0.00137
	(0.000727)	(0.000971)
NPL_count × *Impact*	0.00283	0.00328
	(0.00831)	(0.00671)
Length	-0.00105	-0.00174
	(0.00224)	(0.00187)
_cons	0.297***	0.350***
	(0.0124)	(0.0108)
FEs		
Firm	Y	Y

Year	Y	Y
CPC_subclass	Y	Y
Obs	22,733	23,322
Adj R^2	0.697	0.763

Note: *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The numbers in parentheses represent robust standard errors clustered at the firm level.

5. Conclusions

This paper shows that when information is disclosed and what is disclosed jointly determine how financial markets price patented innovation. Using stage-specific event studies in U.S. listed pharmaceutical firms, we document that market reactions are already sizable at the scientific publication stage for patents linked to peer-reviewed research (PPP), and that pre-grant publication reactions are comparable in magnitude to grant-day reactions. These findings broaden the prevailing grant-centric perspective in KPSS-based studies (Kogan et al., 2017) by establishing that markets incorporate value-relevant patent information earlier in the disclosure timeline.

Mechanistically, we uncover a clear shift in valuation patterns across stages. At the scientific publication stage, the market places a significant premium on novelty—the introduction of previously unseen technical concepts—particularly for PPP inventions, where peer review provides a credible certification of originality and rigor. These finding echoes recent literature emphasizing how credible scientific signals underpin higher expectations of future commercialization and technological breakthroughs (Arora et al., 2021; Marx & Fuegi, 2020). As disclosure progresses to pre-grant publication and grant, technological impact—measured by rapid citation and knowledge diffusion—proxied by the short-run reuse of the invention's novel content—emerges as the dominant factor influencing market valuation. This later-stage emphasis on evidence of diffusion and applicability aligns with the longstanding interpretation of forward-looking citation/usage measures as indicators of economic potential (Trajtenberg, 1990; Griliches, 1990; Hall, Jaffe & Trajtenberg, 2005), and aligns closely with recent work indicating that investors systematically reward technological generality and forward-looking citations as clear, observable evidence of economic potential (Kogan et al., 2017).

Moreover, our science-linkage tests reinforce this interpretation. PPP patents earn a modest premium at pre-grant disclosure, and, crucially, the PPP × Novelty interaction is positive and significant at that stage, showing that markets value novelty more when it is explicitly grounded in peer-reviewed science. By contrast, PPP × Impact is not statistically different from zero, suggesting that once diffusion is observable, markets price impact uniformly regardless of whether the invention is science-linked. Robustness analyses—tightening the PPP matching threshold and replacing the binary PPP indicator with a continuous measure of scientific embeddedness via non-patent literature (NPL) references—leave these conclusions intact: the novelty premium is amplified by scientific grounding, whereas the impact premium is stable across linkage definitions (Marx & Fuegi, 2020; Narin, Hamilton & Olivastro, 1997).

Collectively, our results provide a coherent synthesis: early-stage valuations reflect scientifically certified novelty; later-stage valuations reward observable diffusion; and science–technology linkages shape these trajectories in predictable ways.

The study also provides several potential policy implications. First, the results argue for patient, stage-aware support of science-based novelty. Early-stage, science-intensive inventions can generate substantial market value upon disclosure but typically require time to convert originality into widespread technological impact. Public R&D programs and translational initiatives (e.g., proof-of-concept grants, clinical validation infrastructure) should be structured with longer investment horizons and stable funding commitments that bridge the gap between early scientific signals and later diffusion. Clear and predictable IP rules around pre-grant disclosure—including transparency of pre-grant publication timing and scope—can further reduce information frictions and help investors interpret early signals without compromising appropriability.

Second, for firms, the findings motivate a two-track disclosure strategy. At early stages, firms can enhance valuation by foregrounding the rigor and novelty of their scientific foundations (e.g., publication, replication materials, external validations) to leverage the credibility channel that PPP embodies. At later patent stages, investor communications should pivot toward concrete evidence of diffusion and generality—early reuse by third parties, cross-application within the firm's own portfolio, and complementary assets that enable scale-up—since these signals most strongly drive market value once the technical details are public.

