The Value of Corporate Patent Utilization*

Jarrad Harford, Qiyang He, and Buhui Qiu

This Version: March 2025

Abstract

We use textual analysis of firm patent and product filings to construct a novel measure of patent utilization rate, which reflects the extent to which a firm's patents are applied in new products. We find that new products supported by more patents receive higher announcement returns and are more likely to be breakthrough products. Our firm-level patent utilization rate is positively associated with future new product development, market share growth, profitability, and firm valuation. The effects are predominantly driven by the utilization of high-value patents, and are more pronounced for firms in competitive product markets. We address endogeneity concerns using R&D tax credits as instruments and demonstrate robust findings across various tests. Our findings underscore the detrimental consequences of patent underutilization for firms and highlight the importance of integrating patents into the product development pipeline.

Keywords: Machine Learning, Textual Analysis, Patent Utilization, New Product Development, Market Share, Firm Valuation

JEL classification: G30, L25, O34

*Jarrad Harford is affiliated with The University of Washington; Address: 571 Paccar Hall, Foster School of Business, University of Washington, Seattle, WA 98195-3226, USA; Phone: +1-206-543-4796; Email: jarrad@uw.edu. Qiyang He is affiliated with The University of Sydney; Email: qiyang.he@sydney.edu.au. Buhui Qiu is affiliated with The University of Sydney; Email: buhui.qiu@sydney.edu.au. We thank Baojun Gao, Bin Li, Kaitang Zhou, participants of the 2025 China Accounting and Finance Conference, and seminar participants at Wuhan University and Huazhong Normal University for helpful comments and suggestions. All errors are our own.

"*...There is an incredible amount of technology that's packed into the product. There are 5,000 patents in the product (Vision Pro) and it's, of course, built on many innovations that Apple has spent multiple years on, from silicon to displays and significant AI and machine learning.*"

— Tim Cook, *February 01, 2024, Apple Inc. 2024Q1 Earnings Call*

1 Introduction

Patents are crucial intellectual assets that incentivize technological innovation. In recent years, firms have increasingly engaged in strategic patenting—their main purpose is not to commercialize the associated technologies through new product offerings but rather to restrict competitors from pursuing future innovations (Gilbert and Newbery, 1982; Argente et al., 2020). This defensive action has led to "patent portfolio races," with a dramatic increase in the number of patent production accompanied by a noticeable decline in patent quality and stagnating productivity growth (Hall and Ziedonis, 2001; Choi and Gerlach, 2017; Bloom et al., 2020; Kalyani, 2022).[1] This paper aims to deepen our understanding of firms' patent utilization rate in new product development and its implications for future firm performance.

The underutilization of patent portfolios in new product development can erode firms' competitive positions in product markets and weaken their operating performance. The collapse of Eastman Kodak Company represents a prominent example: In the 1980s, Kodak was a leading camera film producer and the fifth largest patent inventor (Moretti, 2021). However, with the rise of digital photography, Kodak's market share began to decline substantially. It is noteworthy that Kodak engineers Gareth Lloyd and Steven Sasson had developed and patented the first digital camera as early as 1977.[2] Yet, Kodak refused to incorporate this breakthrough innovation into the product pipeline, as the management team feared it would cannibalize the firm's film-based business. This strategic decision ultimately proved to be

[1] Figure A1 shows that since 1996, the number of patents per capita in the U.S. has more than doubled, but the number of highly-cited patents, breakthrough patents, and creative patents per capita have sharply decreased.

[2] The patent is titled "Electronic still camera" (US4131919A), filed in 1977 and granted in 1978. For technical details, see https://patents.google.com/patent/US4131919A/en.

a critical mistake. Digital photography soon emerged as a dominant technology and Kodak encountered a substantial decline in market share. The company eventually filed for bankruptcy in 2012.

Therefore, understanding the extent to which a firm's patent portfolio contributes to new product development offers significant implications for future firm performance. Despite its importance, this area has remained relatively underexplored due to the absence of an appropriate measure. In this paper, we fill this void by quantifying the corporate patent utilization rate in new product designs (patent utilization hereafter) using machine learning.

Specifically, we measure corporate patent utilization based on the premise that a patent is utilized/incorporated in a product if there is a high textual similarity between the patent filing text (obtained from PatentsView) and the new product launch text description (obtained from Capital IQ Key Development database).[3,4] One concern is that the language used in patent filings probably differs significantly from that in product text descriptions. Hence, the traditional "bag-of-words" approach, which requires exact overlap in terms, may inaccurately measure patent-product pair text similarity.

To overcome the challenge, we employ the pre-trained machine learning language model, *FastText* (Bojanowski et al., 2017), that builds on the architecture of the *Word2vec* model (Mikolov et al., 2013a). While both models represent words as semantic vectors, *FastText* takes a step further by accounting for rare or out-of-corpus words, providing a more nuanced understanding when comparing texts from different sources.[5] Leveraging the *FastText* model, we calculate the textual similarity score for each within-firm patent-product pair. Figure 1

[3] The assumption is similar in spirit to the innovation literature investigating knowledge diffusion across firms. Prior studies typically use patent citation a proxy for knowledge diffusion (see., e.g., Jaffe et al., 1993; Thompson and Fox-Kean, 2005; Singh and Marx, 2013; Arora et al., 2021; Fadeev, 2023; Cohen et al., 2023). Our criteria for patent utilization in products are more stringent in that we require high text similarity between patents and products.

[4] We focus on non-process (product) patents, as process patents primarily enhance production efficiency, which is not the focus of the study (Bena and Simintzi, 2022).

[5] The *Word2vec* model fails to provide semantic vector representations for words that are rarely seen or out of the training corpus. As our text data originates from patent filings and new product launch text descriptions, they likely contain extensive technological words that are rarely seen or entirely absent in conventional training corpora, which can lead to absence of vector representations for those words. Details are provided in section B3.3 in Appendix B.

illustrates the process of how we determine whether a patent is utilized in a new product. Suppose that firm A launches two new products in 2015, NP1 and NP2. We then source the three patents (PAT1, PAT2, and PAT3) that firm A has applied for (and later granted) in the preceding five years. We compute the textual similarity for each patent-product pair and consider a patent utilized if its similarity score is above the 80th percentile.[6]

Having identified patent incorporation in new products, we first examine whether new products supported by a greater number of patents exhibit higher quality. We measure product quality with two metrics: economic value, which is the cumulative abnormal stock return in a three-day window surrounding the new product announcement (Kogan et al., 2017), and breakthrough index, a text-based measure capturing a product's impact and novelty (Kelly et al., 2021). Our product-level analyses reveal that new products with greater patent incorporation tend to be of higher quality. These products receive significantly higher announcement returns and are more likely to be breakthrough products.

We next turn to the main research question of this study: does corporate patent utilization rate provide any positive implications for firms' future performance? To answer this question, we first generate a firm-year level patent utilization rate measure, calculated as the number of granted patents applied for (later granted) in the past five years by a firm and have been utilized in the new products launched by the same firm in the current year, scaled by the total number of granted patents applied for (later granted) in the past five years for that firm. This measure represents the proportion of patents from the past five years that a firm has utilized in new product development in the current year. Employing the firm-year level patent utilization rate, we start to investigate the cross-sectional value implications of corporate patent utilization. We shed light on four dimensions of firm performance: new product development, product market performance, profit improvement, and firm values.

First, we examine the relationship between a firm's patent utilization rate and its future new product development. Since patent utilization reflects a firm's propensity to commercialize its

[6] Results are robust to alternative cutoffs (e.g., 70th percentile, 90th percentile) and different product/patent portfolio windows.

intellectual assets, we expect it to positively influence future product development of the firm. Consistent with this expectation, we find that a higher patent utilization rate is associated with an increase in both the number of new products and product announcement returns. Moreover, firms with greater patent utilization are more likely to develop breakthrough products in the future. These relationships remain robust after controlling for a comprehensive set of firm-level characteristics, time-varying economic factors, time-invariant industry characteristics, and time-varying industry trends. In terms of economic significance, a one-standard-deviation increase in a firm's patent utilization rate corresponds, on average, to a 96.60% (80.58%) increase in the number of new (breakthrough) products and a 3.643-percentage-point increase in cumulative abnormal returns around new product announcements in the subsequent year.

Building on the finding that firms develop more and higher-quality new products when patent utilization is high, we further investigate whether higher patent utilization rates lead to improvements in firms' future product market performance. Our analyses confirm this hypothesis: corporate patent utilization rate is positively associated with firms' sales growth and market share growth over the following three years. On average, a one-standard-deviation increase in patent utilization corresponds to a 0.557 to 0.836 percentage-point increase in sales growth and a 0.517 to 0.915 percentage-point increase in market share growth. These findings suggest that firm-level patent utilization can predict short- to medium-term product market performance in the cross-section.

In addition, we find that firms with higher patent utilization rate experience significant improvements in profitability and market valuation. A one-standard-deviation increase in patent utilization rate is associated with an increase in gross profit margin by 0.239 percentage point, return on assets by 0.318 percentage point, operating cash flow by 0.279 percentage point, Tobin's Q by 1.632%, and market-to-book equity ratio by 2.627% in the subsequent year. These results are consistent with prior literature documenting positive relationships between innovation inputs/outputs and future firm performance and valuation (e.g., Lev and Sougiannis, 1996; Hall et al., 2001). Taken together, the baseline findings suggest positive indications between a firm's patent utilization rate and its future product development, product market performance, profitability, and valuations.

With the positive implications of patent utilization on various dimensions of firm performance, a natural question arises: are these effects primarily driven by the utilization of high-value patents? To answer this question, we construct two measures to capture the utilization rates of high-value and low-value patents, based on the economic value of patents estimated by Kogan et al. (2017). Our findings suggest that the utilization of high-value patents is the primary driver of positive firm outcomes. Additionally, we explore the heterogeneous effects of corporate patent utilization based on firms' product market competition. The results indicate that the benefits of patent utilization are more pronounced in competitive product markets.

The uncovered positive relationships between patent utilization rate and firm future performance may be subject to endogeneity concerns. For instance, high-performing firms may have greater incentives to utilize their patents to secure product market shares and maintain their leading positions. Some omitted variables may also correlate to both patent utilization rate and firm performance variables. To address these endogeneity concerns, we employ federal and state-level R&D tax credit variables, constructed by Bloom et al. (2013), as instruments for corporate patent utilization rates in instrumental variable regression analyses. According to Bloom et al. (2013), these R&D tax policy changes are largely random and unlikely to correlate with economic shocks. Consequently, variations in R&D tax credits can be considered exogenous to a firm's patent utilization.[7] The relevance condition is likely satisfied because increased R&D credits (or decreased R&D costs) enhance a firm's innovation capability, making it more likely to incorporate its patents in new product development and thus increasing the patent utilization rate. Our first-stage regression results confirm this hypothesis, showing a significant and negative association between federal and state-level R&D costs and corporate patent utilization rates. The weak-instrument test strongly rejects the null hypothesis of weak instruments.

[7] Similar to our study, Kogan et al. (2017) use R&D tax credits as an instrument for patent outputs to address endogeneity concerns. Hombert and Matray (2018) examine whether R&D-intensive firms are more resilient to trade shocks and also use R&D tax credits to predict R&D investment, mitigating endogenous selection of R&D expenses.

The rationale behind meeting the exclusion restriction lies in the purpose of R&D tax credits, which are intended to incentivize research and development activities rather than directly influencing firm performance metrics such as product market share, firm profitability or market valuation. Therefore, any observed relationship between R&D tax credits and firm performance is likely mediated through their impact on innovation activities, specifically patent utilization. Given that R&D tax credits do not directly affect future firm performance outside their influence on innovation, we can reasonably assume that the exclusion restriction holds. Essentially, R&D tax credits serve as an exogenous source of variation for patent utilization, isolated from direct effects on firm performance metrics. This ensures that our instrument's influence on firm performance is channeled solely through patent utilization rates. Our second-stage regression results show that the positive effects of patent utilization rates on firm performance and valuation remain robust. Additionally, over-identification tests in the second-stage regressions further support the validity of the exclusion restriction.

In addition, we also employ the exogenously determined patent utilization rates of distant rival firms as an instrument for the focal firm's patent utilization rate. The exogenous patent utilization rate of distant rival firms is calculated as follows: for each firm, we first use the federal and state tax credit components of R&D user cost to predict the firm's patent utilization rate (Bloom et al., 2013; Arora et al., 2021). Then, for each focal firm, we average the exogenously predicted patent utilization rates across its distant rival firms, which are defined as those peer firms in the focal firm's TNIC2 classification industry but not in its TNIC3 classification industry (Hoberg and Phillips, 2024). This average predicted distant rival patent utilization rate then serves as the instrumental variable for the focal firm's patent utilization rate. Since distant rivals operate in adjacent product markets but not directly in the focal firm's product market, the exogenously determined patent utilization rate of distant peers can only affect the focal firm's product market performance, profitability, and valuation through influencing the focal firm's patent utilization rate.

Our first-stage results show a significant correlation between the exogenously determined distant rival patent utilization rate and the focal firm's patent utilization rate. The second-stage results consistently indicate that the instrumented corporate patent utilization rate continues to positively correlate with the firm's new product development, product market performance, profit improvements, and overall firm valuation.

Finally, we demonstrate the robustness of our findings by: i) controlling for a firm's past innovation outputs or product similarity score (Hoberg and Phillips, 2016); ii) extending our analyses from the cross-sectional implications of corporate patent utilization to within-firm variation through the inclusion of firm fixed effects; iii) constructing alternative patent utilization measures using different percentile cutoffs, varying the patent portfolio window, or applying a 3-year moving average of the original measure; and iv) conducting intensive margin analyses by limiting the firm-year observations to those with at least one new product launch.

This study contributes to three strands of literature. Recent research demonstrates a puzzling macroeconomic trend that the sharp rise in patent production is accompanied by a significant decline in patent quality and stagnating productivity growth (Hall and Ziedonis, 2001; Bloom et al., 2020; Kalyani, 2022). A potential reason could be attributed to the "patent portfolio races," where firms seek to maintain their competitive positions by patenting ideas preemptively (Choi and Gerlach, 2017; Argente et al., 2020). This strategic purpose primarily aims to block other firms from developing future innovations, but not to commercialize the associated technologies via new product development, which leads to potential underutilization of their patent portfolios. This study employs machine learning to construct a novel measure of patent utilization rate, providing a deeper understanding of the extent to which firms incorporate patents in new product development. Our findings try to raise managers' attention that patent underutilization could bring detrimental consequences to the firms.

Second, our paper extends the literature that investigates the implications of corporate innovation on firm performance and valuation. Prior studies document positive relationships between traditional innovation inputs/outputs and future firm performance (e.g., Bloom and Van Reenen, 2002; Sougiannis, 1994; Lev and Sougiannis, 1996; Chan et al., 2001; Hall et al., 2001; Kogan et al., 2017). Recent studies have explored alternative measures of firms' intangible capabilities, such as innovation efficiency (Hirshleifer et al., 2013), innovation originality (Hirshleifer et al., 2018), research quotient (Cooper et al., 2022), technology differentiation (Arts et al., 2023), and technological obsolescence (Ma, 2021).

We complement the literature by introducing a new dimension to capture a firm's innovation capability. Specifically, we investigate how patent utilization influences new product development, market share, profitability, and firm valuation. Our approach and analyses provide robust empirical evidence on the broader benefits of effectively utilizing patents, highlighting the importance of not just obtaining patents but actively integrating them into the firm's product pipeline. This study also provides a novel metric for assessing a firm's innovation capability and future growth potential, aiding in more accurate valuation and investment decisions.

Third, this study contributes to the literature on textual analysis in economics and finance (e.g., Loughran and McDonald, 2011; Garcia and Norli, 2012; Gentzkow et al., 2019). Prior studies generally use "bag-of-words" approach to measure textual similarity (Hoberg and Phillips, 2016; Kelly et al., 2021; Argente et al., 2020). An emerging literature starts to adopt machine learning techniques to account for word semantics (Mikolov et al., 2013a; Pennington et al., 2014; Bojanowski et al., 2017).[8] For instance, Li et al. (2021) apply the *Word2vec* model to measure corporate culture. Hoberg and Phillips (2024) use the *Doc2vec* model to compute firm product market scope based on a firm's exposure to different industries. Similarly, Kogan et al. (2022) employ a machine learning model, *Glove*, to capture workers' technology exposure by calculating textual similarity between occupation descriptions and patent filings. This study leverages the *FastText* model to link products with patents within each firm, and develop a novel measure of corporate patent utilization.

The remainder of the paper is organized as follows. Section 2 describes our approach to measuring the patent utilization rate. In Section 3, we report the results on product-level patent integration and product announcement returns. Section 4 discusses the implications of corporate patent utilization on new product development, market share, profit improvement, and firm value. Section 5 explores the heterogeneity of the documented effects. Section 6

[8] Seegmiller et al. (2023) show that these machine learning approaches significantly outperform the conventional "bag-of-words" approach.

presents the results from the instrumental variable analysis and various robustness tests. Section 7 concludes. The Appendix A provides variable definitions and additional empirical results. The Appendix B provides technical details of our approach in measuring patent utilization rate.

2 Patent Utilization: Data and Measurement

This section describes how we construct the measure of a firm's patent utilization rate. In Section 2.1, we describe the sources of data used in the study. In Section 2.2, we compute the textual similarity score between a patent filing and a product description text. We regard a patent as utilized in a new product in a firm if the text description of the patent filing is *abnormally similar* to that of the new product description. Finally, we aggregate the patent-level utilization to firm level. Appendix B contains more technical details on the measurement of patent utilization.

2.1 Data

We obtain patent filing text from PatentsView, which provides title, abstract, brief summary text, patent claims, and detailed description sections for each patent granted since 1976. Consistent with prior literature (e.g., Kogan et al., 2022; Kelly et al., 2021), we exploit the full text of patent filings (i.e., aggregate all the five sections of a patent document into a patent-level corpus) for textual analysis. To match patents with the U.S. publicly listed firms, we rely on the linking table developed by Kogan et al. (2017), which matches each patent assignee with a PERMNO ID from CRSP if available. Hence, our final patent text sample consists of 2,544,432 patents generated by U.S. public firms from 1976 to 2022. Figure A2 illustrates an example of patent text filing from the Google Patents website.

We further collect product-related text description data from the Capital IQ Key Development database. After restricting the product-related text descriptions to the sample of U.S. publicly listed firms, we obtain 269,472 product-related announcements from 2002 to 2022. As suggested by Cao et al. (2018), there are generally four types of product-related announcements:

R&D progress, new product introduction, product improvement, and product retirement. In line with prior literature, we focus specifically on the category of new product introduction. To select new product introduction-related announcements, we build upon Cao et al. (2018) by using new-product-launch keywords and employing an advanced natural language processing technique, *FinBert*, to help us determine whether a product-related announcement is related to new product introduction. Please see Section B1 for detailed descriptions of the training sample construction, the *FinBert* fine-tuning process, and the model classification performance.[9] In Figure A3, we demonstrate an example of Apple Inc. announcing a new product in 2020.

After requiring firms to have at least one patent granted throughout their histories, our final sample consists of 125,329 announcements related to new product launches.[10] We follow standard text cleaning procedures (e.g., Kelly et al., 2021; Kogan et al., 2022) to preprocess the patent documents and new product announcement text description, which are discussed step-by-step in Section B2. Finally, we obtain stock return data from the Center for Research in Security Prices (CRSP), financial data from Compustat, and corporate patent quantity and quality data from Kogan et al. (2017). Table 1 reports the summary statistics of the variables used in this study. Table A1 in Appendix A provides detailed variable definitions and data sources.

[Please insert Table 1 about here]

2.2 Measuring Patent-Product Pair Textual Similarity

We assume that a patent is utilized in a new product if the patent-product pair textual similarity is abnormally high.[11] This critical assumption is similar in spirit to the innovation literature

[9] Panel A of Table A2 lists the new product launches keywords. Panel B further tabulates the classification performance in the testing sample. Our fine-tuned *FinBert* model can accurately classify 93% of the headlines. Panel C illustrates some (randomly) selected examples of new-product-introduction-related and non-related headlines predicted by our *FinBert* model.

[10] We also require our sample firms to have at least one new product launch in the key development database. Thus, our final sample contains 3,102 unique firms that have produced patents and launched products.

[11] Verifying whether a patented technology is utilized in a product poses a significant challenge as it requires consultations with technical experts. We acknowledge that high similarity may not indicate definate patent utilization in the new product. However, it does suggest that the new product is very likely to have been heavily influenced by or derived from the patented technology.

that leverages pair-wise patent citations to investigate knowledge diffusion across firms (see., e.g., Jaffe et al., 1993; Thompson and Fox-Kean, 2005; Singh and Marx, 2013; Arora et al., 2021; Fadeev, 2023). In essence, the literature hypothesizes that if a patent of Firm **A** cites a patent of Firm **B**, knowledge is diffused from Firm **B** to Firm **A**.[12] While the current data on new products does not specify information on patent utilization, we can infer the relationship between patents and products through their textual similarity.