Third, for investors and analysts, the stage-dependent pricing we document suggests a refined playbook: treat early novelty (especially with peer-reviewed support) as a credible option on breakthrough potential, but update valuations toward demonstrated impact as patent details diffuse and complementary infrastructure takes shape. Portfolio construction should therefore account for different sources of risk and information across the disclosure lifecycle, rather than concentrating attention at grant alone.

While our evidence is drawn from pharmaceuticals—an industry where patents map more cleanly to products than in, for example, ICT—future research could examine cross-industry heterogeneity in the timing/content mechanism, the role of regulatory milestones that coincide with patent events, and the extent to which market microstructure (e.g., analyst coverage) mediates the novelty-to-impact transition. Designs that further isolate exogenous variation in disclosure timing (e.g., procedural delays in pre-grant publication) would help tighten causal interpretation. Extending text-based measures to capture quality-adjusted novelty and longer-horizon impact may also clarify how early signals translate into durable value.

REFERENCES

Agrawal, A., & Henderson, R. (2002). Putting Patents in Context: Exploring Knowledge Transfer from MIT. *Management Science*, *48*(1), 44–60.

Ahmadpoor, M., & Jones, B. F. (2017). The dual frontier: Patented inventions and prior scientific advance. *Science*, *357*(6351), 583–587.

Ahn, M. J., York, A. S., Wu, W., Suharto, Y., & Daim, T. (2015). On valuing biopharmaceutical product pipelines: An effectuation model and evidence. *Journal of Innovation and Entrepreneurship*, *4*(1), 14.

Ardito, L., & Svensson, R. (2024). Sourcing applied and basic knowledge for innovation and commercialization success. *The Journal of Technology Transfer*, *49*(3), 959–995.

Arora, A., Belenzon, S., & Sheer, L. (2021). Knowledge Spillovers and Corporate Investment in Scientific Research. *American Economic Review*, *111*(3), 871–898.

Arts, S., Hou, J., & Gomez, J. C. (2021). Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures. *Research Policy*, *50*(2), 104144.

Azoulay, P., Graff Zivin, J. S., Li, D., & Sampat, B. N. (2019). Public R&D Investments and Private-sector Patenting: Evidence from NIH Funding Rules. *The Review of Economic Studies*, *86*(1), 117–152.

Beyhaghi, M., Khashabi, P., & Mohammadi, A. (2022). Pre-grant Patent Disclosure and Analyst Forecast Accuracy. *Management Science*, *69*(5), 3140–3155.

Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2000). *Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)* (Working Paper 7552). National Bureau of Economic Research.

Finardi, U. (2011). Time relations between scientific production and patenting of knowledge: The case of nanotechnologies. *Scientometrics*, *89*(1), 37–50.

Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: Evidence from patent data. *Research Policy*, *30*(7), 1019–1039.

Fleming, L., & Sorenson, O. (2004). Science as a map in technological search. *Strategic Management Journal*, *25*(8–9), 909–928.

Franzoni, C., & Scellato, G. (2010). The grace period in international patent law and its effect on the timing of disclosure. *Research Policy*, *39*(2), 200–213.

Grabowski, H. (2002). Patents, Innovation and Access to New Pharmaceuticals. *Journal of International Economic Law*, *5*(4), 849–860.

Griliches, Z. (1990). *Patent Statistics as Economic Indicators: A Survey* (Working Paper 3301; NBER Working Papers). National Bureau of Economic Research, Inc.

Harhoff, D., Narin, F., Scherer, F. M., & Vopel, K. (1999). Citation Frequency and the Value of Patented Inventions. *The Review of Economics and Statistics*, *81*(3), 511–515.

Harhoff, D., Scherer, F. M., & Vopel, K. (2003). Citations, family size, opposition and the value of patent rights. *Research Policy*, *32*(8), 1343–1363.