On this basis, our first step is to measure textual similarity between patents and products. A conventional way to measure textual similarity in economics and finance literature is the "bag-of-words" approach (Hoberg and Phillips, 2016; Gentzkow et al., 2019; Kelly et al., 2021; Chen and Srinivasan, 2023). However, it does not account for semantic similarities between words. That is, words could possess similar meanings even if they are in different forms. For instance, the word "big" is semantically similar to the word "large," but the "bag-of-words" approach will count as a zero match.[13]

Importantly, the underestimation bias could be even more pronounced when comparing two documents from different text sources that exhibit diverse language styles (Seegmiller et al., 2023). In this study, we aim to compare the formal, standardized, and legalistic language used in patent filing text descriptions with the more informal and less structured tone in product announcement text descriptions. If we adopt the "bag-of-words" approach, the contrasting language styles of the two corpora could lead to sparse one-hot vectors, with many elements equal to zero and cosine similarity scores close to zero.

To overcome the issue, we exploit an advanced machine learning technique, *Word2vec* (Mikolov et al., 2013a), that can transform words into semantic, low-dimension, and dense vectors (embeddings) via neural network. Hence, words with similar semantic meanings can

[12] In a similar vein, Cohen et al. (2023) regard a firm as a user of an external patent if the firm has cited the patent previously.

[13] Consider an extreme case: document i contains the phrase "one beautiful house", while document j contains the phrase "a lovely dwelling". As humans, we can discern the closeness of the two documents. However, when using the "bag-of-words" approach, we transform the two documents into two one-hot vectors, $\mathrm{V_i}$ =[1, 1, 1, 0, 0, 0] and $\mathrm{V_j}$ = [0, 0, 0, 1, 1, 1]. We then compute the cosine similarity between the two vectors. In this example, we obtain a cosine similarity score of zero, which indicates that the two documents are unrelated. Please refer to Section B3.1 for more details on the challenges in "bag-of-words" approach.

have close spatial distance even if they are not exactly overlapped. We obtain pre-trained word embeddings from *FastText*, an extension of the *Word2vec* model developed by Bojanowski et al. (2017). In the following paragraphs, we briefly discuss how we use the *FastText* model to measure cosine similarities between patents and product texts. Sections B3.2 to B3.4 contain more details.[14]

First, we aggregate *FastText* word vectors to document (i.e., patent/product text) level using the following equation:

$$D_i = \sum_{X_j \in Z_i} w_{i,j} x_j \tag{1}$$

where D is a vector for document i, measured as the weighted average of the word vectors x for each word j in the set of words Z in document i. Following prior textual analysis literature (see, e.g., Loughran and McDonald, 2011; Hoberg and Phillips, 2016; Li et al., 2021; Kelly et al., 2021), we use the term-frequency-inverse-document-frequency (TFIDF) as our weighting scheme to give different weights w on word vectors based on the importance of the words in our corpus.

After obtaining a dense semantic vector for each document, we use the following equation to measure the cosine similarity between a patent document vector D_p and a product description text D_t within a firm f:

$$Sim_{p,t,f} = \frac{D_{p,f}}{||D_{p,f}||} \cdot \frac{D_{t,f}}{||D_{t,f}||} \tag{2}$$

Equation 2 emphasizes within-firm patent-product pair similarity because we want to measure a firm's self-invented patent utilization in its new product development.[15] It is worth noting that in this study we solely focus on non-process (product) patents (Bena and Simintzi, 2022), as

[14] Please refer to Section B3.2 for technical details and advancement on the *Word2vec* model, Section B3.3 for information on the *FastText* model, and Section B3.4 for thorough description on the measurement of patent-product pair textual similarity using *FastText*.

[15] It is worth noting that we randomly select 250 patent-product pairs from our sample and use OpenAI's new text embedding model (*text-embedding-3-small*) to compute their cosine similarity scores. We further examine the correlation between similarity scores generated by *FastText* and *OpenAI* and find a correlation of approximately 0.61, suggesting that *FastText*, despite being a more cost-effective option, performs reasonably well in capturing semantic similarity.

our goal is to understand whether product patents are utilized in new product development.[16] Moreover, for a firm's self-invented patents, we only focus on the firm's five-year patent application (later granted) portfolio before the launching date of a new product, since patents may become obsolescent as other technologies evolve (Ma, 2021).[17] The calculation process of within-firm patent-product pair similarity is illustrated in Figure 1. Suppose that Firm **A** launched two products in 2015, NP1 and NP2. We then source the three product patents (PAT1, PAT2, and PAT3) that Firm **A** applied (and later granted) in the five years before 2015. For each patent-product pair, we compute its text similarity score using Equation 2.

Next, since a majority of patent–product pairs within a firm have low textual similarity scores and are considered unrelated to one another, we follow the prior literature (e.g., Kogan et al., 2022; Hoberg and Phillips, 2016) to impose a stringent criteria: we only regard a patent as being utilized in a product if the textual similarity score is above 80th percentile of our sample patent-product pair scores.[18] In other words, for each within-firm patent-product pair, we replace the pair score with one if the raw similarity score is above 80th percentile, and otherwise replace it with zero.

In Panel A of Table A3, we demonstrate some examples of within-firm patent-product pair linkage. For each of the three randomly selected products, we show the 5 most (least) similar patents based on the patent-product similarity score. In Panel B of Table A3, we further provide excerpts from the text descriptions of the three new products, along with excerpts from the most and least similar patents for each product. These matching examples illustrate the effectiveness of the *FastText* model. For instance, the patent titled "Multi-functional hand-held device" filed in 2006 by Apple Inc. is most closely associated with the product "Apple IPhone 4," as their texts are semantically similar. In contrast, the patent titled "Transaction ID filtering for buffered programmed input/output (PIO) write acknowledgments" filed in 2009 by Apple

[16] Process-related innovations are of less interest in our study, as these patents primarily focus on improving production processes. For technical details on how to distinguish product innovations from process innovation, please see Section B3.5

[17] The USPTO requires that for patent applications filed after June 8, 1995, the terms of patents will end 20 years after the patent application date. In robustness tests, we also consider the 10-year patent application (later granted) portfolio of a firm and obtain qualitatively similar results.

[18] We also consider alternative percentile cutoffs such as 70th and 90th, and obtain qualitatively similar results.

Inc. is deemed as the least similar as its technical terms differ fundamentally from the iPhone 4 product description.

[Please insert Figure 2 about here]

Finally, having identified whether a patent is utilized by a firm, we can then measure a firm's patent utilization rate, *Pat. Utilization Rate*, as the number of granted patents applied for by a firm in the past five years and utilized in the new products launched by the same firm in the current year, scaled by the total number of granted patents applied for by that firm in the past five years. We replace *Pat. Utilization Rate* with zero if a firm does not launch any new product for a firm-year, but has applied for (later granted) at least one patent in the past five years. Alternatively, if a firm does not apply for (later granted) any patent in the past five years, we set *Pat. Utilization Rate* as missing.

Table 1 shows that the average corporate patent utilization rate is 35.1%, which is analogous to prior literature that surveys inventors to analyze the commercialization outcomes of inventive activity. For instance, using survey data on 3,162 patented inventions, Webster and Jensen (2011) find that around 40% are advanced to subsequent new product launches and production. Similarly, Amesse et al. (1991) document that around 43% of patents are commercialized in Canada. Figure 2 further illustrates the variation of the average corporate patent utilization rate over time, which is fluctuated at around 30% over time and a slight decrease is observed since 2013, indicating a potential increase in defensive patenting (which would decrease utilization). Figure 3 further illustrates the top 10 industries (2-digit SIC) with the highest rates of corporate patent utilization. It shows that five out of the 10 industries are related to the manufacturing sector, with *36: Electronic & Other Electric Equipment* ranked the highest.

[Please insert Figure 3 about here]

3 Patent Integration and Product Quality

Having identified patent incorporation in new products, in this section, we start to examine whether new products supported by a greater number of patents exhibit higher quality. To measure the number of patents utilized in each new product, we aggregate the within-firm patent-product pair scores (either one or zero) to product announcement event level.[19] To capture product quality, we focus on a product's economic value and breakthrough index.

3.1 New Product Economic Value (Announcement Return)

We use product announcement return, *CAR (-1, 1)*, which is the cumulative abnormal stock return during a three-day event window (-1, 1) around the new product announcement event, to proxy for a product's economic value, consistent with Kogan et al. (2017) and Mukherjee et al. (2017). We conduct the event-level regression analyses using the following equation:

$$Y_{i,f,j,t} = \beta_1 Log\ (1 + \#Patents\ Utilized_{i,f,j,t}) + \beta_2 Controls_{i,f,j,q-1} + \theta_j + \mu_t + \epsilon_{i,f,j,t} \tag{3}$$

In Equation 3, Y represents the three-day CAR of product i of firm f in industry j on the product announcement event date t, and *Log (1+#Patents Utilized)* is the natural logarithm of one plus the number patents utilized by the product i of firm f in industry j in event date t. Because a firm may launch multiple products on single event date, we generate two variables to account for the situation of product bundle launching: *Log (1+#Patents Utilized^{Sum})*, which is the natural logarithm of one plus the sum of the number of patents that are utilized in new product(s) for an event date, and *Log (1+#Patents Utilized^{Average})*, which is the natural logarithm of one plus the average number of patents that are utilized in new product(s) for an event date. We further control for a variety of firm characteristics, such as firm size (*Firm Size*), firm age (*Log(Firm Age)*), leverage ratio (*Leverage*), research and development expenses (*R&D*), return on assets (*ROA*), cash holdings (*Cash*), Tobin's Q (*Log(Tobin's Q)*), sales growth (*Sales growth*), and past stock return (*Stock Return*), all measured one quarter before the product announcement quarter. Moreover, we control for the length of the product announcement text (*Log(Product Text Length)*) and the number of new products (*Log(1+#New Products Launched)*) that have already been launched by firm f in the same year before the

[19] For each product announcement event, a firm may launch a single product or a bundle of products.

event date. Finally, we include industry fixed effects θ and event-date fixed effects μ. The results are reported in Table 2.

[Please insert Table 2 about here]

In Column 1 (5), we examine the simple relationship between the number of patents utilized in new products and the new products' announcement return after controlling for the firm and product characteristics. In Column 2 (6), we further include industry fixed effects to control for time-invariant industry characteristics. In Column 3 (7), we include event-date fixed effects to account for the time-varying economic conditions. In Column 4 (8), we replace the industry and event-date fixed effects by industry-by-event-date fixed effects to account for time-varying industrial shocks. We find that, across all specifications, the coefficient estimates of *Log(1+#Patents Utilized^{Sum})* (*Log(1+#Patents Utilized^{Average})*) are positive and statistically significant at least at 5% level, indicating that new products with more patents integrated are valued higher by the stock market. The economic magnitude is meaningful. Take Column 3 as an example: it implies that a 1-percentage-point increase in *Log(1+#Patents Utilized^{Sum})* is associated with a 0.024% (= 0.01*0.024) increase in the product announcement return.[20]

Overall, this finding aligns with the intuitive expectation that more innovative products are highly valued by the market while also offering an objective measure of a product's degree of innovation.

3.2 New Product Breakthrough Index

We further explore whether products that incorporate a greater number of patents are more likely to become breakthrough products. Following Kelly et al. (2021), we define breakthrough products as those that not only introduce novel features but also shape the development of future products. Building on this idea, we construct a text-based breakthrough index that captures both the novelty and impact of a product, offering a fresh perspective on what sets truly transformative products apart.

As aforementioned, a novel product is defined as one that is distinct from prior products.

[20] Note that the average product announcement return is 0.190%.

We follow Kelly et al. (2021) to measure a product's novelty as the inverse of its textual similarity with the prior products, which is as follows:

$$BS^5{}_{\mathrm{j}} = \sum_{i \in \beta^5_{j,m}} \rho_{j,i} \tag{4}$$

where BS denotes the backward similarity of product j. $\rho_{j,i}$ is the pairwise similarity between product j and i. $\beta^5_{j,m}$ denotes the set of previous products that are launched in the 5 years before product j's offering and that are in the same product market m (based on parent firms' TNIC3 classification (Hoberg and Phillips, 2016) as product j). Intuitively, novel products should have low backward similarity (BS) with the prior products.

On the other hand, an impactful product should shape future innovations, exhibiting high similarity with subsequent products. Thus, we measure a product's impact as follows:

$$FS^5{}_{\mathrm{j}} = \sum_{i \in \alpha^5_{j,m}} \rho_{j,i} \tag{5}$$

Similarly, FS denotes the forward similarity of product j. $\rho_{j,i}$ is the pairwise similarity between product j and i, and $\alpha^5_{j,m}$ denotes the set of future products that are launched in the 5 years after product j's offering and that are in the same product market m (based on parent firms' TNIC3 classification) as product j. Thus, an influential product will have high similarity (FS) with future innovations.

Finally, the product breakthrough index, *BreakthroughIndex*, which reflects the novelty (backward similarity BS) and impact (forward similarity FS) of new products, is measured as:

$$BreakthroughIndex^5{}_{\mathrm{j}} = \frac{FS^5{}_{\mathrm{j}}}{BS^5{}_{\mathrm{j}}} \tag{6}$$

The breakthrough index of a product tends to be higher if it exhibits low backward similarity with prior products (which is novel) but high forward similarity with subsequent products (which is impactful). To account for potential time-varying factors—such as fluctuations in the number of new products launched every year and changes in language change over time—we follow Kelly et al. (2021) to adjust the breakthrough index by removing year fixed effects. We present the results between the number of patents utilized in new products and the breakthrough index in Table 3.

[Please insert Table 3 about here]

From Columns 1 to 4, we observe that the coefficients of *Log(1+#Patents Utilized)* are positively and significantly related to the *BreakthroughIndex* at the 1% significance level. These results consistently show that new products supported by a greater number of patents tend to have higher breakthrough indices, suggesting that they are more novel and impactful. Furthermore, we examine whether these patent-embedded new innovations are more likely to be breakthrough products, defined by an indicator variable, *1 (Breakthrough Product)*, that equals one if the *BreakthroughIndex* is above 95th percentile and zero otherwise (Kelly et al., 2021). The results presented in Columns 5 to 8 of Table 3 align with our expectation: New products with more patents embedded are more likely to be breakthrough products.

Taken together, the results in Section 3 indicate that innovative products with a greater number of patents embedded tend to exhibit higher quality. These products are more highly valued by the market and have a greater likelihood of becoming breakthrough products.

4 The Implications of Corporate Patent Utilization

Since Schumpeter introduced the concept of creative destruction, economists have developed various endogenous growth models demonstrating that technological innovation is a central driver of economic growth and firm success (e.g., Aghion and Howitt, 1992; Lentz and Mortensen, 2008; Akcigit and Kerr, 2018). Empirical research consistently shows that innovation capabilities are a key determinant of future firm performance (e.g., Hall et al., 2001; Hirshleifer et al., 2013; Kogan et al., 2017; Cooper et al., 2022). However, recent trends reveal that many firms increasingly file patents as a strategic tool to block competitors rather than to drive genuine innovation. This practice risks stifling research productivity and impeding technological advancement, with potentially adverse effects on long-term economic growth (Bloom et al., 2020; Kalyani, 2022). Over time, unused patents may become obsolete, offering little practical application in new product development (Ma, 2021).

In contrast to simply tracking patent filings, out study introduces a new metric of innovation strength: the proportion of patents that are incorporated into new products. By focusing on patent utilization, we offer a metric of how well a firm translates its innovation efforts into tangible product market outcomes. This approach captures the extent to which a firm's patents contribute to new product development, offering fresh insights into the role of patents in sustaining competitive advantage and driving growth. In this section, we aggregate patent-product pair scores at the firm level to generate a measure of corporate patent utilization rate. This measure effectively captures the extent to which a firm's past patent portfolio is incorporated into new products. We then explore the implications of patent utilization for a firm's future performance, focusing on four key dimensions: new product development, product market performance, profit improvement, and firm value.

4.1 Corporate Patent Utilization and New Product Development

First, we shed light on the association between a firm's patent utilization rate and its future new product development. Since patent utilization rate indicates a firm's proficiency in commercializing patents, we anticipate that firms with higher rates of patent utilization will produce more (and higher-quality) new products. To investigate this research question, we employ the following firm-year regression model:

$$Y_{f,j,t+1} = \beta_1 Pat.\ Utilization\ Rate_{f,j,t} + \beta_2 Controls_{f,j,t} + \theta_j + \mu_t + \epsilon_{f,j,t} \tag{7}$$

The dependent variable Y represents the new products development of firm f in industry j in year t+*1*. To evaluate a firm's new product development, we follow Mukherjee et al. (2017) to concentrate on a firm's number of new products launched, which is measured as the raw number of new product announcements (*#New Products*) a firm releases in a year.[21] To capture the quality of new products, we follow Kogan et al. (2017) and Mukherjee et al. (2017) to estimate the economic value of a new product using the cumulative abnormal returns around the product announcement. We then generate the variable, *Sum CARs*, which is calculated as the sum of all positive three-day cumulative abnormal stock returns of the new products a firm launches in a year. Moreover, we create another outcome variable, *#Breakthrough Products*, that measures the number of breakthrough products a firm develops in a year.[22]

The independent variable, *Pat.Utilization Rate*, represents the patent utilization rate of firm f in industry j in year t. We also include a variety of standard firm-level controls: firm size (*Firm Size*), firm age (*Log(Firm Age)*), leverage ratio (*Leverage*), research and development expenses (*R&D*), return on assets (*ROA*), cash holdings (*Cash*), Tobin's Q (*Log(Tobin's Q)*), sales growth (*Sales Growth*), and past stock return (*Past Stock Return*), all measured in year t. Additionally, we control for a firm's new product intensity (*#New Products/Sales*), as the patent utilization rate may be positively correlated with the number of new products launched in the same year.[23] Lastly, we include industry fixed effects (θ) and year fixed effects (μ) to account for time-invariant industry characteristics and time-varying economic factors. The results are presented in Table 4.

[Please insert Table 4 about here]

In Columns 1 and 2, we investigate the relationship between a firm's patent utilization rate and its one-year-ahead raw number of new products (*#New Products*); in Columns 3-4,

[21] Consistent with Mukherjee et al. (2017), we primarily focus on major new product introductions by firms. Specifically, we count only new products with cumulative abnormal returns (CARs) above the 80th percentile in a given calendar year. To test the robustness of our results, Table A4 examines highly valued new products, defined as those with CARs above the 95th percentile. We construct two outcome variables: *#Highly-Valued New Products*, which represents the number of highly valued new products a firm develops in a year, and *1(Highly-Valued New Products)*, a binary indicator that equals one if a firm launches at least one highly valued new product in a year and zero otherwise. Our findings remain consistent, showing that firms with higher patent utilization rates tend to develop more highly valued new products and are more likely to introduce such products in the following year. Additionally, we confirm that our results remain qualitatively similar even when considering all new product launches.

[22] Breakthrough products are defined as those products with breakthrough indices above the 95th percentile.

[23] To further address the concern that the relation between patent utilization and future firm performance may depend on whether firms launch new products in a given year, we also conduct intensive margin analyses as a robustness check by restricting the firm-year observations to those with at least one new product launch, and the results remain qualitatively similar.

we further shed light on how a firm's patent utilization rate is related to its one-year ahead new product CARs (*Sum CARs*); in Columns 5-6, we explore whether firms with higher patent utilization rate will develop more breakthrough products (*#Breakthrough Products*) or not. Cohn et al. (2022) find that when the dependent variables are count-based (Columns 1-2 and 5-6), OLS regressions can lead to biased estimates. To address the biased estimation issue, we follow their suggestion by employing the fixed-effects Poisson model in Columns 1-2 and 5-6 to produce more reasonably efficient estimates. Finally, Columns 1, 3, and 5 include industry and year fixed effects, while Columns 2, 4, and 6 include industry-by-year fixed effects to further account for time-varying industrial economic changes.

Consistent with our expectation, the results in Table 4 show that *Pat.Utilization Rate* is positive and significantly (all at 1% level) associated with the firm's new product quantity and quality in the subsequent year. The economic magnitude is meaningful. A one-standard-deviation increase in a firm's patent utilization rate, on average, will lead to a 0.676-log-point (i.e., 0.398*1.698) or 96.60% (i.e., exp(0.676)-1) increase in its one-year-ahead number of new products, a 3.643-percentage-point (i.e., 0.398*9.149) increase in subsequent year's cumulative abnormal returns of new products, and a 0.591-log-point (i.e., 0.398*1.485) or 80.58% (i.e., exp(0.591)-1) increase in the number of breakthrough products in the subsequent year.

Overall, the results suggest that a higher patent utilization rate is linked to both a greater quantity and higher quality of future new products. This finding highlights the predictive power of patent utilization in shaping a firm's future product development at the cross-sectional level.

4.2 Corporate Patent Utilization and Product Market Performance

In the prior subsection, we demonstrated that corporate patent utilization rate is associated with higher quantity and better quality of future new products. According to endogenous growth theories, firms' performance will be enhanced as they introduce more innovative new products. Building upon the documented results and the theoretical underpinnings, in this subsection, we investigate how a firm's short to medium-term product market performance evolves with its patent utilization rate.

Similarly, we use Equation 7 to examine the relation between firms' patent utilization and future product market performance. To measure firms' product market performance, we follow the prior literature (e.g., Campello, 2006; Fresard, 2010; Billett et al., 2017) and concentrate on three outcome variables: sales growth (*Sales Growth*), which is measured as the natural logarithm of total sales of a firm in the current year minus that of the previous year; market share growth (*MSG(FF49 or SIC4)*), which is measured as the sales growth of a firm in a year minus the industry (Fama-French 49 industries or 4-digit SIC) median sales growth in the same year. The results are reported in Table 5.

[Please insert Table 5 about here]

In Panel A, we shed light on sales growth, while in Panels B and C, we further look into the market share growth of the firms. Across the three panels, we find that the coefficient estimates on *Pat.Utilization Rate* are positive and statistically significant, indicating that patent utilization rate is associated with better product market performance in the next three years. The results are robust to controlling for industry and year fixed effects (Columns 1, 3, and 5 in each panel) or industry-by-year fixed effects (Columns 2, 4, and 6 in each panel). In terms of economic magnitude, a one-standard-deviation increase in *Pat.Utilization Rate*, on average, corresponds to a 0.557 to 0.836 percentage-point increase (i.e., 0.398*0.014 to 0.398*0.021) in sales growth, and a 0.517 to 0.915 percentage-point increase (i.e., 0.398*0.013 to 0.398*0.023) in market share growth in the subsequent three years.

Importantly, we control for a firm's new product intensity (*#New Products/Sales*) across all panels. As expected, new product intensity is positively and significantly associated with both future sales growth and market share growth. However, the significant relationship is concentrated in the first year and diminishes over the following two years. These findings suggest that the *Pat.Utilization Rate* offers incremental information beyond a firm's new product introductions. While new product launches drive short-term product market outcomes, patent-backed product introductions have more sustained impacts on future product market performance, especially over the longer term.

In summary, this subsection shows that a firm's patent utilization rate positively predicts future product market performance. These results are in line with the endogenous growth theories that creative destruction (in this study, patent utilization in new product development) contributes to future firm growth.

4.3 Corporate Patent Utilization and Profit Improvement

Having documented the positive relationships between a firm's patent utilization rate, future new product development, and product market performance, we further explore the implications of patent utilization rate on a firm's future profitability changes. We employ three variables to measure changes in a firm's profitability: the change of gross profit margin ($\triangle GPM$), the change of return on assets ($\triangle ROA$), and the change of operating cash flow ($\triangle OCF$), all measured in year *t+1*. Similar to Section 4.2, we conduct the analyses using Equation 7 while also controlling for a firm's new product intensity (*#New Products/Sales*). The empirical results are presented in Table 6.

[Please insert Table 6 about here]

Again, we find a positive relationship between patent utilization rate and future profitability change. The results from Table 6 show that an increase in patent utilization rate of a firm is associated with positive and significant improvement in its one-year-ahead profitability, measured by $\triangle GPM$, $\triangle ROA$, and $\triangle OCF$ respectively. A one-standard-deviation increase in *Pat.Utilization Rate* is associated with an increase in gross profit margin by 0.239 (0.398*0.006) percentage point, return on assets by 0.318 (0.398*0.008) percentage point, and operating cash flow by 0.279 (0.398*0.007) percentage point of a firm in the subsequent year. Interestingly, although positive, we do not find a statistically significant relation between a firm's new product intensity and its profitability improvement in the subsequent year, with the exception of Column 4, where the coefficient on *#New Products/Sales* is positive but only marginally significant. Overall, the results indicate that firms utilizing more patents have significantly higher profit improvements in the future.

4.4 Corporate Patent Utilization and Firm Value

We next examine the relationship between corporate patent utilization rate and future firm values. A large strand of literature has investigated the implications of innovation inputs and/or outputs on the market valuations of firms. For example, regarding the innovation inputs, prior studies show that R&D expenditures and R&D intensity can positively predict future firm values and stock returns (see, e.g., Sougiannis, 1994; Lev and Sougiannis, 1996; Chan et al., 2001). On the innovation output side, several studies find that firms that generate more patent citations and patent economic values are associated with higher market valuations (see, e.g., Hall et al., 2001; Kogan et al., 2017). Recent studies also explore alternative measures of firms' intangible capabilities, such as innovation efficiency (Hirshleifer et al., 2013) and research quotient (Cooper et al., 2022). Both measures imply a positive relationship between firms' innovation strengths and future firm valuations.

Our constructed patent utilization rate can be regarded as a firm's ability to transform its patent portfolio into new product development. Therefore, a higher corporate patent utilization rate may be favorably valued by investors as it indicates the firm's capacity for patent commercialization, thereby increasing the economic value of its intangible assets. In this regard, we hypothesize that patent utilization rate of a firm is positively associated with its future firm valuations. Following prior literature, we measure firm values based on two proxies: Tobin's Q (*Log(Tobin's Q)*) and market-to-book ratio (*Log(MTB)*). Similar to previous subsections, we use Equation 7 to conduct the test. The results are reported in Table 7.

[Please insert Table 6 about here]

We find that the alternative measure of a firm's innovation strength, proxied by *Pat.Utilization Rate*, is positively and significantly associated with one-year-ahead firm valuation. A one-standard-deviation increase in a firm's patent utilization rate, on average, will lead to a 1.632% (2.627%) increase in its Tobin's Q (market-to-book ratio) in the subsequent year. Consistent with the literature, the results show a strong positive correlation between a firm's R&D intensity (*R&D*) and future firm valuation. It indicates that patent utilization rate has incremental information beyond a firm's innovation inputs.[24]

To summarize, the results in Section 4 show that higher patent utilization rate of a firm is associated with more and higher-quality new products, better product market performance, improved profitability, and higher firm values in the future. These findings together demonstrate that corporate patent utilization rate has positive value implications on future firm performance.

5 Heterogeneity of Patent Utilization Rate

In this section, we further explore the heterogeneous effects of corporate patent utilization on future firm performance. We first shed light on the economic values of patents utilized in firms' new products. We next investigate the role of product market competition.

5.1 High-value versus Low-value Patent Utilization

Patent quality varies with its scientific and economic value (Hall et al., 2001; Kogan et al., 2017). While scientifically advanced patents can attract future forward citations and generate positive knowledge externalities, economically significant patents should be the ones that most impact future firm financial performance. Therefore, we hypothesize that the documented positive implications of patent utilization on firms' future new product development, product market performance, profit improvement, and overall valuations would be primarily driven by the use of patents that possess significant economic value.

We conduct the analyses by acquiring patent economic value data from Kogan et al. (2017), which assesses the economic significance of each innovation by analyzing the stock market reaction around the patent grant date. To distinguish whether patents are economically meaningful, we regard a patent as high-value (low-value) patent if its economic value is above (below) the 80th percentile across the sample patent economic values. We then construct two separate

[24] There is no relationship between a firm's new product intensity and its future firm valuation. It is worth mentioning that we further control for a firm's patent outputs (both citation-weighted patent count and patent economic values) and related results are reported in Table 11 of the robustness check section. We continue to obtain qualitatively similar results that corporate patent utilization rate is positively and significantly related to future firm valuation.

measures: *High-Value (Low-Value) Pat.Utilization Rate*, which are measured as the number of granted high-value (low-value) patents that are applied for in the past five years by a firm and that are utilized in the new products launched by the same firm in the current year, scaled by the total number of granted patents applied for in the past five years for that firm.[25] Finally, we estimate Equation 7 to compare the role of *High-Value (Low-Value) Pat.Utilization Rate* on future firm performance. The results are reported in Table 8.

[Please insert Table 8 about here]

Consistent with our expectations, we find that the utilization of high-economic-value patents is the primary driver of our results. While the utilization of low-economic-value patents also contributes to future product development (Panel A), its economic impact is significantly smaller than that of high-value patent utilization. More importantly, the results indicate that the effects of low-value patent utilization on product market performance (Panel B), profitability changes (Panel C), and valuations (Panel D) are almost negligible. These pronounced positive effects of high-value patent utilization on ex-post firm performance reinforce the idea that market investors can effectively distinguish between high- and low-value patents at the time of their announcement.

5.2 Product Market Competition

Next, we investigate the role of product market competition. In competitive product markets, firms often struggle to differentiate their products from those of competitors. Therefore, corporate innovation and patent utilization are crucial for these firms to survive intense competition and gain market share. We thus hypothesize that the positive effects of patent utilization on firms' ex-post performance to be stronger for firms in competitive product markets.

We measure a firm's product market competition using TNIC HHI (Hoberg and Phillips, 2016), a sales-based Herfindahl-Hirschman index for the firm's industry that is defined by text-based network industry classifications. We further sort the full sample based on the median value of a firm's TNIC HHI, and split the full sample into the above-median (High) and below-

[25] The average high-value (low-value) patent utilization rate is 0.171 (0.180).

median (Low) subsamples. Finally, we re-estimate Equation 7 for each subsample and further report the p-values from Chow tests for the differences in the regression coefficient estimates between the two subsamples. The regression results are presented in Table 9.

[Please insert Table 9 about here]

We find that firms utilizing more patents in highly competitive product markets (lower TNIC HHI) tend to develop a greater number of higher-quality new products (Panel A), achieve stronger product market performance (Panel B), generate higher profitability (Panel C), and attain greater valuations (Panel D) compared to firms in less competitive markets (higher TNIC HHI). These results align with our expectation that the benefits of patent utilization are particularly pronounced for firms operating in highly competitive product markets.

Overall, the findings in this section suggest that the positive effects of patent utilization on firms' future performance are driven by the use of high-value patents and are more pronounced for firms in competitive product markets.

6 Additional Analyses

In this section, we discuss and address natural concerns surrounding the endogeneity of patent utilization. We partially alleviate the endogeneity issues using exogenously determined R&D cost variables as instruments. We further perform a series of additional checks to demonstrate the robustness of the findings.

6.1 Addressing Endogeneity Concerns

In the previous sections, we document positive implications between a firm's patent utilization and its future new product development, product market performance, profitability improvement, and valuations. However, these relations may be subject to endogeneity concerns. For instance, high-performing firms may have greater incentives to utilize their patents to secure product market shares and maintain their leading positions. Moreover, some unobserved factors may correlate with both the firm's patent utilization rate and future firm performance, leading to spurious relationships.

To help alleviate these endogeneity concerns, we follow Bloom et al. (2013), Hombert and Matray (2018), and Arora et al. (2021) and use the tax-induced changes in R&D user costs at the firm-level to instrument for corporate patent utilization rate. The R&D tax credit policy was first initiated at the federal level in the U.S. in 1981. Since then, state-level R&D tax credits have been gradually adopted across states. According to Bloom et al. (2013), a firm's R&D user cost is estimated using the Hall-Jorgenson user cost of capital formula, which incorporates both federal and state-level R&D tax credits components. The formula is as follows:

$$P_{f,t} = \frac{1 - D_{f,t}}{1 - \tau_{f,t}}[I_{\mathrm{t}} + \delta - \frac{\triangle P_{\mathrm{t}}}{P_{\mathrm{t\text{-}1}}}] \tag{8}$$

where $P_{f,t}$ is the Hall-Jorgenson user cost of R&D capital for firm f in year t, $D_{f,t}$ is the discounted tax credits and depreciation allowances, $\tau_{f,t}$ is a firm's tax rate, $\mathrm{I_t}$ is the real interest rate, δ is the depreciation rate of R&D capital, and $\frac{\triangle P_{\mathrm{t}}}{P_{\mathrm{t\text{-}1}}}$ is the growth of the R&D asset price. As $[I_{\mathrm{t}} + \delta - \frac{\triangle P_{\mathrm{t}}}{P_{\mathrm{t\text{-}1}}}]$ is common to all firms, only the tax price component of the R&D user cost $\frac{1-D_{f,t}}{1-\tau_{f,t}}$ is considered. Moreover, it can be inferred from this equation that the higher R&D tax credits ($D_{f,t}$), the lower the R&D user costs will be.

As in Bloom et al. (2013), $P_{f,t}$ can be further decomposed into two factors: federal tax credit component $P^F_{f,t}$, and the state tax credit component $P^S_{f,t}$, respectively. Specifically, the federal tax credit component ($P^F_{f,t}$) is determined by firm-level interactions with the federal tax rules (Hall, 1993). Firms benefit differently from federal R&D tax policies due to varying firm-specific requirements, making the federal tax-driven R&D user cost a firm-specific component.[26] Moreover, state R&D tax credit policies vary over time. Firms are thus exposed to these policies differently depending on the location of their R&D activities. Following Bloom et al. (2013), the state-level tax credit component of R&D user cost ($P^S_{f,t}$) is estimated by the interaction between state-specific R&D tax rules at a given time, and a firm's R&D activities, proxied by

[26] Bloom et al. (2013) list three reasons why federal R&D tax credit constitutes a firm-specific component: i) the tax credit allowance is based on the difference between a firm's actual R&D expenses and firm-specific tax base; ii) the tax credit depends on the taxable profit of a firm, and iii) these firm-level components are all further interacted with changes in the aggregated tax credit rate, deduction rules, and corporate tax rate.

the geographic distribution of its patent inventors.[27]

The prior literature suggests that changes in R&D tax credits are largely random and unlikely to be endogenous to shocks in the economic or political environment. Similarly, Bloom et al. (2013) and Hombert and Matray (2018) find no evidence that economic or political conditions can predict R&D policies. Therefore, these policies provide pseudo-random variation in a firm's R&D user cost.[28]

Our two-stage least squares (2SLS) instrumental-variable (IV) strategy uses the variation in a firm's R&D user costs as a source of exogenous variation in the firm's patent utilization intention. Importantly, the R&D tax credits decrease the marginal cost of R&D expenses for a firm, but do not directly affect the firm's product market performance. We conjecture that greater R&D tax credits (or the lower R&D user costs) can enhance a firm's innovation capability. Increased innovation capacity enables firms to allocate more resources to explore potential applications of their existing patents in new product development. Additionally, firms with stronger innovation capacity may develop more coherent innovation strategies and thus can build up a more applicable patent portfolio over time. Collectively, these mechanisms will contribute to a higher corporate patent utilization rate.

On this basis, we implement the 2SLS-IV approach as follows: in the first stage, we regress the patent utilization rate of firm i in year t, on both the federal and state-level tax credit components of R&D user costs of the firm in year *t-1*, with the same set of firm controls and fixed effect structures as Equation 7. In the second stage, we further regress each of the dependent variables used in the prior sections on the fitted value $Pat.\ \widehat{Utilization}\ Rate$. The results are reported in Table 10.

[Please insert Table 10 about here]

The first-stage results in Column 1 of Table 10 suggest that the federal and state-level R&D user costs of a firm (as discussed earlier, R&D tax credits are translated to R&D user costs using

[27] Following Bloom et al. (2013), the state tax credit component of R&D user cost can be formally estimated as follows: $P^S_{f,t} = \sum_s \theta_{i,s,t}\rho^S_{s,t}$. $\theta_{i,s,t}$ is 10-year moving average share of firm i's patent inventors in state s in year t. $\rho^S_{s,t}$ is state s's R&D tax price in year t.

[28] In a similar vein, Kogan et al. (2017) and Arora et al. (2021) use R&D tax credit as an instrument for the patent outputs of a firm to alleviate endogeneity concerns.

the Hall-Jorgenson formula) are both significantly and negatively associated with corporate patent utilization rate. The Cragg-Donald Wald F Statistics in the first-stage regression is over 38, strongly rejecting the null that the instruments are weak. Columns 2-12 report the second-stage results of regressing each dependent variable on the instrumented patent utilization rate. The results confirm that the positive impact of patent utilization on firms' new product development, product market performance, profitability, and valuations remain qualitatively unchanged. Notably, the Hansen J statistics in the second-stage regressions are well above 0.10 (except the positive but insignificant coefficient in Column 8), further supporting the validity of the exclusion restriction.

Following Bloom et al. (2013) and Arora et al. (2021), we also use the exogenously determined distant industry peers' patent utilization rate as an instrument for the focal firm's patent utilization rate. The exogenous patent utilization rate of the distant peer firms is measured as follows: first, we use the federal and state tax credit components of R&D user cost to predict each firm's patent utilization rate. Then, for each focal firm, we take the average of the exogenously predicted patent utilization rate across its distant industry peer firms. Importantly, following Hoberg and Phillips (2024), we define distant peers as those within the focal firm's TNIC2 classification but outside its TNIC3 classification.[29] These distant peers operate in adjacent markets but not directly within the focal firm's product market. As suggested by network econometrics literature (Bramoullé et al., 2009; Cohen-Cole et al., 2014) and Hoberg and Phillips (2024), the endogenous effects of patent utilization can be mitigated when variation comes from distant peers. Finally, the average exogenously predicted patent utilization rate of distant peers is used as an instrumental variable for the focal firm's patent utilization rate.

We expect that the relevance condition of this instrumental variable is satisfied, as an increase in distant peers' patent utilization should motivate the focal firm's use of its patents. The exclusion restriction is also likely satisfied because the exogenously determined patent utilization rate from the distant product market peers is unlikely to directly and positively affect

[29] Hoberg and Phillips (2016) develop the text-based industry classification (TNIC) based on the product description sections of firms' 10-K filings. TNIC2 is analogous to two-digit SIC industries, while TNIC3 corresponds to three-digit SIC industries.

the focal firm's performance unless through the variation in the focal firm's patent utilization rate.[30] The reported results in Table A5 suggest that it is indeed the case. The first-stage results in column 1 show that the exogenously determined rival patent utilization rate is significantly and positively correlated with the focal firm's patent utilization rate. The second-stage results in Columns 2-12 further indicate that the instrumented corporate patent utilization rate continues to be positively and significantly related to future firm performance.

In summary, we perform two sets of 2SLS-IV regression analyses by using either the R&D tax credits or the exogenously determined distant peers' patent utilization rate as instrumental variables for the focal firm's patent utilization rate. We continue to obtain qualitatively similar estimates that patent utilization rate possesses positive implications for a firm's future performance. These results partially alleviate the endogeneity concerns and help establish causal interpretations.

6.2 Robustness Checks

6.2.1 Controlling for Firms' Patent Outputs

One might be concerned that our patent utilization measure, which is partly based on firms' patent application portfolio (the denominator), reflects the patenting outputs of these firms and thus, our findings are driven by innovation outputs and not related to the degree of patent utilization.

To address this concern, we again employ Equation 7 to compare a firm's patent utilization rate to its citation-weighted ($\#CW\ Patents/AT$) and economic values ($Patent\ Values/AT$) of patents. If our findings are primarily driven by the corporate patenting outputs, we would expect to observe insignificant associations between patent utilization rate and future firm performance. However, if a firm's patent utilization captures additional information beyond the traditional patenting output measures, we should continue to observe results similar to those shown in prior sections. The results are reported in Panel A of Table 11.

[30] By definition, the distant peer firms are not in the same product market as the focal firm.

[Please insert Table 11 about here]

Even after controlling for a firm's patent outputs, we continue to find a positive and statistically significant relationship between patent utilization rate and firm performance in the subsequent year. The economic magnitudes of *Pat.Utilization Rate* remain close to the baseline results. Consistent with Kogan et al. (2017) and Ma (2021), we also find that patent economic value is strongly and positively associated with future firm performance, whereas the relationship between citation-weighted patent counts and firm performance appears fragile and ambiguous.[31] These findings imply that our measure of corporate patent utilization provides information beyond traditional innovation output metrics. While a firm's patent portfolio reflects the stock of its innovation outputs, the patent utilization rate captures the extent to which the firm leverages these patents in new product development.

6.2.2 Controlling for Firm Fixed Effect

Another concern is that the positive implications of patent utilization for a firm's future performance might be driven by time-invariant firm characteristics. To address this concern, we replace the industry fixed effect in the baseline regressions with firm fixed effect. In this case, we are exploring within-firm variation between patent utilization rate and future firm performance. As such, any time-invariant firm characteristics will be absorbed.

The results in Panel B of Table 11 show that our findings are largely remained. We continue to observe qualitatively similar evidence that corporate patent utilization has positive and significant effects on future firms' performance.

6.2.3 Controlling for Product Similarity Score

One might also be concerned that our patent utilization ratio overlaps with the product similarity score constructed by Hoberg and Phillips (2016), which is based on the business description section of firms' 10-K filings. However, while their measure primarily captures product

[31] The results suggest that citation-weighted patent stocks may even negatively relate to future product market outcomes. This could be because citation-weighted measures capture the scientific value of inventions, which may generate positive externalities for society without necessarily benefiting the firm's product market performance.

market competition, our measure differs both conceptually and methodologically. Specifically, we aim to quantify a firm's patent portfolio utilization rate rather than its competitive positioning. To achieve this, we compare the textual content of the firm's new products with its patent portfolio from the past five years, generating a *within-firm* patent-product similarity score. A patent is considered utilized in a product if its similarity score falls above the 80th percentile. In contrast, Hoberg and Phillips (2016)'s measure emphasizes *firm-to-firm* product market similarity.

Nevertheless, to formally address this concern, we control for the product similarity score in addition to our baseline control variables. The regression results, presented in Table A6, remain qualitatively consistent, showing that patent utilization rate remains positively and significantly associated with future firm performance. Perhaps not surprisingly, product similarity score is negatively related to future firm performance.

6.2.4 Alternative Measures of Patent Utilization

In the baseline regressions, we only consider the past 5 years' patent portfolio of a firm and classify a patent as being utilized in a product if the textual similarity score is above 80th percentile across our sample's patent-product pair scores. In this subsection, we further construct alternative measures of patent utilization rate using either 10-year patent portfolio of a firm or alternative patent-product pair score cutoffs (70th or 90th). In addition, we further employ a 3-year moving average approach to generate an alternative measure of firm-level patent utilization rate, accounting for the possibility that a firm's patent portfolio utilization in new product development may be more stable over the medium term. Similarly, instead of focusing on a 1-year (i.e., current year) patent-product incorporation rate over the past five-year patent portfolio window, we extend the patent usage window to three years. That is, we count the number of unique patents that have been incorporated into new products launched over the past three years (including the current year) from year *t-2* to year *t*, scaled by the total number of unique patents applied for and later granted by the firm from year *t-6* to year *t*.