Hegde, D., & Sampat, B. (2015). Can Private Money Buy Public Science? Disease Group Lobbying and Federal Funding for Biomedical Research. *Management Science*, *61*(10), 2281–2547.

Jaffe, A. B., & Trajtenberg, M. (2002). *Patents, Citations, and Innovations: A Window on the Knowledge Economy*. MIT Press.

Jeon, D., Ahn, J. M., Kim, J., & Lee, C. (2022). A doc2vec and local outlier factor approach to measuring the novelty of patents. *Technological Forecasting and Social Change*, *174*, 121294.

Ke, Q. (2020). Technological impact of biomedical research: The role of basicness and novelty. *Research Policy*, *49*(7), 104071.

Kelly, B., Papanikolaou, D., Seru, A., & Taddy, M. (2021). Measuring Technological Innovation over the Long Run. *American Economic Review: Insights*, *3*(3), 303–320.

Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2017). Technological Innovation, Resource Allocation, and Growth. *The Quarterly Journal of Economics*, *132*(2), 665–712.

Kwon, S. (2024). Underappreciated government research support in patents. *Science*, *385*, 936–938.

Lanjouw, J. O., & Schankerman, M. (2004). Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators. *The Economic Journal*, *114*(495), 441–465.

Lerner, J. (1994). The Importance of Patent Scope: An Empirical Analysis. *The RAND Journal of Economics*, *25*(2), 319–333.

Mansfield, E. (1991). Academic research and industrial innovation. *Research Policy*, *20*(1), 1–12.

Martens, T. (2023). The disclosure function of the U.S. patent system: Evidence from the PTDL program and extreme snowfall. *Review of Accounting Studies*, *28*(1), 237–264.

Marx, M., & Fuegi, A. (2020). Reliance on science: Worldwide front-page patent citations to scientific articles. *Strategic Management Journal*, *41*(9), 1572–1594.

Marx, M. & Fuegi, A. (2022). "Reliance on Science by Inventors: Hybrid Extraction of in-Text Patent-to-Article Citations." Journal of Economics & Management Strategy 31 (2): 369–392.

Masclans, R., Hasan, S., & Cohen, W. M. (2025). Measuring the commercial potential of science. *Strategic Management Journal*, *46*(9), 2199–2236.

Mohammad, A., & Jones, B. F. (2017). The dual frontier: Patented inventions and prior scientific advance. *Science*, *357*(6351), 583–587.

Mohammadi, A., & Khashabi, P. (2021). Patent disclosure and venture financing: The impact of the American Inventor's Protection Act on corporate venture capital investments. *Strategic Entrepreneurship Journal*, *15*(1), 73–97.

Murray, F. (2002). Innovation as co-evolution of scientific and technological networks: Exploring tissue engineering. *Research Policy*, *31*(8–9), 1389–1403.

Narin, F., Hamilton, K. S., & Olivastro, D. (1997). The increasing linkage between U.S. technology and public science. *Research Policy*, *26*(3), 317–330.

Packalen, M., & Bhattacharya, J. (2015). *New Ideas in Invention* (Working Paper 20922). National Bureau of Economic Research.

Poege, F., Harhoff, D., Gaessler, F., & Baruffaldi, S. (2019). Science quality and the value of inventions. *Science Advances*, *5*(12), 7323.

Putnam, J. D. (1996). *The value of international patent right*, Phd. Dissertation, Yale University, May 1996

Simeth, M., & Cincera, M. (2015). Corporate Science, Innovation, and Firm Value. *Management Science*, *62*(7), 1970–1981.