The robustness results, reported in Table A7, show that the findings of the study remain qualitatively similar across different alternative measures of patent utilization rate. In product-level analyses (Panels A and B), we continue to find that new products embedded with more patents will have significantly higher product announcement returns (CARs) and quality (breakthrough index) across different measures of patent utilization. In firm-year level analyses (Panel C), the alternative measures of corporate patent utilization rate continue to positively and significantly affect firms' new product development, product market performance, profit improvement, and valuations.

6.2.5 Restrict Firm-year Observations with at least One New Product Launch

Finally, one might be concerned that the findings in this study are dominated by the extensive margin; that is, the positive implications of patent utilization on firm performance heavily depend on whether firms launch new products in a given year. Although we have controlled for a firm's new product intensity in the baseline regressions, we further alleviate this issue by restricting the firm-year observations to those with at least one new product launch. In this case, we are exploring the effects of patent utilization on firms' future performance, conditioned on firms launching at least one product in a given year.

The results reported in Table A8 show that a firm's patent utilization rate continues to positively and significantly predict its future new product development, product market sales, profitability changes, and valuations. Therefore, the findings in this study are robust to intensive margin analyses.

7 Conclusion

This study introduces a machine learning approach to capture the extent to which a firm's patent portfolio contributes to new product development, based on the textual analysis of firm patent and product texts. Leveraging the pre-trained *FastText* model, we generate a product- and firm-level patent utilization rate for 3,102 firms from 2002 to 2022. We exploit this novel measure to deepen our understanding of firms' patent utilization rate in new product development, and its implications for firm performance.

Our findings show that, at the product level, new products incorporating more patents tend to be of higher quality, as measured by cumulative abnormal stock returns in the three-day window surrounding the announcement, and by the text-based breakthrough index, which captures a product's novelty and impact. At the firm level, we find that corporate patent utilization is positively and significantly associated with new product development, product market performance, profitability improvements, and firm valuation in the following year. Heterogeneity tests further reveal that these positive effects are primarily driven by the utilization of high-value patents. Moreover, firms operating in competitive product markets benefit more from effective patent utilization, highlighting the role of market dynamics in shaping the value derived from intellectual assets.

To address potential endogeneity concerns, we employ federal and state-level R&D tax credit variables as instruments for patent utilization rates. Our instrumental variable analysis confirms the robustness of the findings, reinforcing the potential causal relationship between patent utilization and firm performance.

Our study speaks to the emerging "patent portfolio races," where firms strike for patents not to commercialize the associated technologies through new products, but solely to block competitors from pursuing future innovations. Our findings underscore the detrimental consequences of patent underutilization for firms, and highlight the importance of integrating patents into the product development pipeline. This paper offers practical insights for corporate managers to reconsider the use of their patent portfolios, and for policymakers to design initiatives that promote the effective adoption of intellectual assets.

References

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

Akcigit, U. and Kerr, W. R. (2018). Growth through heterogeneous innovations. *Journal of Political Economy*, 126(4):1374–1443.

Amesse, F., Desranleau, C., Etemad, H., Fortier, Y., and Seguin-Dulude, L. (1991). The individual inventor and the role of entrepreneurship: A survey of the canadian evidence. *Research policy*, 20(1):13–27.

Argente, D., Baslandze, S., Hanley, D., and Moreira, S. (2020). Patents to products: Product innovation and firm dynamics.

Arora, A., Belenzon, S., and Sheer, L. (2021). Knowledge spillovers and corporate investment in scientific research. *American Economic Review*, 111(3):871–898.

Arts, S., Cassiman, B., and Hou, J. (2023). Position and differentiation of firms in technology space. *Management Science*, 69(12):7253–7265.

Bena, J., Ortiz-Molina, H., and Simintzi, E. (2022). Shielding firm value: Employment protection and process innovation. *Journal of Financial Economics*, 146(2):637–664.

Bena, J. and Simintzi, E. (2022). Machines could not compete with chinese labor: Evidence from us firms' innovation. *Available at SSRN 2613248*.

Billett, M. T., Garfinkel, J. A., and Yu, M. (2017). The effect of asymmetric information on product market outcomes. *Journal of Financial Economics*, 123(2):357–376.

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

Bloom, N., Schankerman, M., and Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4):1347–1393.

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

Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the association for computational linguistics*, 5:135–146.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of econometrics*, 150(1):41–55.

Campello, M. (2006). Debt financing: Does it boost or hurt firm performance in product markets? *Journal of Financial Economics*, 82(1):135–172.

Cao, S. S., Ma, G., Tucker, J. W., and Wan, C. (2018). Technological peer pressure and product disclosure. *The Accounting Review*, 93(6):95–126.

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, W. and Srinivasan, S. (2023). Going digital: Implications for firm value and performance. *Review of Accounting Studies*, pages 1–47.

Choi, J. P. and Gerlach, H. (2017). A theory of patent portfolios. *American Economic Journal: Microeconomics*, 9(1):315–351.

Cohen, L., Gurun, U. G., Moon, S. K., and Suh, P. (2023). Patent hunters. *Available at SSRN 4635609*.

Cohen-Cole, E., Kirilenko, A., and Patacchini, E. (2014). Trading networks and liquidity provision. *Journal of Financial Economics*, 113(2):235–251.

Cohn, J. B., Liu, Z., and Wardlaw, M. I. (2022). Count (and count-like) data in finance. *Journal of Financial Economics*, 146(2):529–551.

Cooper, M., Knott, A. M., and Yang, W. (2022). Rq innovative efficiency and firm value. *Journal of Financial and Quantitative Analysis*, 57(5):1649–1694.

Fadeev, E. (2023). Creative construction: Knowledge sharing and cooperation between firms.

Firth, J. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis*, pages 10–32.

Fresard, L. (2010). Financial strength and product market behavior: The real effects of corporate cash holdings. *The Journal of finance*, 65(3):1097–1122.

Garcia, D. and Norli, Ø. (2012). Geographic dispersion and stock returns. *Journal of Financial*

Economics, 106(3):547–565.

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3):535–574.

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

Hall, B. H. (1993). R&d tax policy during the 1980s: success or failure? *Tax policy and the economy*, 7:1–35.

Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2001). Market value and patent citations: A first look.

Hall, B. H. and Ziedonis, R. H. (2001). The patent paradox revisited: an empirical study of patenting in the us semiconductor industry, 1979-1995. *rand Journal of Economics*, pages 101–128.

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.

Hoberg, G. and Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124(5):1423–1465.

Hoberg, G. and Phillips, G. (2024). Scope, scale and concentration: The 21st century firm. *The Journal of Finance*, Forthcoming.

Hombert, J. and Matray, A. (2018). Can innovation help us manufacturing firms escape import competition from china? *The Journal of Finance*, 73(5):2003–2039.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics*, 108(3):577–598.

Kalyani, A. (2022). The creativity decline: Evidence from us patents. *Available at SSRN 4318158*.

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., Schmidt, L., and Seegmiller, B. (2022). Technology, vintage-specific human capital, and labor displacement: Evidence from linking patents with occupations. *Available at SSRN 3983906*.

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.

Lentz, R. and Mortensen, D. T. (2008). An empirical model of growth through product innovation. *Econometrica*, 76(6):1317–1373.

Lev, B. and Sougiannis, T. (1996). The capitalization, amortization, and value-relevance of r&d. *Journal of accounting and economics*, 21(1):107–138.

Li, K., Mai, F., Shen, R., and Yan, X. (2021). Measuring corporate culture using machine learning. *The Review of Financial Studies*, 34(7):3265–3315.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of finance*, 66(1):35–65.

Ma, S. (2021). Technological obsolescence.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., and Joulin, A. (2017). Advances in pre-training distributed word representations. *arXiv preprint arXiv:1712.09405*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.

Moretti, E. (2021). The effect of high-tech clusters on the productivity of top inventors. *American Economic Review*, 111(10):3328–3375.

Mukherjee, A., Singh, M., and Žaldokas, A. (2017). Do corporate taxes hinder innovation? *Journal of Financial Economics*, 124(1):195–221.

Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language* *processing (EMNLP)*, pages 1532–1543.

Seegmiller, B., Papanikolaou, D., and Schmidt, L. D. (2023). Measuring document similarity with weighted averages of word embeddings. *Explorations in Economic History*, 87:101494.

Singh, J. and Marx, M. (2013). Geographic constraints on knowledge spillovers: Political borders vs. spatial proximity. *Management Science*, 59(9):2056–2078.

Sougiannis, T. (1994). The accounting based valuation of corporate r&d. *Accounting review*, pages 44–68.

Thompson, P. and Fox-Kean, M. (2005). Patent citations and the geography of knowledge spillovers: A reassessment. *American Economic Review*, 95(1):450–460.

Webster, E. and Jensen, P. H. (2011). Do patents matter for commercialization? *The Journal of Law and Economics*, 54(2):431–453.

Wilson, D. J. (2009). Beggar thy neighbor? the in-state, out-of-state, and aggregate effects of r&d tax credits. *The Review of Economics and Statistics*, 91(2):431–436.

Figure 1. Patent-Product Pair Matching Process Illustration

This figure illustrates the patent-product pair matching process. The patent filing text data is obtained from PatentsView, while the new product launch text data is obtained from Capital IQ Key Development Database. The similarity score of patent-product pair is generated via the *Word2vec* model which compares the text similarities between patent filings and new product launch text description.

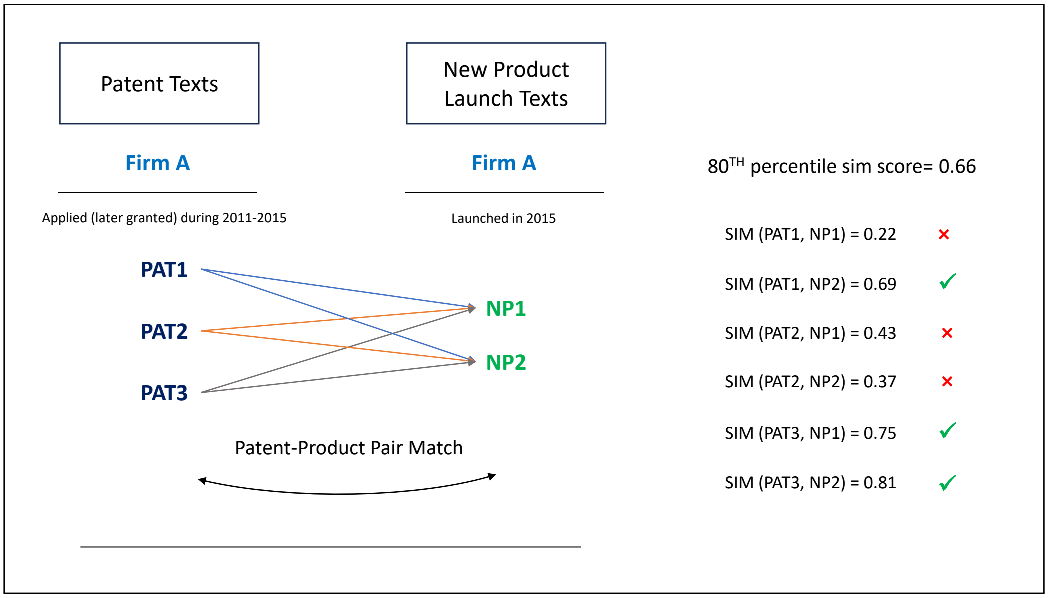


Figure 2. Annual Variation of Patent Utilization Rate

This figure illustrates the annual variation of corporate patent utilization rate by year from 2002 to 2022.

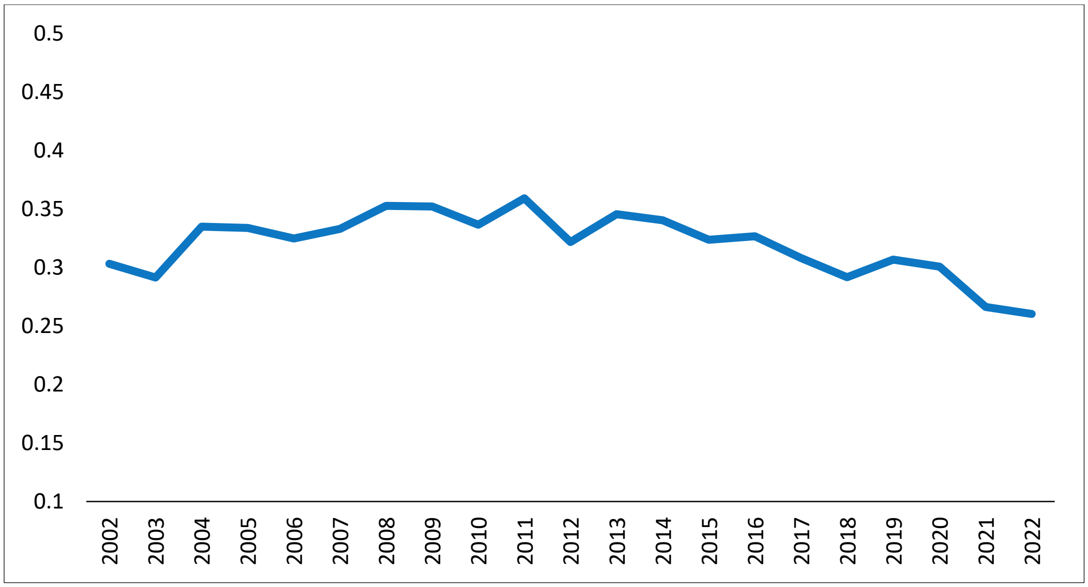


Figure 3. Industry Variation of Patent Utilization Rate

The figure illustrates the top 10 industries (2-digit SIC) with the highest rates in patent utilization. The y-axis denotes the 2-digit SIC and the related industry classification, and the x-axis reports the rate of patent utilization.

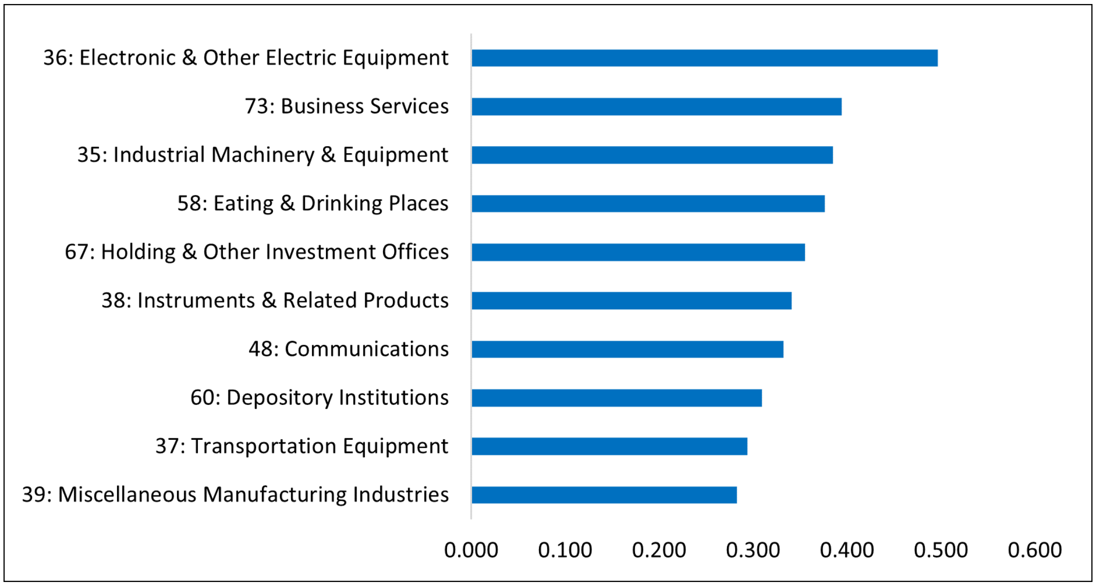


Table 1. Summary Statistics

This table reports the summary statistics for the new product announcement event sample (Panel A) and firm-year regression sample (Panel B) The sample period starts from 2002 to 2022. We report the number of observations, standard deviation, mean, 25th percentile, median, and 75th percentile for each of the variables used in the study. All continuous variables are winsorized at the 1st and 99th percentiles. Table A1 in Appendix A provides detailed variable definitions.

Panel A. New Product Announcement Event Sample

Variables	Obs.	Std.	Mean	P25	Median	P75
Dependent Variables						
CAR (-1,1)	94,239	4.156	0.190	-1.699	0.030	1.872
Breakthrough Index	105,196	0.797	0.000	-0.211	-0.106	0.077
1 (Breakthrough Product)	105,196	0.155	0.025	0.000	0.000	0.000
Independent Variables						
Log(1+#Patents Utilized$^{\text{Sum}}$)	94,239	2.367	2.795	0.693	2.565	4.635
Log(1+#Patents Utilized$^{\text{Average}}$)	94,239	2.305	2.732	0.405	2.565	4.543
Log(1+#Patents Utilized)	105,196	2.333	2.819	0.693	2.639	4.663
Log(Product Text Length)	94,239	0.587	5.025	4.635	5.011	5.425
Log(1+#New Products Launched)	94,239	1.263	1.777	0.693	1.609	2.708
Firm Size	94,239	2.324	6.206	4.567	6.207	8.093
Log(Firm Age)	94,239	0.936	2.769	2.197	2.890	3.497
Leverage	94,239	0.161	0.163	0.002	0.133	0.267
R&D	94,239	0.023	0.023	0.007	0.019	0.032
ROA	94,239	0.038	0.026	0.014	0.029	0.044
Cash	94,239	0.192	0.266	0.112	0.225	0.383
Log(Tobin's Q)	94,239	0.530	0.672	0.273	0.614	1.002
Sales Growth	94,239	0.185	0.014	-0.042	0.022	0.082
Past Stock Return	94,239	0.273	0.034	-0.100	0.019	0.137

Panel B. Firm-year Sample

Variables	Obs.	Std.	Mean	P25	Median	P75
Dependent Variables						
#New Products	21,453	3.181	1.258	0.000	0.000	1.000
#Breakthrough Products	21,453	0.298	0.035	0.000	0.000	0.000
Sum CARs	21,453	14.236	6.242	0.000	0.106	6.621
Sales Growth	21,403	0.318	0.072	-0.030	0.068	0.175
MSG(FF49)	21,403	0.303	0.002	-0.089	0.000	0.092
MSG(SIC4)	21,403	0.295	0.001	-0.081	0.000	0.081
△GPM	21,450	0.170	-0.003	-0.049	0.001	0.046
△ROA	21,441	0.134	0.000	-0.033	0.000	0.031
△OCF	21,438	0.144	0.000	-0.043	-0.002	0.037
Log(Tobin's Q)	21,453	0.563	0.661	0.250	0.573	0.998
Log(MTB)	21,453	0.834	1.017	0.454	0.955	1.485
Independent Variables						
Pat. Utilization Rate	21,453	0.398	0.351	0.000	0.071	0.750
#New Products/Sales	21,453	2.614	0.070	0.000	0.000	0.002
Firm Size	21,453	2.579	6.344	4.572	6.342	8.166
Log(Firm Age)	21,453	0.852	2.744	2.197	2.833	3.367
Leverage	21,453	0.167	0.167	0.002	0.135	0.281
R&D	21,453	0.123	0.093	0.016	0.055	0.120
ROA	21,453	0.246	0.054	0.017	0.108	0.171
Cash	21,453	0.236	0.277	0.084	0.207	0.418
Past Stock Return	21,453	0.654	0.156	-0.196	0.070	0.349

Table 2. Number of Patents Utilized in New Products and New Product Announcement Return