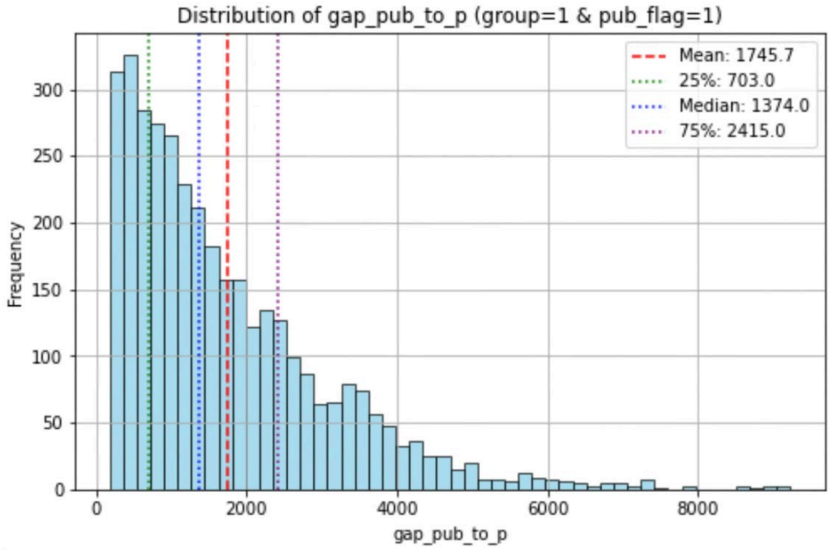
Trajtenberg, M. (1990). A Penny for Your Quotes: Patent Citations and the Value of Innovations. *The RAND Journal of Economics*, *21*(1), 172–187.

Zhao, P., Xing, Q., Dou, K., Tian, J., Tai, Y., Yang, J., Cheng, M.-M., & Li, X. (2025). From Words to Worth: Newborn Article Impact Prediction with LLM. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(1), 1183–1191.

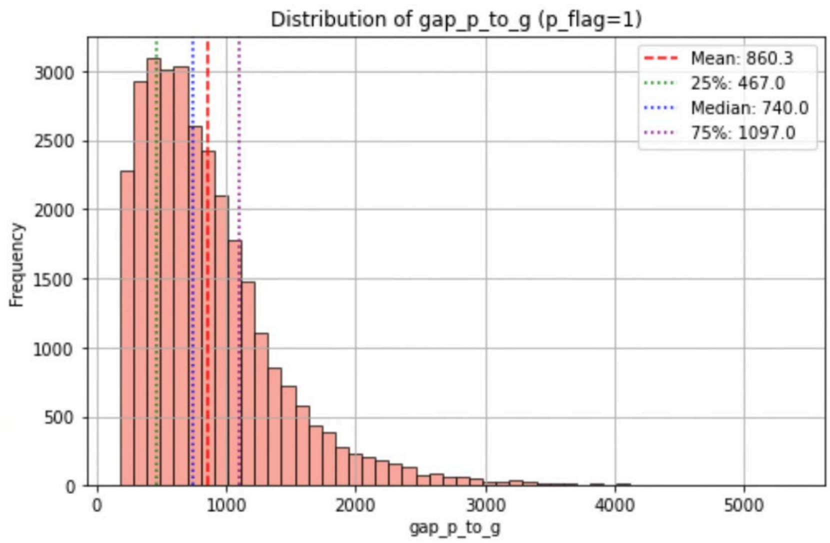
Zheng, Y. (2025). Can investors learn from patent documents? Evidence from textual analysis. *Contemporary Accounting Research*, *42*(2), 1331–1358.

Zhu, K., Malhotra, S., & Li, Y. (2022). Technological diversity of patent applications and decision pendency. *Research Policy*, *51*(1), 104364.