This table reports the regression results that investigate the association between the number of unique patents utilized in a new product and the product's announcement return. The dependent variable *CAR (-1, 1)* is the cumulative abnormal stock returns during a three-day event window of (-1, 1) following the new product announcement event. Because a firm may launch multiple products on one event date, we then generate two independent variables to account for the situation of product bundle launching: *Log(1+#Patents Utilized^{Sum})*, which is the natural logarithm of one plus the sum of the number of patents that are utilized in new product(s) for an event-date, and *Log(1+#Patents Utilized^{Average})*, which is the natural logarithm of one plus the average number of patents that are utilized in new product(s) for an event-date. We define a patent is utilized in a new product if the patent-product pair text similarity score is above 80th percentile. All regression specifications include product and firm level control variables. Columns 1 and 5 do not include any fixed effects. Columns 2 and 6 include industry fixed effects. Columns 3 and 7 include both industry fixed effects and event-date fixed effects. Columns 4 and 8 include industry-by-event-date fixed effects. Table A1 in Appendix A provides detailed variable definitions. Robust standard errors clustered at the firm and event-date level are provided in parentheses. All financial variables are winsorized at the 1st and 99th percentiles. ***, **, and * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	CAR (-1,1)				CAR (-1,1)			
Log (1+#Patents Used^{Sum})	**0.016****	**0.026*****	**0.024*****	**0.020****				
	(0.008)	**(0.008)**	**(0.008)**	**(0.010)**				
Log (1+#Patents Used^{Average})					**0.014***	**0.024*****	**0.023*****	**0.020****
					(0.008)	**(0.009)**	**(0.008)**	**(0.010)**
Log (Product Text Length)	-0.052*	-0.043	-0.041	-0.037	-0.052*	-0.043	-0.041	-0.037
	(0.029)	(0.030)	(0.028)	(0.035)	(0.029)	(0.030)	(0.028)	(0.035)
Log(1+# New Products Launched)	-0.047***	-0.035*	-0.013	-0.045**	-0.046***	-0.034*	-0.012	-0.044*
	(0.017)	(0.018)	(0.019)	(0.022)	(0.017)	(0.018)	(0.019)	(0.022)
Firm Size $_{q-1}$	-0.062***	-0.074***	-0.081***	-0.072***	-0.061***	-0.072***	-0.080***	-0.071***
	(0.012)	(0.014)	(0.014)	(0.016)	(0.012)	(0.014)	(0.014)	(0.016)
Log (Firm Age) $_{q-1}$	-0.025	-0.027	-0.028	-0.028	-0.025	-0.027	-0.028	-0.028
	(0.023)	(0.024)	(0.023)	(0.026)	(0.023)	(0.024)	(0.023)	(0.026)
Leverage $_{q-1}$	0.282**	0.295**	0.286**	0.293*	0.281**	0.295**	0.286**	0.294*
	(0.121)	(0.128)	(0.124)	(0.155)	(0.121)	(0.128)	(0.124)	(0.155)
R&D $_{q-1}$	4.410***	4.693***	4.321***	3.863***	4.442***	4.714***	4.334***	3.870***
	(1.135)	(1.164)	(1.140)	(1.236)	(1.136)	(1.165)	(1.141)	(1.237)
ROA $_{q-1}$	0.631	0.976	0.717	0.920	0.632	0.970	0.709	0.913
	(0.771)	(0.778)	(0.774)	(0.898)	(0.771)	(0.778)	(0.774)	(0.898)
Cash $_{q-1}$	0.171	0.220*	0.139	0.084	0.173	0.222*	0.140	0.085
	(0.117)	(0.120)	(0.116)	(0.133)	(0.117)	(0.120)	(0.116)	(0.133)

Tobin's Q $_{q-1}$	-0.143***	-0.176***	-0.110**	-0.134**	-0.143***	-0.175***	-0.109**	-0.133**
	(0.049)	(0.052)	(0.051)	(0.061)	(0.049)	(0.052)	(0.051)	(0.061)
Sales Growth $_{q-1}$	0.296***	0.300***	0.382***	0.465***	0.295***	0.300***	0.382***	0.465***
	(0.112)	(0.113)	(0.115)	(0.137)	(0.112)	(0.113)	(0.115)	(0.138)
Stock Return $_{q-1}$	0.099	0.110	0.097	0.005	0.099	0.110	0.097	0.004
	(0.095)	(0.095)	(0.090)	(0.103)	(0.095)	(0.095)	(0.090)	(0.103)
Industry FE		✓	✓			✓	✓	
Event-Date FE			✓				✓	
Industry-Event-Date FE				✓				✓
Obs.	94,239	94,236	94,108	74,845	94,239	94,236	94,108	74,845
Adj. R2	0.003	0.003	0.055	0.087	0.003	0.003	0.055	0.087

Table 3. Number of Patents Utilized in New Products and Product Breakthrough

This table reports the regression results that investigate the association between the number of unique patents utilized in a new product and the product's breakthrough index and the likelihood of being a breakthrough product. The dependent variable *Breakthrough Index* is a text-based measurement that captures product significance. Following Kelly et al. (2021), the breakthrough index considers a product's novelty and impact, which is constructed as: $BreakthroughIndex^5{}_{\mathrm{j}} = \frac{FS^5{}_{\mathrm{j}}}{BS^5{}_{\mathrm{j}}}$, where $BS^5{}_j$ measures the backward similarity (novelty dimension) and $FS^5{}_j$ measures the forward similarity (impact dimension). Specifically, $BS^5{}_{\mathrm{j}} = \sum_{i \in \beta^5_{j,m}} \rho_{j,i}$, where $\rho_{j,i}$ is the pairwise similarity between product j and i, and $\beta^5_{j,m}$ denotes the set of previous products that are launched in the 5 years before product j's offering and that are in the same product market m (based on parent firms' TNIC3 classification (Hoberg and Phillips, 2016) as product j). Similarly, the forward similarity $FS^5{}_{\mathrm{j}} = \sum_{i \in \alpha^5_{j,m}} \rho_{j,i}$. $\rho_{j,i}$ is the pairwise similarity between product j and i, and $\alpha^5_{j,m}$ denotes the set of future products that are launched in the 5 years after product j's offering and that are in the same product market m (based on parent firms' TNIC3 classification) as product j. Thus, a product with low backward similarity with the prior products (which is novel) but high forward similarity with the subsequent products (which is impactful) has a high breakthrough index. The other dependent variable, *1 (Breakthrough Product)*, is an indicator that equals 1 if the product's breakthrough index is above the 95th percentile, otherwise equals 0. The independent variable *Log(1+#Patents Utilized)* is the natural logarithm of one plus the number of patents that are utilized in the new product. We define a patent is utilized in a new product if the patent-product pair text similarity score is above 80th percentile. All regression specifications include product and firm level control variables. Columns 1 and 5 do not include any fixed effects. Columns 2 and 6 include industry fixed effects. Columns 3 and 7 include both industry fixed effects and event-date fixed effects. Columns 4 and 8 include industry-by-event-date fixed effects. Table A1 in Appendix A provides detailed variable definitions. Robust standard errors clustered at the firm and event-date level are provided in parentheses. All financial variables are winsorized at the 1st and 99th percentiles. ***, **, and * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Breakthrough Index				1 (Breakthrough Product)			
Log (1+#Patents Utilised)	**0.025*** **	**0.024*** **	**0.021*** **	**0.020*** **	**0.003*** **	**0.002*** **	**0.001** **	**0.001** **
	(0.007)	**(0.007)**	**(0.006)**	**(0.006)**	**(0.001)**	**(0.001)**	**(0.001)**	**(0.001)**
Log (Product Text Length)	-0.065***	-0.054**	-0.003	-0.005	-0.021***	-0.019***	-0.002	-0.004***
	(0.024)	(0.022)	(0.014)	(0.009)	(0.003)	(0.003)	(0.002)	(0.001)
Log(1+# New Products Launched)	-0.008	0.011	0.011	0.018	-0.005***	-0.002	-0.004**	-0.004***
	(0.010)	(0.011)	(0.016)	(0.016)	(0.001)	(0.001)	(0.002)	(0.002)
Firm Size $_{\text{q-1}}$	-0.028***	-0.033***	-0.029***	-0.036***	-0.003***	-0.004***	-0.002	-0.001
	(0.009)	(0.011)	(0.011)	(0.012)	(0.001)	(0.001)	(0.001)	(0.001)
Log (Firm Age) $_{\text{q-1}}$	0.041**	0.026	0.032	0.028	0.004*	0.001	0.003	0.003
	(0.021)	(0.020)	(0.021)	(0.022)	(0.002)	(0.002)	(0.002)	(0.002)
Leverage $_{\text{q-1}}$	0.115	0.068	0.112	0.179*	0.011	0.001	0.006	0.004
	(0.097)	(0.096)	(0.090)	(0.096)	(0.012)	(0.011)	(0.010)	(0.010)
R&D $_{\text{q-1}}$	-1.501**	-1.064	-1.074*	-1.168*	-0.195**	-0.184**	-0.178**	-0.144

	(0.653)	(0.658)	(0.632)	(0.673)	(0.089)	(0.084)	(0.083)	(0.089)
ROA $_{q\text{-}1}$	-0.201	0.141	0.058	-0.008	0.017	0.033	0.009	0.002
	(0.411)	(0.457)	(0.416)	(0.436)	(0.043)	(0.045)	(0.043)	(0.035)
Cash $_{q\text{-}1}$	-0.189**	-0.118	-0.133*	-0.160*	-0.038***	-0.030***	-0.035***	-0.023***
	(0.088)	(0.085)	(0.078)	(0.085)	(0.012)	(0.010)	(0.009)	(0.008)
Tobin's Q $_{q\text{-}1}$	0.017	0.000	0.011	0.028	-0.003	-0.006*	-0.000	-0.002
	(0.036)	(0.029)	(0.031)	(0.035)	(0.004)	(0.003)	(0.004)	(0.004)
Sales Growth $_{q\text{-}1}$	-0.004	-0.004	-0.014	-0.032	-0.005	-0.004	-0.005	-0.007
	(0.028)	(0.026)	(0.026)	(0.028)	(0.005)	(0.005)	(0.005)	(0.005)
Stock Return $_{q\text{-}1}$	0.021	0.029	-0.007	-0.040	0.016***	0.017***	0.001	-0.002
	(0.024)	(0.022)	(0.021)	(0.026)	(0.004)	(0.004)	(0.004)	(0.004)
Industry FE		✓	✓			✓	✓	
Event-Date FE			✓				✓	
Industry-Event-Date FE				✓				✓
Obs.	105,196	105,193	105,067	86,661	105,196	105,193	105,067	86,661
Adj. R2	0.012	0.041	0.060	0.146	0.014	0.030	0.084	0.201

Table 4. Corporate Patent Utilization Rate and New Product Development

This table reports the regression results that investigate the association between a firm's patent utilization rate and one-year-ahead new product development. The dependent variable *#New Products* is measured as the number of new products (with three-day CARs above 80th percentile) a firm launches in a year. *Sum CARs* is measured as the sum of all positive three-day cumulative abnormal stock returns of the new products that a firm launches in a year. *#Breakthrough Products* is measured as the number of breakthrough products (new products with breakthrough indexes above 95th percentile) a firm launches in a year. The independent variable *Pat.Utilization Rate* is measured as the number of patents that are applied (and later granted) in the past five years by a firm and that are utilized in the new products launched by the same firm in the current year, further scaled by the total number of patents applied (and later granted) in the past five years for that firm. We regard a patent as utilized in a product if the patent-product pair similarity score is above 80th percentile. All specifications include firm controls. Columns 1, 3, and 5 (2, 4, and 6) include industry fixed effects and year fixed effects (industry-by-year fixed effects). The results on columns 1-2 and 5-6 are estimated with Poisson regressions, while the results on columns 3-4 are estimated with OLS regressions. Table A1 in Appendix A provides detailed variable definitions. Robust standard errors clustered at the firm level are provided in parentheses. All financial variables are winsorized at the 1st and 99th percentiles. ***, **, and * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	#New Products $_{t+1}$		Sum CARs $_{t+1}$		#Breakthrough Products $_{t+1}$	
Pat. Utilization Rate	**1.697*****	**1.698*****	**9.107*****	**9.149*****	**1.525*****	**1.485*****
	(0.071)	**(0.072)**	**(0.471)**	**(0.478)**	**(0.148)**	**(0.145)**
#New Products/Sales	0.012***	0.013***	0.047**	0.049***	0.003	0.005
	(0.003)	(0.003)	(0.019)	(0.019)	(0.006)	(0.006)
Firm Size	0.369***	0.373***	1.497***	1.531***	0.154***	0.156***
	(0.021)	(0.021)	(0.203)	(0.208)	(0.032)	(0.032)
Log(Firm Age)	-0.022	-0.027	-0.076	-0.068	0.156**	0.167**
	(0.046)	(0.046)	(0.311)	(0.320)	(0.077)	(0.077)
Leverage	-0.505***	-0.505***	-2.372**	-2.404**	-0.038	-0.033
	(0.178)	(0.177)	(1.098)	(1.126)	(0.474)	(0.471)
R&D	2.379***	2.441***	13.190***	14.191***	0.310	0.334
	(0.238)	(0.247)	(1.716)	(1.778)	(0.649)	(0.667)
ROA	-0.377**	-0.385***	-2.175***	-2.109**	-0.248	-0.196
	(0.147)	(0.147)	(0.826)	(0.832)	(0.370)	(0.368)
Cash	0.865***	0.866***	4.560***	4.226***	-0.154	-0.089
	(0.127)	(0.130)	(1.009)	(1.019)	(0.387)	(0.388)
Log(Tobin's Q)	-0.020	-0.027	0.180	0.278	0.042	0.063
	(0.048)	(0.049)	(0.311)	(0.321)	(0.127)	(0.127)
Sales Growth	0.180***	0.179***	0.031	-0.026	0.020	0.065
	(0.052)	(0.054)	(0.177)	(0.183)	(0.191)	(0.178)
Past Stock Return	-0.044*	-0.038*	0.073	0.102	-0.116	-0.018
	(0.023)	(0.022)	(0.140)	(0.144)	(0.117)	(0.108)
Model	Poisson	Poisson	OLS	OLS	Poisson	Poisson
Industry FE	✓		✓		✓	
Year FE	✓		✓		✓	
Industry-Year FE		✓		✓		✓
Obs.	21,453	21,453	21,453	21,453	7,881	7,204
Pseudo/Adj. R2	0.365	0.375	0.199	0.196	0.219	0.245

Table 5. Corporate Patent Utilization Rate and Product Market Performance

This table reports the regression results that investigate the association between a firm's patent utilization and one-year-ahead product market performance. In Panel A, we report the relationship between a firm's patent utilization rate and sales growth. In Panel B and C, we shed light on the association between a firm's patent utilization rate and market share growth. The dependent variable *Sales Growth* is measured as the natural logarithm of total sales for a firm in year *t+1* minus the natural logarithm of total sales for that firm in year *t*. *MSG(FF49)* is measured as the sales growth of a firm in year *t+1* minus the industry (Fama-French 49 industries) median sales growth in the same year. *MSG(SIC4)* is measured as the sales growth of a firm in year *t+1* minus the industry (4-digit SIC) median sales growth in the same year. The independent variable *Pat.Utilization Rate* is measured as the number of patents that are applied (and later granted) in the past five years by a firm and that are utilized in the new products launched by the same firm at the current year, further scaled by the total number of patents applied (and later granted) in the past five years for that firm. We regard a patent as utilized in a product if the patent-product pair similarity score is above 80th percentile. All specifications include firm controls. Columns 1, 3 and 5 include industry fixed effects and year fixed effects. Columns 2, 4 and 6 include industry-by-year fixed effects. Table A1 in Appendix A provides detailed variable definitions. Robust standard errors clustered at the firm level are provided in parentheses. All financial variables are winsorized at the 1st and 99th percentiles. ***, **, and * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A. Patent Utilization Rate and Firm Future Sales Growth

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Sales Growth $_{t+1}$		Sales Growth $_{t+2}$		Sales Growth $_{t+3}$	
Pat. Utilization Rate	**0.013****	**0.014****	**0.018****	**0.018****	**0.021***	**0.021***
	(0.006)	**(0.006)**	**(0.009)**	**(0.009)**	**(0.012)**	**(0.012)**
#New Products/Sales	0.005***	0.005***	0.016	0.019	0.024	0.027
	(0.001)	(0.001)	(0.021)	(0.022)	(0.020)	(0.022)
Firm Size	-0.005***	-0.005***	-0.001	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Log(Firm Age)	-0.024***	-0.023***	-0.039***	-0.040***	-0.055***	-0.056***
	(0.003)	(0.003)	(0.005)	(0.005)	(0.008)	(0.008)
Leverage	0.069***	0.067***	0.078***	0.073***	0.108***	0.101***
	(0.015)	(0.016)	(0.024)	(0.025)	(0.036)	(0.037)
R&D	-0.330***	-0.324***	-0.656***	-0.645***	-0.747***	-0.736***
	(0.053)	(0.053)	(0.067)	(0.067)	(0.097)	(0.097)
ROA	-0.192***	-0.192***	-0.213***	-0.208***	-0.189***	-0.183***
	(0.032)	(0.032)	(0.038)	(0.038)	(0.054)	(0.054)
Cash	-0.055***	-0.056***	-0.011	-0.013	0.013	0.007
	(0.019)	(0.019)	(0.027)	(0.027)	(0.038)	(0.039)
Log(Tobin's Q)	0.117***	0.117***	0.189***	0.190***	0.249***	0.251***
	(0.007)	(0.007)	(0.010)	(0.010)	(0.014)	(0.014)
Past Stock Return	0.038***	0.038***	0.039***	0.038***	0.044***	0.042***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)
Model	OLS	OLS	OLS	OLS	OLS	OLS
Industry FE	✓		✓		✓	
Year FE	✓		✓		✓	
Industry-Year FE		✓		✓		✓
Obs.	21,403	21,403	19,589	19,585	17,780	17,764
Adj. R2	0.100	0.103	0.144	0.156	0.154	0.162

Panel B. Patent Utilization Rate and Firm Future Market Share Growth (FF49)

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	MSG(FF49) $_{t+1}$		MSG(FF49) $_{t+2}$		MSG(FF49) $_{t+3}$	
Pat. Utilization Rate	**0.014****	**0.014****	**0.019****	**0.019****	**0.023***	**0.023***
	(0.006)	**(0.006)**	**(0.009)**	**(0.009)**	**(0.012)**	**(0.012)**
#New Products/Sales	0.006***	0.006***	0.017	0.017	0.028	0.029
	(0.002)	(0.002)	(0.021)	(0.021)	(0.022)	(0.023)
Firm Size	-0.005***	-0.005***	-0.000	0.000	0.000	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)
Log(Firm Age)	-0.023***	-0.023***	-0.038***	-0.039***	-0.055***	-0.055***
	(0.003)	(0.003)	(0.005)	(0.005)	(0.007)	(0.008)
Leverage	0.070***	0.070***	0.070***	0.071***	0.097***	0.097***
	(0.015)	(0.015)	(0.024)	(0.025)	(0.036)	(0.037)
R&D	-0.326***	-0.326***	-0.664***	-0.663***	-0.754***	-0.758***
	(0.050)	(0.050)	(0.067)	(0.067)	(0.096)	(0.096)
ROA	-0.179***	-0.180***	-0.194***	-0.199***	-0.158***	-0.167***
	(0.031)	(0.031)	(0.038)	(0.038)	(0.054)	(0.055)
Cash	-0.059***	-0.061***	-0.023	-0.030	-0.003	-0.013
	(0.018)	(0.019)	(0.027)	(0.027)	(0.038)	(0.039)
Log(Tobin's Q)	0.111***	0.112***	0.177***	0.182***	0.232***	0.239***
	(0.006)	(0.006)	(0.010)	(0.010)	(0.014)	(0.014)
Past Stock Return	0.029***	0.030***	0.037***	0.038***	0.042***	0.044***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)
Model	OLS	OLS	OLS	OLS	OLS	OLS
Industry FE	✓		✓		✓	
Year FE	✓		✓		✓	
Industry-Year FE		✓		✓		✓
Obs.	21,403	21,403	19,589	19,585	17,780	17,764
Adj. R2	0.064	0.055	0.094	0.093	0.108	0.103

Panel C. Patent Utilization Rate and Firm Future Market Share Growth (SIC4)

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	MSG(SIC4) $_{t+1}$		MSG(SIC4) $_{t+2}$		MSG(SIC4) $_{t+3}$	
Pat. Utilization Rate	**0.013****	**0.013****	**0.019****	**0.018****	**0.020***	**0.021***
	(0.005)	**(0.006)**	**(0.008)**	**(0.008)**	**(0.012)**	**(0.012)**
#New Products/Sales	0.006***	0.006***	0.021	0.021	0.031	0.032
	(0.002)	(0.002)	(0.021)	(0.021)	(0.022)	(0.024)
Firm Size	-0.004**	-0.003**	0.001	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Log(Firm Age)	-0.021***	-0.021***	-0.033***	-0.034***	-0.047***	-0.047***
	(0.003)	(0.003)	(0.005)	(0.005)	(0.007)	(0.007)
Leverage	0.060***	0.060***	0.063**	0.063**	0.084**	0.082**
	(0.015)	(0.015)	(0.025)	(0.025)	(0.035)	(0.036)
R&D	-0.313***	-0.313***	-0.649***	-0.650***	-0.738***	-0.747***
	(0.049)	(0.049)	(0.066)	(0.066)	(0.095)	(0.096)
ROA	-0.177***	-0.179***	-0.190***	-0.195***	-0.160***	-0.172***
	(0.030)	(0.031)	(0.037)	(0.037)	(0.052)	(0.053)
Cash	-0.060***	-0.062***	-0.028	-0.036	-0.014	-0.024

	(0.018)	(0.018)	(0.027)	(0.027)	(0.038)	(0.038)
Log(Tobin's Q)	0.103***	0.104***	0.168***	0.174***	0.219***	0.227***
	(0.006)	(0.006)	(0.010)	(0.010)	(0.014)	(0.014)
Past Stock Return	0.023***	0.025***	0.032***	0.034***	0.039***	0.040***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)
Model	OLS	OLS	OLS	OLS	OLS	OLS
Industry FE	✓		✓		✓	
Year FE	✓		✓		✓	
Industry-Year FE		✓		✓		✓
Obs.	21,403	21,403	19,589	19,585	17,780	17,764
Adj. R2	0.056	0.043	0.085	0.080	0.097	0.090