APPENDIX

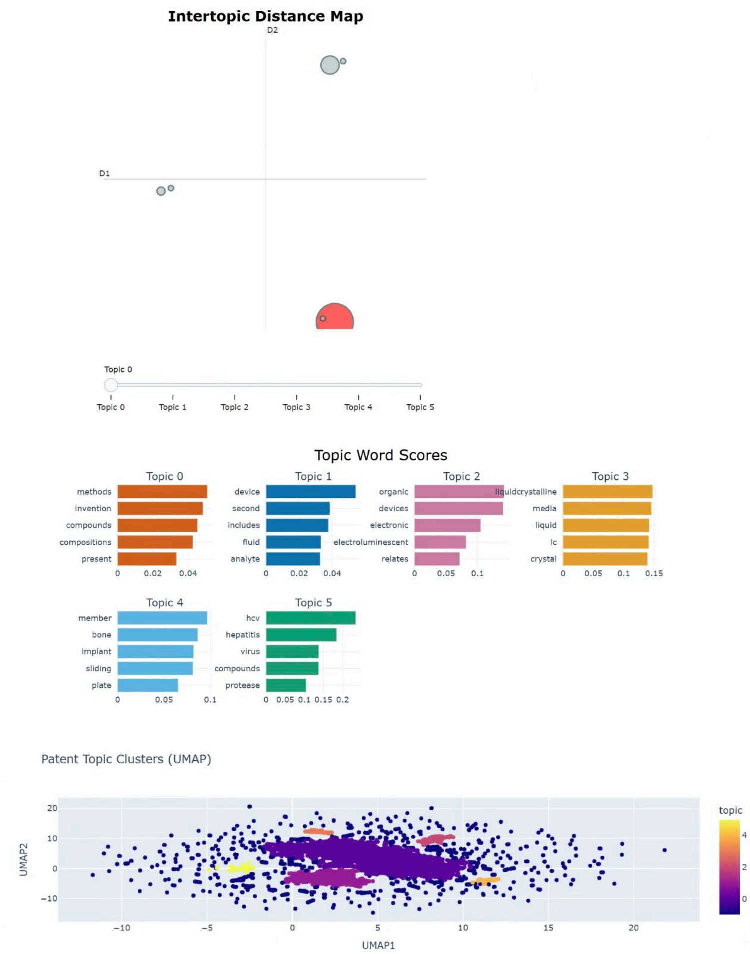


(a) Time from Scientific Publication to Patent Publication (PPP only)

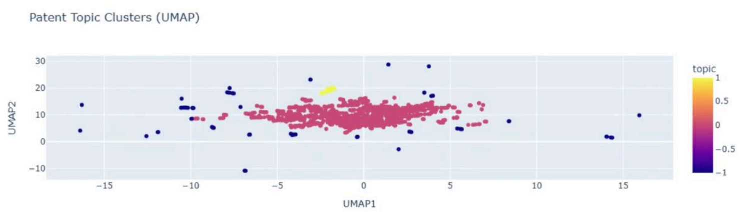


(b) Time from Patent Publication to Grant

Appendix Figure 1. Durations Between Disclosure Events



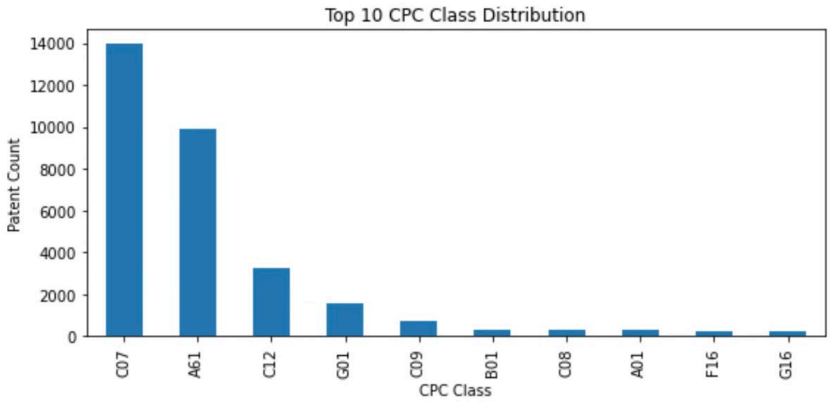
(a) Topic Clusters of Filtered Samples & UMAP Projection of Filtered Samples



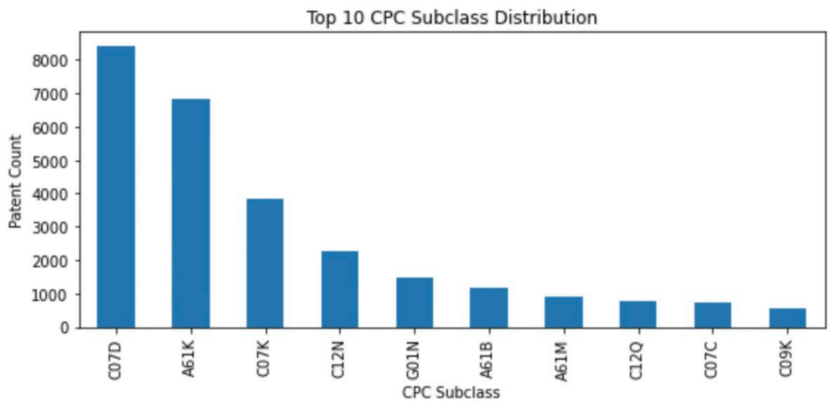
(b) UMAP Projection of PPP Patents Only

Appendix Figure 2. Semantic Clustering of Patents (BERTopic)[10]

[10] We use BERTopic to get the semantic clusters. For the whole filtered sample in (a), we restrict each cluster contains more than 250 patents; for PPP, the threshold is 50 patents and concentrated in Topic 0 and 1 in (a). From



(a) CPC Class Distribution

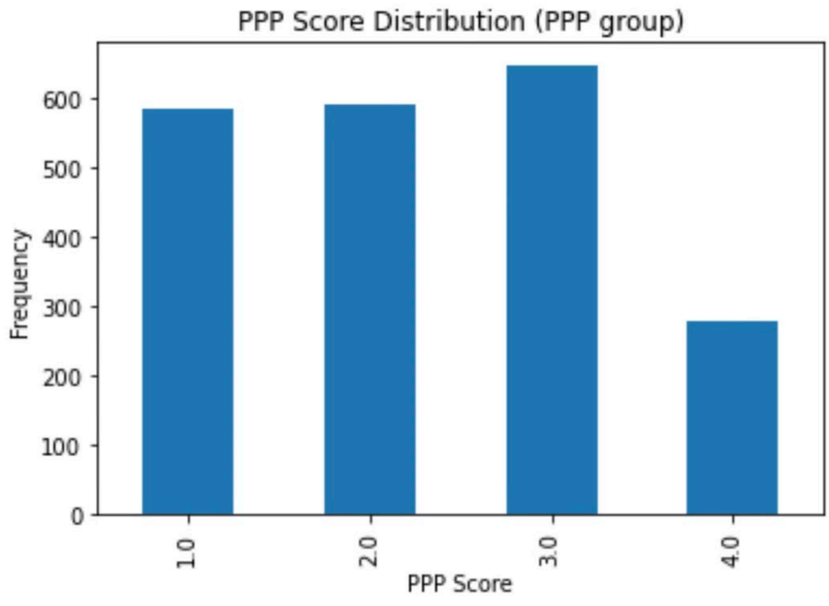


(b) CPC Subclass Distribution

Appendix Figure 3. Distribution of Technological Fields (CPC Classification)[11]

the content perspective, we don't find too much difference among our samples. Thus, we don't prefer to use the content-based controllers.

[11] Most sample patents are concentrated in a few CPC subclasses (even for CPC main class), particularly C07 (e.g., C07D, C07K – chemical compounds) and A61 (e.g., A61K – pharmaceutical preparations). To account for technological heterogeneity, we classify patents based on the top five most frequent CPC subclasses in our sample and group all remaining subclasses into an "other" category, resulting in six fixed-effect categories. This approach captures the majority of technological variation while ensuring model tractability. As a robustness check, we also tested specifications using CPC main sections instead of subclasses, and our results remain consistent.



Appendix Figure 4. Distribution of PPP Matching Scores[12]

[12] Marx & Scharfmann (2024) provide general information about the PPP dataset including four categories for PPP score ranged 1 to 4. Score 3 means more than 0.90 model prediction (high confidence) and 4 is more 0.98 model prediction and no false positives on testing set. Thus, we choose score 3 as the threshold.