Table 6. Corporate Patent Utilization Rate and Profit Improvements

This table reports the regression results that investigate the association between a firm's patent utilization rate and one-year-ahead firm profit improvements. The dependent variable $\triangle GPM$ is measured as the change in the gross profit margin of a firm between year t and $t{+}1$; gross profit margin is defined as a firm's sales minus cost of goods sold, further divided by the firm's book value of total assets at the beginning of the year. $\triangle ROA$ is measured as the change in the return on assets (ROA) of a firm between year t and $t{+}1$; ROA is defined as a firm's operating income before depreciation divided by the firm's book value of total assets at the beginning of the year. $\triangle OCF$ is measured as the change in the operating cash flow of a firm between year t and $t{+}1$; operating cash flow is defined as a firm's operating cash flow divided by the firm's book value of total assets at the beginning of the year. The independent variable *Pat. Utilization Rate* is measured as the number of patents that are applied (and later granted) in the past five years by a firm and that are utilized in the new products launched by the same firm in the current year, further scaled by the total number of patents applied (and later granted) in the past five years for that firm. We regard a patent as utilized in a product if the patent-product pair similarity score is above 80th percentile. All specifications include firm controls. Columns 1, 3 and 5 include industry fixed effects and year fixed effects. Columns 2, 4 and 6 include industry-by-year fixed effects. Table A1 in Appendix A provides detailed variable definitions. Robust standard errors clustered at the firm level are provided in parentheses. All financial variables are winsorized at the 1st and 99th percentiles. ***, **, and * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	$\triangle$GPM $_{t+1}$		$\triangle$ROA $_{t+1}$		$\triangle$OCF $_{t+1}$	
Pat. Utilization Rate	**0.006****	**0.006****	**0.008*****	**0.008*****	**0.007*****	**0.007*****
	(0.003)	**(0.003)**	**(0.002)**	**(0.002)**	**(0.002)**	**(0.002)**
#New Products/Sales	0.000	0.001	0.001	0.001*	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Firm Size	-0.003***	-0.003***	-0.004***	-0.004***	-0.005***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log(Firm Age)	0.000	0.001	0.000	0.000	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Leverage	0.011	0.010	0.037***	0.036***	0.041***	0.041***
	(0.008)	(0.008)	(0.006)	(0.006)	(0.006)	(0.007)
R&D	0.104***	0.105***	-0.072***	-0.070***	-0.023	-0.020
	(0.022)	(0.022)	(0.021)	(0.021)	(0.022)	(0.022)
Cash	-0.033***	-0.035***	0.017**	0.016**	-0.037***	-0.037***
	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)
Log(Tobin's Q)	0.009***	0.010***	0.001	0.001	0.008***	0.007***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Sales Growth	-0.074***	-0.074***	-0.021***	-0.020***	-0.007	-0.006
	(0.008)	(0.008)	(0.006)	(0.006)	(0.007)	(0.007)
Past Stock Return	-0.009***	-0.010***	0.004	0.004	-0.010***	-0.010***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Model	OLS	OLS	OLS	OLS	OLS	OLS
Industry FE	✓		✓		✓	
Year FE	✓		✓		✓	
Industry-Year FE		✓		✓		✓
Obs.	21,450	21,450	21,441	21,441	21,438	21,438
Adj. R2	0.056	0.054	0.019	0.015	0.009	0.002

Table 7. Corporate Patent Utilization Rate and Firm Values

This table reports the regression results that investigate the association between a firm's patent utilization rate and one-year-ahead firm values. The dependent variable *Log(Tobin's Q)* is measured as the natural logarithm of a firm's book value of assets minus book value of equity plus market value of equity further divided by the book value of total assets. *Log(MTB)* is measured as the natural logarithm of a firm's market value of assets divided by the book value of total assets. The independent variable *Pat.Utilization Rate* is measured as the number of patents that are applied (and later granted) in the past five years by a firm and that are utilized in the new products launched by the same firm at the current year, further scaled by the total number of patents applied (and later granted) in the past five years for that firm. We regard a patent as utilized in a product if the patent-product pair similarity score is above 80th percentile. All specifications include firm controls. Columns 1 and 3 include industry fixed effects and year fixed effects. Columns 2 and 4 include industry-by-year fixed effects. Table A1 in Appendix A provides detailed variable definitions. Robust standard errors clustered at the firm level are provided in parentheses. All financial variables are winsorized at the 1st and 99th percentiles. ***, **, and * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	(1)	(2)	(3)	(4)
VARIABLES	Log(Tobin's Q) $_{t+1}$		Log(MTB) $_{t+1}$	
Pat. Utilization Rate	**0.039****	**0.041****	**0.064*****	**0.066*****
	(0.016)	**(0.017)**	**(0.024)**	**(0.024)**
#New Products/Sales	-0.000	0.000	-0.001	-0.001
	(0.001)	(0.001)	(0.002)	(0.002)
Firm Size	0.013***	0.013**	0.040***	0.039***
	(0.005)	(0.005)	(0.007)	(0.008)
Log(Firm Age)	0.001	0.002	0.015	0.018
	(0.011)	(0.011)	(0.016)	(0.017)
Leverage	-0.008	-0.005	1.128***	1.132***
	(0.047)	(0.048)	(0.077)	(0.077)
R&D	1.156***	1.147***	1.751***	1.730***
	(0.084)	(0.086)	(0.116)	(0.118)
ROA	0.276***	0.279***	0.170**	0.164**
	(0.053)	(0.054)	(0.071)	(0.072)
Cash	0.574***	0.580***	0.696***	0.700***
	(0.044)	(0.045)	(0.061)	(0.062)
Sales Growth	0.103***	0.106***	0.124***	0.129***
	(0.014)	(0.014)	(0.018)	(0.019)
Past Stock Return	0.150***	0.147***	0.198***	0.194***
	(0.008)	(0.008)	(0.011)	(0.011)
Model	OLS	OLS	OLS	OLS
Industry FE	✓		✓	
Year FE	✓		✓	
Industry-Year FE		✓		✓
Obs.	21,453	21,453	21,453	21,453
Adj. R2	0.290	0.292	0.260	0.266

Table 8. High-Value versus Low-Value Patent Utilization Rate

This table compares the effects of high-value versus low-value patent utilization rate on future firms' new product development (panel A), product market performance (panel B), profit improvements (panel C), and firm values (panel D). We define a patent as high value if its economic value is above 80th percentile across all patents' values. Similarly, a patent is regarded as low value if the economic value is below 80th percentile. The independent variable *High-Value (Low-Value) Pat. Utilization Rate* is measured as the number of high-value (low-value) patents that are applied (and later granted) in the past five years by a firm and that are utilized in the new products launched by the same firm in the current year, further scaled by the total number of patents applied (and later granted) in the past five years for that firm. We regard a patent as utilized in a product if the patent-product pair similarity score is above 80th percentile. All specifications include firm controls, industry fixed effects, and year fixed effects. Table A1 in Appendix A provides detailed variable definitions. Robust standard errors clustered at the firm level are provided in parentheses. All financial variables are winsorized at the 1st and 99th percentiles. ***, **, and * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A. New Product Development

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	#New Products $_{t+1}$		Sum CARs $_{t+1}$		#Breakthrough Products $_{t+1}$	
Patent Values	High	Low	High	Low	High	Low
High-Value (Low-Value)	1.521***	1.304***	11.624***	6.711***	1.521***	1.042***
Pat. Utilization Rate	(0.076)	(0.078)	(0.759)	(0.546)	(0.166)	(0.224)
P Value of Diff.	0.025		0.000		0.109	
Model	Poisson	Poisson	OLS	OLS	Poisson	Poisson
Firm Controls	✓	✓	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Obs.	21,453	21,453	21,453	21,453	7,881	7,881
Pseudo/Adj. R2	0.326	0.309	0.188	0.160	0.213	0.184

Panel B. Product Market Performance

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Sales Growth $_{t+1}$		MSG (FF49) $_{t+1}$		MSG (SIC4) $_{t+1}$	
Patent Values	High	Low	High	Low	High	Low
High-Value (Low-Value)	0.027***	-0.000	0.026***	0.002	0.024***	0.002
Pat. Utilization Rate	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)
P Value of Diff.	0.012		0.026		0.037	
Model	OLS	OLS	OLS	OLS	OLS	OLS
Firm Controls	✓	✓	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Obs.	21,403	21,403	21,403	21,403	21,403	21,403
Adj. R2	0.100	0.100	0.064	0.064	0.056	0.055

Panel C. Profit Improvements

VARIABLES	(1) $\triangle$GPM $_{t+1}$	(2)	(3) $\triangle$ROA $_{t+1}$	(4)	(5) $\triangle$OCF $_{t+1}$	(6)
Patent Values	High	Low	High	Low	High	Low
High-Value (Low-Value)	0.009**	0.003	0.011***	0.005	0.010***	0.005
Pat. Utilization Rate	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
P Value of Diff.	0.285		0.198		0.279	
Model	OLS	OLS	OLS	OLS	OLS	OLS
Firm Controls	✓	✓	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Obs.	21,450	21,450	21,441	21,441	21,438	21,438
Adj. R2	0.056	0.056	0.019	0.019	0.009	0.008

Panel D. Firm Values

VARIABLES	(1) Log (Tobin's Q) $_{t+1}$	(2)	(3) Log (MTB) $_{t+1}$	(4)
Patent Values	High	Low	High	Low
High-Value (Low-Value)	0.113***	-0.031	0.138***	-0.008
Pat. Utilization Rate	(0.022)	(0.022)	(0.030)	(0.033)
P Value of Diff.	0.000		0.001	
Model	OLS	OLS	OLS	OLS
Firm Controls	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Obs.	21,453	21,453	21,453	21,453
Adj. R2	0.292	0.289	0.261	0.259

Table 9. Heterogeneous Tests: Product Market Competition and Patent Utilization Rate

This table investigates the heterogeneous effects of patent utilization rate on future firms' new product development (panel A), product market performance (panel B), profit improvements (panel C), and firm values (panel D) based on firms' product market competition, which is proxied by the Herfindahl-Hirschman index based on text-based network industry classification. The independent variable *Pat. Utilization Rate* is measured as the number of patents that are applied (and later granted) in the past five years by a firm and that are utilized in the new products launched by the same firm in the current year, further scaled by the total number of patents applied (and later granted) in the past five years for that firm. We regard a patent as utilized in a product if the patent-product pair similarity score is above the 80th percentile. All specifications include firm controls, industry fixed effects, and year fixed effects. Table A1 in Appendix A provides detailed variable definitions. Robust standard errors clustered at the firm level are provided in parentheses. All financial variables are winsorized at the 1st and 99th percentiles. ***, **, and * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A. New Product Development

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	#New Products $_{t+1}$		Sum CARs $_{t+1}$		#Breakthrough Products $_{t+1}$	
TNIC HHI	High	Low	High	Low	High	Low
Pat. Utilization Rate	1.641***	1.767***	6.723***	10.326***	1.489***	1.811***
	(0.085)	(0.117)	(0.533)	(0.736)	(0.173)	(0.205)
P Value of Diff.	0.359		0.000		0.373	
Model	Poisson	Poisson	OLS	OLS	Poisson	Poisson
Firm Controls	✓	✓	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Obs.	9,483	8,795	9,491	8,820	3,303	3,093
Pseudo R2/Adj. R2	0.293	0.446	0.160	0.252	0.189	0.379

Panel B. Product Market Performance

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Sales Growth $_{t+1}$		MSG (FF49) $_{t+1}$		MSG (SIC4) $_{t+1}$	
TNIC HHI	High	Low	High	Low	High	Low
Pat. Utilization Rate	0.005	0.031***	0.007	0.031***	0.007	0.031***
	(0.008)	(0.010)	(0.008)	(0.010)	(0.007)	(0.010)
P Value of Diff.	0.026		0.036		0.036	
Model	OLS	OLS	OLS	OLS	OLS	OLS
Firm Controls	✓	✓	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Obs.	9,478	8,786	9,478	8,786	9,478	8,786
Adj. R2	0.109	0.101	0.055	0.070	0.043	0.063

Panel C. Profit Improvements

VARIABLES	(1) $\triangle$GPM $_{t+1}$	(2)	(3) $\triangle$ROA $_{t+1}$	(4)	(5) $\triangle$OCF $_{t+1}$	(6)
TNIC HHI	High	Low	High	Low	High	Low
Pat. Utilization Rate	-0.001	0.013**	0.003	0.016***	0.004	0.013***
	(0.004)	(0.005)	(0.003)	(0.004)	(0.003)	(0.005)
P Value of Diff.	0.031		0.015		0.116	
Model	OLS	OLS	OLS	OLS	OLS	OLS
Firm Controls	✓	✓	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Obs.	9,489	8,820	9,487	8,817	9,487	8,813
Adj. R2	0.073	0.053	0.024	0.025	0.009	0.012

Panel D. Firm Values

VARIABLES	(1) Log (Tobin's Q) $_{t+1}$	(2)	(3) Log (MTB) $_{t+1}$	(4)
TNIC HHI	High	Low	High	Low
Pat. Utilization Rate	0.022	0.031	0.031	0.065*
	(0.022)	(0.026)	(0.032)	(0.036)
P Value of Diff.	0.791		0.457	
Model	OLS	OLS	OLS	OLS
Firm Controls	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Obs.	9,491	8,820	9,491	8,820
Adj. R2	0.278	0.307	0.263	0.299

Table 10. 2SLS IV Regressions: Federal and State R&D Tax Credit

This table presents the two-stage least squares (2SLS) instrumental-variable (IV) regression results. We follow Bloom et al. (2013) and use federal and state-level R&D tax credits as instruments for a firm's patent utilization rate. The instrument variables in the first stage (column 1) are *Log(Fed. tax credit comp. of R&D user cost)*, which is the natural logarithm of the federal tax credit component of R&D user cost for firm i in year *t-1*, and *Log(State tax credit comp. of R&D user cost)*, which is the natural logarithm of the state tax credit component of R&D user cost for firm i in year *t-1*. Columns 2-4 report the second-stage results of firms' new product developments. Columns 5-7 report the second-stage results on firms' product market outcomes. Columns 8-10 report the second stage results on firms' profit improvement. Columns 11-12 report the second-stage results on firm values. All specifications include firm controls, industry fixed effects, and year fixed effects. Table A1 in Appendix A provides detailed variable definitions. Robust standard errors clustered at the firm level are provided in parentheses. All financial variables are winsorized at the 1st and 99th percentiles. ***, **, and * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

VARIABLES	(1) Pat.Util.Rate$_{t}$	(2) #New Products$_{t+1}$	(3) Sum CARs$_{t+1}$	(4) #Break-through Products$_{t+1}$	(5) Sales Growth$_{t+1}$	(6) MSG (FF49)$_{t+1}$	(7) MSG (SIC4)$_{t+1}$	(8) $\triangle$GPM$_{t+1}$	(9) $\triangle$ROA$_{t+1}$	(10) $\triangle$OCF$_{t+1}$	(11) Log (Tobin's Q)$_{t+1}$	(12) Log (MTB)$_{t+1}$
Log(Fed. tax credit comp. of R&D user cost)$_{t-1}$	-1.944***											
	(0.349)											
Log(State tax credit comp. of R&D user cost)$_{t-1}$	-0.047**											
	(0.023)											
Pat. $\widehat{\textit{Utilization}}$ *Rate*		4.617***	18.797***	0.197*	0.281***	0.228**	0.196**	0.043	0.147***	0.125***	0.580**	0.663*
		(1.171)	(7.557)	(0.118)	(0.096)	(0.092)	(0.089)	(0.049)	(0.044)	(0.044)	(0.254)	(0.355)
Cragg-Donald Wald F statistics	38.311											
P Value of Over-Identification Test		0.874	0.435	0.169	0.171	0.393	0.305	0.083	0.677	0.613	0.387	0.617
Firm Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	12,862	12,862	12,862	12,862	12,862	12,862	12,862	12,862	12,857	12,854	12,862	12,862

Table 11. Robustness Checks

This table conducts robustness checks for the baseline results. In panel A, we further control for *#CW Patents/AT*, which is the citation-weighted number of granted patents that are applied for by a firm in a year scaled by the firm's book value of total assets at the beginning of the year, and *Patent Values/AT*, which is the economic values of granted patents that are applied for by a firm in a year scaled by the firm's book value of total assets at the beginning of the year. In panel B, we replace industry fixed effects with firm fixed effects to further account for any time-invariant firm heterogeneity. All specifications include firm controls but are omitted for succinctness. Table A1 in Appendix A provides detailed variable definitions. Robust standard errors clustered at the firm level are provided in parentheses. All financial variables are winsorized at the 1st and 99th percentiles. ***, **, and * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A. Controlling Patent Count and Patent Values

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
VARIABLES	#New Products$_{t+1}$	Sum CARs$_{t+1}$	#Break-through Products$_{t+1}$	Sales Growth $_{t+1}$	MSG(FF49) $_{t+1}$	MSG(SIC4) $_{t+1}$	$\triangle$GPM$_{t+1}$	$\triangle$ROA$_{t+1}$	$\triangle$OCF$_{t+1}$	Log(Tobin's Q)$_{t+1}$	Log(MTB)$_{t+1}$
Pat.Utilization Rate	**1.699*** **	**9.081*** **	**1.507*** **	**0.014** **	**0.014** **	**0.013** **	**0.007** **	**0.008*** **	**0.007*** **	**0.025* **	**0.047** **
	(0.071)	**(0.470)**	**(0.150)**	**(0.006)**	**(0.006)**	**(0.005)**	**(0.003)**	**(0.002)**	**(0.002)**	**(0.015)**	**(0.022)**
#CW Patents/AT	-0.380	-2.410	2.443***	-0.145***	-0.125***	-0.122***	-0.093***	-0.027	-0.025	0.445***	0.540***
	(0.232)	(2.137)	(0.550)	(0.044)	(0.043)	(0.041)	(0.022)	(0.022)	(0.023)	(0.098)	(0.127)
Patent Values/AT	0.327***	5.318**	0.340	0.132***	0.130***	0.126***	0.018	0.064***	0.066***	0.582***	0.688***
	(0.097)	(2.422)	(0.237)	(0.026)	(0.025)	(0.025)	(0.017)	(0.011)	(0.014)	(0.079)	(0.095)
Model	Poisson	OLS	Poisson	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Other Firm Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	21,453	21,453	7,881	21,403	21,403	21,403	21,450	21,441	21,438	21,453	21,453
Pseudo R2/Adj. R2	0.366	0.203	0.229	0.105	0.069	0.061	0.058	0.026	0.015	0.338	0.290

Panel B. Adding Firm Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
VARIABLES	#New Products$_{t+1}$	Sum CARs$_{t+1}$	#Break-through Products$_{t+1}$	Sales Growth $_{t+1}$	MSG(FF49) $_{t+1}$	MSG(SIC4) $_{t+1}$	$\triangle$GPM $_{t+1}$	$\triangle$ROA $_{t+1}$	$\triangle$OCF $_{t+1}$	Log(Tobin's Q)$_{t+1}$	Log(MTB)$_{t+1}$
Pat.Utilization Rate	0.668***	2.286***	0.586***	0.015*	0.014**	0.015**	0.010**	0.007**	0.006	0.006	0.004
	(0.076)	(0.287)	(0.227)	(0.007)	(0.007)	(0.007)	(0.005)	(0.004)	(0.004)	(0.010)	(0.015)
Firm Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	18,022	21,229	1,652	21,178	21,178	21,178	21,228	21,220	21,217	21,229	21,229
Pseudo R2/Adj. R2	0.502	0.518	0.307	0.224	0.192	0.181	0.053	0.018	-0.007	0.665	0.636

Appendix A

Table A1. Variable Definition

Variables	Definition
Dependent Variables	
CAR(-1, 1)	Cumulative abnormal stock return within a three-day event window of (-1, 1) following the new product announcement event. **Source:** CRSP
Breakthrough Index	A text-based measurement that captures the novelty and impact of a product, following Kelly et al. (2021). Please refer to Section 3.2 of the paper for detailed technical construction of the variable. **Source:** PatentsView and Capital IQ Key Development Database.
1(Breakthrough Product)	An indicator that equals 1 if a product's breakthrough index (*BreakthroughIndex*) is above 95th percentile and equals zero otherwise. Please refer to Section 3.2 of the paper for detailed technical construction of the *BreakthroughIndex*. **Source:** PatentsView and Capital IQ Key Development Database.
#New Products	The number of new products (with three-day CARs above 80th percentile) a firm launches in a year. **Source:** CRSP and Capital IQ Key Development Database.
#Breakthrough Products	The number of breakthrough products (with the breakthrough index (*BreakthroughIndex*) above 95th percentile) a firm launches in a year. Please refer to Section 3.2 of the paper for detailed technical construction of the *BreakthroughIndex*. **Source:** CRSP, PatentsView and Capital IQ Key Development Database.
Sum CARs	The sum of all positive three-day cumulative abnormal stock returns of the new products that a firm launches in a year. **Source:** CRSP and Capital IQ Key Development Database.
Sales Growth	Natural logarithm of total sales for a firm in year $t+1$ minus the natural logarithm of total sales for that firm in year t. **Source:** Compustat.
MSG(FF49)	The sales growth of a firm in year t minus the industry (Fama-French 49 industries) median sales growth in the same year. **Source:** Compustat.
MSG(SIC4)	The sales growth of a firm in year t minus the industry (4-digit SIC) median sales growth in the same year. **Source:** Compustat.
$\triangle$GPM	The change in the gross profit margin of a firm between year $t+1$ and t; the gross profit margin is defined as a firm's sales minus cost of goods sold, further divided by the firm's book value of total assets at the beginning of the year. **Source:** Compustat.
$\triangle$ROA	The change in the return on assets (ROA) of a firm between year $t+1$ and t; the ROA is defined as a firm's operating income before depreciation divided by the firm's book value of total assets at the beginning of the year. **Source:** Compustat.
$\triangle$OCF	The change in the operating cash flow of a firm between year $t+1$ and t; the operating cash flow is defined as a firm's operating cash flow divided by the firm's book value of total assets at the beginning of the year. **Source:** Compustat.
Log(Tobin's Q)	Natural logarithm of a firm's book value of assets minus book value of equity plus market value of equity further divided by the book value of total assets. **Source:** Compustat.