Appendix Table 1. Summary of Variable Construction and Definitions

Vars	Definition
ξ	The estimated patent market value (1982 USD) in paper publication (ξ_{pub_adate}), pre-grant publication (ξ_{pdate}) and grant (ξ_{gdate}).
Novelty	The sum of extracted four types of newness comparing with all prior arts (both pre-grant and grant documents) at the time of the focal patent filed into USPTO, which is also constant across different stages.
Impact	The window impact measures how the sum of subsequent patents filed after the focal patent in 180-day window reusing focal newness in 3 stages respectively. Although we don't donate varied symbols for them, they are different in each stage.
Length	The number of words of the focal patent includes title, abstract and claims (cleaned texts exclude stopwords, symbols, etc).
Lag	Different durations between paper publication to patent publication (for PPP group) and patent publication to grant.
PPP	The binary variable represents whether the focal patent is linked with a peer-reviewed scientific publication in the PPP dataset.
NPL_count	The sum of NPL citations when the focal patent filed into USPTO.

Note: all of forementioned variables are transformed into logarithm except the binary *PPP*.

We also test using number of unique words as the *Length* controller, our conclusions still stand.

Appendix Table 2. PPP-only Regressions by Disclosure Stage

	ξ_{pub_adate}	ξ_{pdate}	ξ_{gdate}
Novelty	0.0926**	0.00411**	0.00222
	(0.0438)	(0.00189)	(0.00231)
Impact	-0.0511	0.00693	0.00706
	(0.0403)	(0.00579)	(0.00558)
Length	-0.0725***	-0.00210	-0.00678
	(0.0218)	(0.00579)	(0.00631)
_cons	3.972***	0.245***	0.319***
	(0.211)	(0.0345)	(0.0319)
FEs			
Firm	Y	Y	Y
Year	Y	Y	Y
CPC_subclass	Y	Y	Y
Obs	2,082	2,082	2,082
Adj R^2	0.887	0.723	0.767

Note: *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The numbers in parentheses represent robust standard errors clustered at the firm level.

Appendix Table 2 (a). Pairwise Correlation Matrix of Key Variables

	ξ_{pub_adate}	*Novelty*	*Impact*	*Length*
ξ_{pub_adate}	1.000			
Novelty	0.066	1.000		
Impact	-0.087	0.127	1.000	
Length	-0.120	0.302	0.015	1.000

	ξ_{pdate}	*Novelty*	*Impact*	*Length*
ξ_{pdate}	1.000			
Novelty	0.033	1.000		
Impact	0.059	0.167	1.000	
Length	-0.046	0.353	0.064	1.000

	ξ_{gdate}	*Novelty*	*Impact*	*Length*
ξ_{gdate}	1.000			
Novelty	0.055	1.000		
Impact	0.043	0.169	1.000	
Length	-0.046	0.354	0.068	1.000

Appendix Table 2 (b). Multicollinearity Diagnostics: Average and Maximum VIF

Variables	Paper Publication	Pre-grant Publication	Grant
Novelty	1.43	1.87	1.97
Impact	1.40	1.48	1.53
Length	1.02	1.44	1.47
Mean VIF	1.28	1.44	1.66

Appendix Table 3. PPP-only Regressions by Disclosure Stage

	ξ_{pub_adate}	ξ_{pdate}	ξ_{gdate}
Novelty	0.0926**	0.00411**	0.00222
	(0.0438)	(0.00189)	(0.00231)
Impact	-0.0511	0.00693	0.00706
	(0.0403)	(0.00579)	(0.00558)
Length	-0.0725***	-0.00210	-0.00678
	(0.0218)	(0.00579)	(0.00631)
_cons	3.972***	0.245***	0.319***
	(0.211)	(0.0345)	(0.0319)
FEs			
Firm	Y	Y	Y
Year	Y	Y	Y
CPC_subclass	Y	Y	Y
Obs	2,082	2,082	2,082
Adj R^2	0.887	0.723	0.767

Note: *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The numbers in parentheses represent robust standard errors clustered at the firm level.