Log(MTB)	Natural logarithm of a firm's market value of assets divided by the book value of total assets. **Source:** Compustat.
Independent Variables	
Log(1+#Patents Utilized$^{\text{Sum}}$)	Natural logarithm of one plus the total number of unique patents utilized in a new product (or a series of new products launched on the same date) by a firm. **Source:** PatentsView and Capital IQ Key Development Database.
Log(1+#Patents Utilized$^{\text{Average}}$)	Natural logarithm of one plus the average number of unique patents utilized in a new product (or a series of new products launched on the same date) by a firm. **Source:** PatentsView and Capital IQ Key Development Database.
Log(1+#Patents Utilized)	Natural logarithm of one plus the total number of patents utilized in a new product by a firm. **Source:** PatentsView and Capital IQ Key Development Database.
Pat. Utilization Rate	The number of patents that are utilized in products scaled by the total number of patents applied (and later granted) in the past five years for a firm in a year. We regard a patent as utilized in a product if the patent-product pair similarity is above 80th percentile. **Source:** PatentsView and Capital IQ Key Development Database.
Log(Product Text Length)	Natural logarithm of the new product announcement text description length. **Source:** Capital IQ Key Development Database.
Log(1+#New Products Launched)	Natural logarithm of one plus the number of new products that have been launched in the year. **Source:** Capital IQ Key Development Database.
#New Products/Sales	The number of new products (with three-day CARs above 80th percentile) a firm launches in a year divided by the sales of that firm in year *t-1*. **Source:** CRSP, Capital IQ Key Development Database and Compustat.
Firm Size	Natural logarithm of the sales of a firm in a year. **Source:** Compustat.
Log(Firm Age)	Natural logarithm of one plus the current year of observation minus the first year a firm appears in Compustat. **Source:** Compustat.
Leverage	The sum of a firm's current liabilities and long-term debt divided by the book value of total assets of the firm. **Source:** Compustat.
R&D	The research and development expenses of a firm in a year divided by the book value of total assets of that firm. **Source:** Compustat.
ROA	A firm's operating income before depreciation divided by the firm's book value of total assets at the beginning of the year. **Source:** Compustat.
Cash	A firm's cash holdings divided by the book value of assets. **Source:** Compustat.
Past Stock Return	Buy-and-hold stock return of a firm. **Source:** CRSP and Compustat.
Log(Federal tax credit component of R&D user cost)	The natural logarithm of federal tax credit component of R&D user cost, $P^F_{f,t}$, for a firm f in year t. As discussed in Section 6.1, a firm's R&D user cost is estimated using the Hall-Jorgenson user cost of capital formula: $\frac{1-D_{f,t}}{1-\tau_{f,t}}$, which incorporates both federal ($P^F_{f,t}$) and state-level ($P^S_{f,t}$) R&D tax credits components. This federal tax-driven R&D user cost $P^F_{f,t}$ is constructed by Hall (1993) and Bloom et al. (2013) based on the interaction between federal R&D tax credit rules and firm-specific factors (e.g., federal R&D tax credits depend on firm-specific "base"). Please refer to Section 6.1 for more details. **Source:** Hall (1993) and Bloom et al. (2013)

Log(State tax credit component of R&D user cost)	The natural logarithm of state tax credit component of R&D user cost, $P^S_{f,t}$, for a firm f in year t. As discussed in Section 6.1, a firm's R&D user cost is estimated using the Hall-Jorgenson user cost of capital formula: $\frac{1-D_{f,t}}{1-\tau_{f,t}}$, which incorporates both federal ($P^F_{f,t}$) and state-level ($P^S_{f,t}$) R&D tax credits components. The state component of the tax price $P^S_{f,t}$ can be formally estimated as follows: $P^S_{f,t} = \sum_s \theta_{i,s,t}\rho^S_{s,t}$. $\theta_{i,s,t}$ is 10-year moving average share of firm i's patent inventors in state s in year t. $\rho^S_{s,t}$ is state i's R&D tax price in year t estimated by Wilson (2009). Please refer to Section 6.1 for more details. **Source:** Wilson (2009) and Bloom et al. (2013)

Table A2. Selecting New Product Launch Announcement Texts

This table reports the keywords that are used to select the training sample of new product launch announcement texts (panel A), the FinBert's classification performance (panel B), and a randomly selected sample of new product launch headlines predicted by the fine-tuned FinBert model (panel C).

Panel A. Keywords about New Product Launches

launch, product, introduce, begin, unveil, release, debut and their variants.

Panel B. FinBert Classification Performance

	Precision	Recall	F1-score	# Headlines
Negative	0.94	0.90	0.92	139
Positive	0.92	0.95	0.93	161
Overall Accuracy			**0.93**	300
Macro Average	0.93	0.92	0.93	300
Weighted Average	0.93	0.93	0.93	300

Panel C. Randomly Selected New-Product-Launch Headlines that are Predicted by the Fine-tuned FinBert Model

Headline	Company Name	Date	New Product Launch
Sally Beauty Holdings, Inc. Announces Nationwide Launch of the Vernon Franois Collection	Sally Beauty Holdings, Inc.	2019-07-25	YES
Quest Diagnostics, Inc. Announces the Availability of OVA1 Blood Test to Aid Pre-Surgical Evaluation of Women for Ovarian Cancer	Quest Diagnostics Inc.	2010-03-09	YES
Thermo Fisher Scientific Launches 300mm FT-IR Metrology Tool	Thermo Fisher Scientific Inc.	2007-07-30	YES
Zoom(R) Modems Ship With ENERGY STAR(R) Qualified Adapters	ZOOM Technologies, Inc.	2007-11-13	YES
Lockheed Martin Offers Advanced Electro-Optical Targeting System for the F-35 Lightning II	Lockheed Martin Co.	2015-09-10	YES
Anthera Pharmaceuticals, Inc. Provides Clinical Program Updates for Blisibimod and Sollpura	Anthera Pharmaceuticals, Inc.	2016-06-28	NO
Acura Pharmaceuticals Provides Update on FDA Discussions Surrounding Development of Aversion Hydrocodone with Acetaminophen Tablet	Acura Pharmaceuticals, Inc.	2014-08-15	NO
MAIA Biotechnology, Inc. Announces HREC Approval in Australia for its THIO-101 Phase 2 Trial for NSCLC	MAIA Biotechnology, Inc.	2022-03-15	NO
Northern Vertex Mining Corp. Announces to Report Recent Results from Its Multi-Phase Infill and Resource Expansion Drilling Program At the Moss Mine in Nw Arizona	Elevation Gold Mining Co.	2021-06-10	NO
Delta Expands Trans-Pacific Service with Nonstop Shanghai-Atlanta Flight	Delta Air Lines, Inc.	2017-07-20	NO

Table A3. Overview of Within-Firm Patent-Product Linkage

This table demonstrates some randomly selected samples of within-firm patent-product linkages. Panel A shows three products and their 5 most (least) similar patents based on the patent-product similarity scores. For each patent-product linkage, we report the patent title, patent number, patent filing year, product name, product launching year, and the similarity score. In Panel B, we further report the excerpts of the three new product descriptions. For each product, we also report the most (least) similar patent text excerpt. The similarity score is estimated based on the *Word2vec* model which computes text similarities between patent text filings and product descriptions.

Panel A. Top-5 (green) and bottom-5 (red) patent-product linkages based on the pair similarity score

Patent Title	Patent Num. (Filing Year)	Product Name (Launching Year)	Similarity
Content analytics system configured to support multiple tenants	9183230 (2012)	IBM Watson Analytics (2014)	0.880
Concurrent execution of request processing and analytics of requests	8819183 (2009)		0.843
Automatic log sensor tuning	9507847 (2013)		0.821
Analytics platform spanning unified subnet	9342345 (2014)		0.791
Analytic solution integration	9098821 (2013)		0.786
Dynamic scan	8516318 (2010)		0.171
Immersion-cooled and conduction-cooled electronic system	8947873 (2012)		0.171
Dynamically reconfiguring time zones in real-time using plural time zone libraries	9740176 (2014)		0.207
Non-uniformity evaluation apparatus, non-uniformity evaluation method, and display inspection apparatus and program	8368750 (2009)		0.208
Land grid array interposer with compressible conductors	8672688 (2012)		0.211
Techniques to transfer data among hardware devices	11132326 (2020)	Nvidia GeForce RTX 3060 (2021)	0.770
Technique for sharing context among multiple threads	11080111 (2020)		0.741
Asynchronous data movement pipeline	11294713 (2020)		0.728
Graphics processing unit systems for performing data analytics operations in data science	11307863 (2019)		0.717
Real-time hardware-assisted GPU tuning using machine learning	10909738 (2018)		0.699
Cross talk reduction differential cross over routing systems and methods	10600730 (2018)		0.345
System and method for procedurally synthesizing datasets of objects of interest for training machine-learning models	10643106 (2018)		0.350
Three state latch	10009027 (2017)		0.385
System and method for cooperative game control	10252171 (2016)		0.387
Resistance and capacitance balancing systems and methods	10685925 (2018)		0.389
Multi-functional hand-held device	11275405 (2006)	Apple IPhone 4 (2010)	0.814
Establishing a video conference during a phone call	8744420 (2010)		0.814
Integrated touch screen	8390582 (2009)		0.813
In conference display adjustments	8502856 (2008)		0.813
Multipoint touchscreen	8125463 (2008)		0.812
Technique for reducing wasted material on a printed circuit board panel	8650744 (2010)		0.395
Low power peer detection circuit	8291241 (2009)		0.393
System and method for internet connected service providing heterogeneous mobile systems with situational location relevant content	8538685 (2007)		0.375
Methods and apparatus for shielding circuitry from interference	8071893 (2009)		0.372
Transaction ID filtering for buffered programmed input/output (PIO) write acknowledgements	8032673 (2009)		0.317

Panel B. Excerpts of patent and product Text

Patent Name & Excerpt	**Product Name & Excerpt**
1. Patent Name: Content analytics system configured to support multiple tenants (**Similarity Score:** 0.880) **Excerpt:** Techniques are disclosed for a software as a ***service (SaaS) provider*** to host a ***content analytics*** tool used to ***evaluate data collections*** for multiple customers (referred to as tenants) using one dedicated and expandable computing infrastructure, without requiring that the service provider obtain, install, license, and manage a separate copy of the content analytics tools for each tenant. ***Customers are provided access to resources dedicated to their enterprise***, but do not have access, or even awareness, of data collections or analytics resources hosted for other customers. That is, embodiments presented herein allow a provider to host content analytics tools used by customers to evaluate their enterprise data in a secure and timely manner. **2. Patent Name:** Land grid array interposer with compressible conductors (**Similarity Score:** 0.211) **Excerpt:** An electrical interconnect is provided for use within, for example, a land grid array (LGA) interposer such as a module-to-board connector. The electrical interconnect includes an electrically-conductive, compressible conductor which has a first conductor end portion and a second conductor end portion. The first and second conductor end portions physically contact in slidable relation each other with compression of the compressible conductor to facilitate inhibiting rotation of the compressible conductor. In one embodiment, the first end portion includes at least one first leg and the second end portion includes at least two second legs, and the at least one first leg and at least two second legs are interdigitated. Further, in one embodiment, the first end portion and the second end portion are each in slidable contact with an inner-facing surface of the compressible conductor.	**Product Name:** IBM Watson Analytics. **Excerpt:** IBM announced Watson Analytics, a natural language-based ***cognitive service*** that can provide instant access to powerful predictive and visual analytic tools for businesses. Watson Analytics is designed to make ***advanced and predictive analytics*** easy to acquire and use for anyone. The first release of Watson Analytics will include a freemium version of its cloud-based service designed to run on desktop and mobile devices. Watson Analytics offers a full range of ***self-service*** analytics, including access to easy to use ***data refinement and data warehousing services*** that make it easier for business users to acquire and prepare data - beyond simple spreadsheets - for analysis and visualization that can be acted upon and interacted with.
1. Patent Name: Techniques to transfer data among hardware devices (**Similarity Score:** 0.770) **Excerpt:** Apparatuses, systems, and techniques to route data transfers between hardware devices. In at least one embodiment, a path over which to ***transfer data from a first hardware component of a computer system to a second hardware component of a computer system*** is determined based, at least in part, on one or more characteristics of different paths usable to transfer the data. In at least one embodiment, first ***CPU*** is communicatively coupled with a first ***peripheral component interconnect (PCI) express (PCIe)*** switch, and second CPU is communicatively coupled with a second PCIe switch. A first ***graphics processing unit (GPU)***, designated at GPU 0, is coupled with third PCIe switch, and a second GPU, designated as GPU 1, is coupled with fourth PCIe switch. In at least one embodiment, memory can include various types of memory devices including ***graphics double data rate ("GDDR") memory***. **2. Patent Name:** Resistance and capacitance balancing systems and methods (**Similarity Score:** 0.389) **Excerpt:** Systems and methods that facilitate resistance and capacitance balancing are presented. In one embodiment, a system comprises: a plurality of ground lines configured to ground components; and a plurality of signal bus lines interleaved with the plurality of ground lines, wherein the interleaving is configured so that plurality of signal bus lines and plurality of ground lines are substantially evenly spaced and the plurality of signal bus lines convey a respective plurality of signals have similar resistance and capacitance constants that are balanced. The plurality of signals can see a substantially equal amount ground surface and have similar amounts of capacitance. The plurality of signal bus lines can have similar cross sections and lengths with similar resistances. The plurality of signal bus lines interleaved with the plurality of ground lines can be included in a two copper layer interposer design with one redistribution layer (RDL).	**Product Name:** Nvidia GeForce RTX 3060. **Excerpt:** NVIDIA Corporation announced that it is bringing the NVIDIA Ampere architecture to millions more PC gamers with the new GeForce RTX 3060 ***GPU***. When combined with a compatible motherboard, this advanced ***PCI Express*** technology enables ***all of the GPU memory to be accessed by the CPU at once***, providing a performance boost in many games. The RTX 3060's key specifications include: 13 shader-TFLOPs; 25 RT-TFLOPs for ray tracing; 101 tensor-TFLOPs to power NVIDIA DLSS (Deep Learning Super Sampling); 192-bit memory interface; 12GB of ***GDDR6 memory***.

1. Patent Name: Multi-functional hand-held device (**Similarity Score:** 0.814) **Excerpt:** The term “multi-functional” is used to define a device that has the capabilities of two or more traditional devices in a single device. The multi-functional device may, for example, include two or more of the following device functionalities: ***cell phone, music player, video player, game player, digital camera, handtop, Internet terminal, GPS or remote control***. The multi-functional ***hand-held device*** also incorporates a variety of input mechanisms, including ***touch sensitive screens***, touch sensitive housings, display actuators, audio input, etc. The device also incorporates a user-configurable GUI for each of the multiple functions of the devices. **2. Patent Name:** Transaction ID filtering for buffered programmed input/output (PIO) write acknowledgments (**Similarity Score:** 0.317) **Excerpt:** A PIO transaction unit includes an input buffer, a response buffer, and a control unit. The input buffer may receive and store PIO write operations sent by one or more transactons sources. Each PIO write operation may include a source identifier that identifies the transaction source. The response buffer may store response operations corresponding to respective PIO write operations that are to be transmitted to the transaction source identified by the identifier. The control unit may store a particular response operation corresponding to the given PIO write operation in the response buffer prior to the given PIO write operation being sent from the input buffer. The control unit may store the particular response operation within the response buffer if the given PIO write operation is bufferable and there is no non-bufferable PIO write operation having a same source identifier stored in the input buffer.	**Product Name:** Apple IPhone 4. **Excerpt:** Apple Inc. presented the new iPhone 4 featuring FaceTime, which makes the dream of ***video calling*** a reality, and Apple’s stunning new ***Retina display***, the highest resolution display ever built into a phone, resulting in super crisp ***text, images and video***. In addition, iPhone 4 features a 5 megapixel ***camera*** with LED flash, ***HD video recording***, Apple’s A4 processor, a 3-axis gyro and up to 40% longer talk time in a beautiful all-new design of glass and stainless steel that is the thinnest smartphone in the world. It can shoot high-definition video, catching up to some other smart phones. It has a gyroscope in addition to other sensors, to enable more advanced motion-sensing applications, such as ***games and mapping services***. The 3.5-inch screen is the same size as on previous models but features 326 pixels per inch, four times more pixels than the earlier iPhones.

Table A4. Alternative Measures of New Products

This table reports the regression results that investigate the association between a firm's patent utilization rate and alternative measures of new product count of the firm. The dependent variable *#Highly-Valued New Products* is measured as the number of highly-valued new products (with three-day CARs above 95th percentile) a firm launches in a year. *1(Highly-Valued New Products)* is an indicator that equals 1 if a firm launches at least one highly-valued new product (with three-day CARs above 95th percentile) in a year. The independent variable *Pat.Utilization Rate* is measured as the number of patents that are applied (and later granted) in the past five years by a firm and that are utilized in the new products launched by the same firm in the current year, further scaled by the total number of patents applied (and later granted) in the past five years for that firm. We regard a patent as utilized in a product if the patent-product pair similarity score is above 80th percentile. All specifications include firm controls. Columns 1 and 4 include industry fixed effects and year fixed effects. Columns 2 and 5 include industry-by-year fixed effects. Columns 3 and 6 include firm and year fixed effects. The results on columns 1-3 are estimated with Poisson regressions, while the results on columns 4-6 are estimated with OLS regressions. Table A1 in Appendix A provides detailed variable definitions. Robust standard errors clustered at the firm level are provided in parentheses. All financial variables are winsorized at the 1st and 99th percentiles. ***, **, and * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	#Highly-Valued New Products $_{t+1}$			1 (Highly-Valued New Product) $_{t+1}$		
Pat. Utilization Rate	1.505***	1.514***	0.515***	0.222***	0.223***	0.056***
	(0.093)	(0.093)	(0.154)	(0.010)	(0.010)	(0.010)
Model	Poisson	Poisson	Poisson	OLS	OLS	OLS
Firm Controls	✓	✓	✓	✓	✓	✓
Industry FE	✓			✓		
Year FE	✓		✓	✓		✓
Industry-Year FE		✓			✓	
Firm FE			✓			✓
Obs.	21,338	19,909	13,628	21,453	21,453	21,229
Pseudo/Adj. R2	0.260	0.262	0.366	0.139	0.138	0.261

Table A5. 2SLS IV Regressions: Exogenous Distant Rivals' Patent Utilization Rate

This table presents the two-stage least squares (2SLS) instrumental-variable (IV) regression results. The instrumental variable in the first stage (column 1), *Distant Rival Pat. Utilization Rate*, is the exogenously determined average distant peers' patent utilization rate. Columns 2-4 report the second-stage results on firms' new product developments. Columns 5-7 report the second stage results on firms' product market outcomes. Columns 8-10 report the second-stage results on firms' profit improvement. Columns 11-12 report the second-stage results on firm values. All specifications include firm controls, industry fixed effects and year fixed effects. Table A1 in Appendix A provides detailed variable definitions. Robust standard errors clustered at the firm level are provided in parentheses. All financial variables are winsorized at the 1st and 99th percentiles. ***, **, and * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	Pat. Util.Rate	#New Products $_{t+1}$	Sum $CARs_{t+1}$	#Break-through Products $_{t+1}$	Sales Growth $_{t+1}$	MSG(FF49) $_{t+1}$	MSG(SIC4) $_{t+1}$	$\triangle GPM_{t+1}$	$\triangle ROA_{t+1}$	$\triangle OCF_{t+1}$	Log(Tobin's $Q)_{t+1}$	Log(MTB) $_{t+1}$
Distant Rivals' Pat. Utilization Rate	0.321***											
	(0.362)											
$\widehat{Pat.\ Utilization\ Rate}$		3.680***	21.548***	0.276***	0.109**	0.072	0.046	0.003	0.068***	0.078***	0.403**	0.601**
		(0.722)	(4.994)	(0.051)	(0.049)	(0.047)	(0.023)	(0.022)	(0.020)	(0.143)	(0.215)	
Cragg-Donald Wald F statistics	288.012											
Firm Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Industry FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Obs.	17,592	17,592	17,592	17,592	17,592	17,592	17,592	17,592	17,584	17,581	17,592	17,592

Table A6. Control for TNIC Product Similarity

This table conducts robustness checks for the baseline results where we control for the average product similarity (*TNIC Product Sim.*) of a firm in a year. All specifications include firm controls but are omitted for succinctness. Table A1 in Appendix A provides detailed variable definitions. Robust standard errors clustered at the firm level are provided in parentheses. All financial variables are winsorized at the 1st and 99th percentiles. ***, **, and * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
VARIABLES	#New Products_{t+1}	Sum CARs_{t+1}	#Break-through Products_{t+1}	Sales Growth $_{t+1}$	MSG(FF49) $_{t+1}$	MSG(SIC4) $_{t+1}$	$\triangle \text{GPM}_{t+1}$	$\triangle \text{ROA}_{t+1}$	$\triangle \text{OCF}_{t+1}$	Log(Tobin's $\text{Q})_{t+1}$	Log(MTB)_{t+1}
Pat.Utilization Rate	**1.552*** **	**8.510*** **	**1.419*** **	**0.011***	**0.013** **	**0.013** **	**0.006***	**0.008*** **	**0.007*** **	**0.047*** **	**0.072*** **
	(0.069)	**(0.473)**	**(0.157)**	**(0.006)**	**(0.006)**	**(0.006)**	**(0.003)**	**(0.002)**	**(0.003)**	**(0.018)**	**(0.025)**
TNIC Product Sim.	-1.002***	-2.417***	-0.081	-0.038***	-0.026***	-0.024***	-0.005	-0.012***	-0.015***	0.003	0.042
	(0.188)	(0.506)	(0.254)	(0.009)	(0.009)	(0.008)	(0.004)	(0.004)	(0.004)	(0.037)	(0.057)
Model	Poisson	OLS	Poisson	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Other Firm Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	18,819	18,819	6,868	18,771	18,771	18,771	18,816	18,809	18,805	18,819	18,819
Pseudo/Adj. R2	0.365	0.206	0.247	0.100	0.065	0.056	0.056	0.018	0.009	0.253	0.242

Table A7. Alternative Measures of Corporate Patent Utilization Rate

This table examines the robustness of the main results using alternative measures of patent utilization rate. We construct alternative measures of patent utilization rate using either 10-year patent portfolio of a firm or alternative patent-product pair score cutoffs (70th or 90th). In addition, we employ a 3-year moving average approach to generate an alternative measure of firm-level patent utilization rate. Moroever, instead of focusing on a 1-year (i.e., current year) patent-product incorporation rate over the past five-year patent portfolio window, we extend the patent usage window to three years. That is, we count the number of unique patents that have been incorporated into new products launched over the past three years (including the current year) from year *t-2* to year *t*, scaled by the total number of unique patents applied for and later granted by the firm from year *t-6* to year *t*. In Panel A (B), we report the results that investigate the relationship between the number of unique patents utilized in a new product and the product's announcement return (breakthrough index). In Panel C, we further investigate the relationship between a firm's patent utilization and the firm's new product development, product market performance, profit improvement, and firm values. All specifications include firm/product controls but are omitted for succinctness. Table A1 in Appendix A provides detailed variable definitions. Robust standard errors clustered at the firm level are provided in parentheses. All financial variables are winsorized at the 1st and 99th percentiles. ***, **, and * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

Panel A. Number of Patents Utilized and New Product Announcement Return

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	CAR (-1,1)				CAR (-1,1)			
				70th Percentile Cutoff				
Log(1+#Patents Utilized$^{\text{Sum}}$)	0.019**	0.031***	0.030***	0.027***				
	(0.008)	(0.009)	(0.009)	(0.010)				
Log(1+#Patents Utilized$^{\text{Average}}$)					0.017**	0.030***	0.028***	0.026**
					(0.008)	(0.009)	(0.009)	(0.011)
Obs.	94,239	94,236	94,108	74,845	94,239	94,236	94,108	74,845
Adj. R2	0.003	0.003	0.055	0.087	0.003	0.003	0.055	0.087
				90th Percentile Cutoff				
Log(1+#Patents Utilized$^{\text{Sum}}$)	0.014*	0.021***	0.020**	0.015*				
	(0.008)	(0.008)	(0.008)	(0.009)				
Log(1+#Patents Utilized$^{\text{Average}}$)					0.012	0.020**	0.018**	0.014
					(0.008)	(0.008)	(0.008)	(0.009)
Obs.	94,239	94,236	94,108	74,845	94,239	94,236	94,108	74,845
Adj. R2	0.003	0.003	0.055	0.087	0.003	0.003	0.055	0.087
				10-Year Patent Portfolio Window				
Log(1+#Patents Utilized$^{\text{Sum}}$)	0.018*	0.034***	0.032***	0.025**				
	(0.009)	(0.011)	(0.010)	(0.012)				
Log(1+#Patents Utilized$^{\text{Average}}$)					0.016*	0.032***	0.030***	0.024*
					(0.009)	(0.011)	(0.010)	(0.012)
Obs.	94,239	94,236	94,108	74,845	94,239	94,236	94,108	74,845
Adj. R2	0.003	0.003	0.055	0.087	0.003	0.003	0.055	0.087
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Industry FE		✓	✓			✓	✓	
Event-Date FE			✓				✓	
Industry-Event-Date FE				✓				✓

Panel B. Number of Patents Utilized and Breakthrough Product (Index and Indicator)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Breakthrough Index				1 (Breakthrough Product)			
				70th Percentile Cutoff				
Log(1+#Patents Utilized)	0.025***	0.024***	0.020***	0.020***	0.003***	0.002***	0.001*	0.002**
	(0.007)	(0.008)	(0.007)	(0.007)	(0.001)	(0.001)	(0.001)	(0.001)
Obs.	105,196	105,193	105,067	86,661	105,196	105,193	105,067	86,661
Adj. R2	0.012	0.041	0.060	0.146	0.014	0.030	0.084	0.201
				90th Percentile Cutoff				
Log(1+#Patents Utilized)	0.026***	0.024***	0.021***	0.020***	0.003***	0.002***	0.001**	0.001**
	(0.006)	(0.007)	(0.006)	(0.006)	(0.001)	(0.001)	(0.001)	(0.001)
Obs.	105,196	105,193	105,067	86,661	105,196	105,193	105,067	86,661
Adj. R2	0.012	0.041	0.060	0.146	0.014	0.030	0.084	0.201
				10-Year Patent Portfolio Window				
Log(1+#Patents Utilized)	0.031***	0.032***	0.029***	0.030***	0.003***	0.003***	0.002**	0.003**
	(0.010)	(0.012)	(0.011)	(0.011)	(0.001)	(0.001)	(0.001)	(0.001)
Obs.	105,196	105,193	105,067	86,661	105,196	105,193	105,067	86,661
Adj. R2	0.013	0.042	0.061	0.147	0.014	0.031	0.084	0.202
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Industry FE		✓	✓			✓	✓	
Event-Date FE			✓				✓	
Industry-Event-Date FE				✓				✓

Panel C. Patent Utilization Rate and firms' new product development, product market performance, profit improvement, and firm values

VARIABLES	(1) #New Products_{t+1}	(2) Sum CARs_{t+1}	(3) #Break-through Products_{t+1}	(3) Sales Growth $_{t+1}$	(4) MSG(FF49) $_{t+1}$	(5) MSG(SIC4) $_{t+1}$	(6) $\triangle\text{GPM}_{t+1}$	(7) $\triangle\text{ROA}_{t+1}$	(8) $\triangle\text{OCF}_{t+1}$	(9) Log(Tobin's $\text{Q})_{t+1}$	(10) Log(MTB)_{t+1}
70th Percentile Cutoff											
Pat. Utilization Rate	1.754***	8.356***	1.507***	0.011**	0.012**	0.011**	0.005*	0.008***	0.008***	0.038**	0.069***
	(0.067)	(0.393)	(0.150)	(0.005)	(0.005)	(0.005)	(0.003)	(0.002)	(0.002)	(0.015)	(0.022)
Obs.	21,453	21,453	7,881	21,403	21,403	21,403	21,450	21,441	21,438	21,453	21,453
Pseudo/Adj. R2	0.367	0.197	0.220	0.100	0.064	0.055	0.056	0.019	0.009	0.290	0.260
90th Percentile Cutoff											
Pat. Utilization Rate	1.637***	10.376***	1.551***	0.015**	0.016**	0.014**	0.009**	0.009***	0.008***	0.039**	0.053*
	(0.075)	(0.644)	(0.155)	(0.007)	(0.006)	(0.006)	(0.003)	(0.003)	(0.003)	(0.019)	(0.027)
Obs.	21,453	21,453	7,881	21,403	21,403	21,403	21,450	21,441	21,438	21,453	21,453
Pseudo/Adj. R2	0.355	0.198	0.214	0.100	0.064	0.055	0.057	0.019	0.009	0.290	0.259
10-Year Window											
Pat. Utilization Rate	1.521***	9.228***	1.588***	0.014**	0.014**	0.013**	0.008**	0.010***	0.010***	0.041**	0.072**
	(0.084)	(0.560)	(0.159)	(0.007)	(0.007)	(0.007)	(0.003)	(0.003)	(0.003)	(0.020)	(0.029)
Obs.	21,453	21,453	7,881	21,403	21,403	21,403	21,450	21,441	21,438	21,453	21,453
Pseudo/Adj. R2	0.335	0.185	0.215	0.100	0.064	0.055	0.056	0.019	0.009	0.290	0.259
3-Year Moving Average											
Pat. Utilization Rate	2.051***	11.867***	1.588***	0.013**	0.013**	0.010	0.007**	0.009***	0.010***	0.046**	0.080**
	(0.084)	(0.609)	(0.163)	(0.007)	(0.007)	(0.006)	(0.003)	(0.002)	(0.002)	(0.022)	(0.032)
Obs.	21,447	21,447	7,880	21,398	21,398	21,398	21,445	21,436	21,434	21,447	21,447
Pseudo/Adj. R2	0.378	0.213	0.218	0.100	0.064	0.055	0.056	0.019	0.009	0.290	0.260
3-Year Patent-Product Incorporation over 5-Year Patent Portfolio											
Pat. Utilization Rate	1.455***	7.010***	1.407***	0.011*	0.010*	0.009	0.001	0.005***	0.003	0.028*	0.063***
	(0.071)	(0.376)	(0.146)	(0.006)	(0.005)	(0.005)	(0.003)	(0.002)	(0.002)	(0.016)	(0.024)
Obs.	21,453	21,453	7,881	21,403	21,403	21,403	21,450	21,441	21,438	21,453	21,453
Pseudo/Adj. R2	0.340	0.181	0.222	0.100	0.064	0.055	0.056	0.019	0.008	0.290	0.260
Model	Poisson	OLS	Poisson	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Firm Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table A8. Robustness Check: Restrict Firm-year Observations with at least One New Product Launch

This table conducts robustness checks for the baseline results by requiring firm-year observations to have at least one new product launch. All specifications include firm controls but are omitted for succinctness. Table A1 in Appendix A provides detailed variable definitions. Robust standard errors clustered at the firm level are provided in parentheses. All financial variables are winsorized at the 1st and 99th percentiles. ***, **, and * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
VARIABLES	#New Products$_{t+1}$	Sum CARs$_{t+1}$	#Break-through Products$_{t+1}$	Sales Growth$_{t+1}$	MSG (FF49)$_{t+1}$	MSG (SIC4)$_{t+1}$	$\triangle$GPM$_{t+1}$	$\triangle$ROA$_{t+1}$	$\triangle$OCF$_{t+1}$	Log(Tobin's Q)$_{t+1}$	Log (MTB)$_{t+1}$
Pat. Utilization Rate	1.097***	7.497***	1.120***	0.017**	0.016**	0.014**	0.007*	0.007**	0.004	0.040**	0.034
	(0.066)	(0.612)	(0.187)	(0.007)	(0.007)	(0.007)	(0.004)	(0.003)	(0.003)	(0.019)	(0.028)
Model	Poisson	OLS	Poisson	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Firm Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	13,967	13,967	5,131	13,950	13,950	13,950	13,964	13,959	13,957	13,967	13,967
Pseudo/Adj. R2	0.230	0.178	0.240	0.138	0.090	0.078	0.084	0.022	0.012	0.284	0.262

Figure A1. (High-Quality) Patent Count per capita in the United States

This figure illustrates the (high-quality) patents per capita in the United States from 1976 to 2016. The blue line represents the number of total patents granted per capita; the purple line is the number of highly-cited patents (with citations above 95th percentile) granted per capita; the grey line denotes the number of breakthrough patents granted per capita, where breakthrough patents are defined by Kelly et al. (2021); the green line shows the number of creative patents granted per capita, where creative patents are measured by Kalyani (2022).

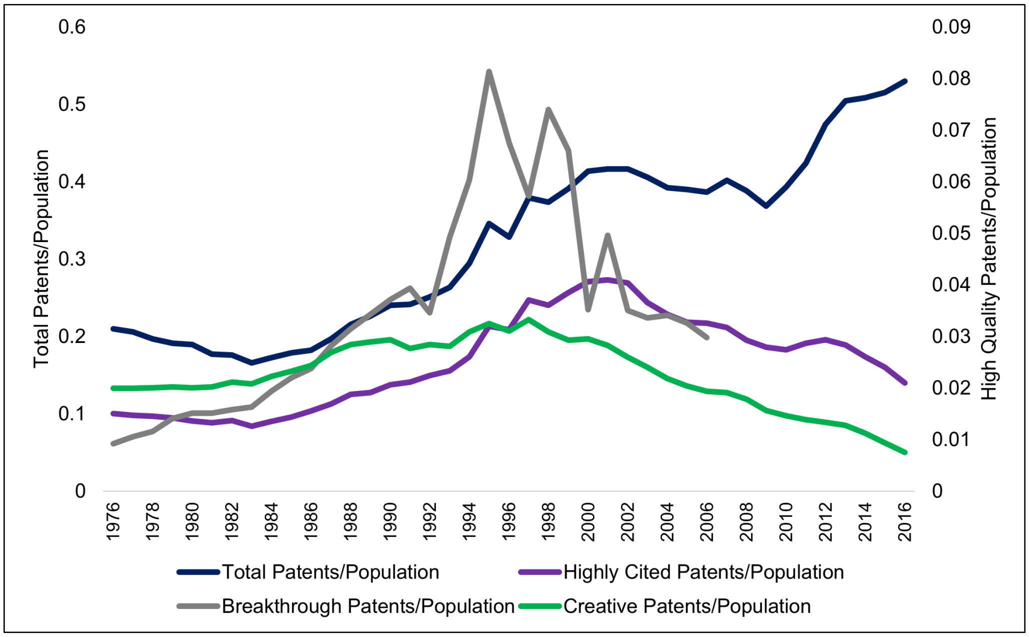
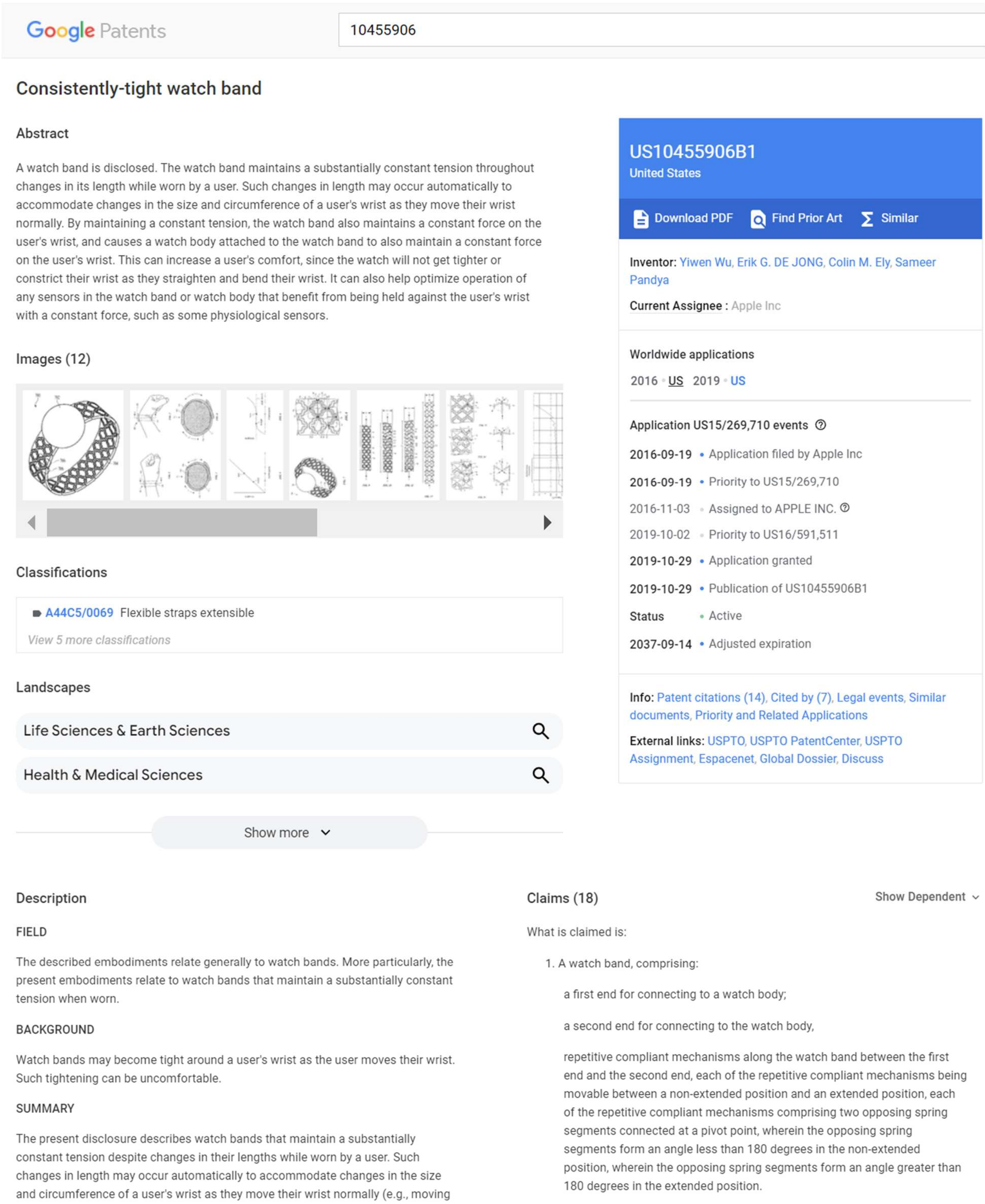


Figure A2. An Example of Patent Text Description

This figure illustrates the text description of the patent "Consistently-tight watch band" applied by Apple Inc. in 2016. The patent text description web page is sourced from Google Patent.



Google Patents 10455906

Consistently-tight watch band

Abstract

A watch band is disclosed. The watch band maintains a substantially constant tension throughout changes in its length while worn by a user. Such changes in length may occur automatically to accommodate changes in the size and circumference of a user's wrist as they move their wrist normally. By maintaining a constant tension, the watch band also maintains a constant force on the user's wrist, and causes a watch body attached to the watch band to also maintain a constant force on the user's wrist. This can increase a user's comfort, since the watch will not get tighter or constrict their wrist as they straighten and bend their wrist. It can also help optimize operation of any sensors in the watch band or watch body that benefit from being held against the user's wrist with a constant force, such as some physiological sensors.

Images (12)

Classifications

A44C5/0069 Flexible straps extensible

View 5 more classifications

Landscapes

Life Sciences & Earth Sciences

Health & Medical Sciences

Show more

US10455906B1

United States

Download PDF Find Prior Art Similar

Inventor: Yiwen Wu, Erik G. DE JONG, Colin M. Ely, Sameer Pandya

Current Assignee : Apple Inc

Worldwide applications

2016 US 2019 US

Application US15/269,710 events

2016-09-19 • Application filed by Apple Inc

2016-09-19 • Priority to US15/269,710

2016-11-03 • Assigned to APPLE INC.

2019-10-02 • Priority to US16/591,511

2019-10-29 • Application granted

2019-10-29 • Publication of US10455906B1

Status • Active

2037-09-14 • Adjusted expiration

Info: Patent citations (14), Cited by (7), Legal events, Similar documents, Priority and Related Applications

External links: USPTO, USPTO PatentCenter, USPTO Assignment, Espacenet, Global Dossier, Discuss

Description

FIELD

The described embodiments relate generally to watch bands. More particularly, the present embodiments relate to watch bands that maintain a substantially constant tension when worn.

BACKGROUND

Watch bands may become tight around a user's wrist as the user moves their wrist. Such tightening can be uncomfortable.

SUMMARY

The present disclosure describes watch bands that maintain a substantially constant tension despite changes in their lengths while worn by a user. Such changes in length may occur automatically to accommodate changes in the size and circumference of a user's wrist as they move their wrist normally (e.g., moving

Claims (18)

Show Dependent

What is claimed is:

1. A watch band, comprising:

a first end for connecting to a watch body;

a second end for connecting to the watch body,

repetitive compliant mechanisms along the watch band between the first end and the second end, each of the repetitive compliant mechanisms being movable between a non-extended position and an extended position, each of the repetitive compliant mechanisms comprising two opposing spring segments connected at a pivot point, wherein the opposing spring segments form an angle less than 180 degrees in the non-extended position, wherein the opposing spring segments form an angle greater than 180 degrees in the extended position.

Figure A3. An Example of Product Text Description

This figure illustrates the text description of Apple Watch SE by Apple Inc. in 2020. The product text description is obtained from Capital IQ key development database.

S&P Capital IQ PRO

Apple Inc. | Key Development Details

NASDAQGS: AAPL (MI KEY: 4004205; SPCIQ KEY: 24937)

Apple Inc. Announces Apple Watch® SE

Apple Inc. announced Apple Watch® SE, packing the essential features of Apple Watch into a modern design — all at a more affordable price. The most advanced Retina® display allows customers to easily see more details and the information that matters most, right on their wrist. Apple Watch SE features the same accelerometer, gyroscope, and always-on altimeter as AppleWatch Series 6, and with the latest motion sensors and microphone, it offers robust health and safety capabilities including fall detection, Emergency SOS, international emergency calling, and the Noise app. With watchOS® 7, users can take advantage of powerful new features including Family Setup, which allows kids or older family members without an iPhone® to enjoy Apple Watch, plus sleep tracking, automatic handwashing detection, and new workouts. Apple Watch SE is available in three beautiful case finishes made of 100% recycled aluminum, and compatible with all Apple Watch bands including the new Solo Loop and Braided Solo Loop. Apple Watch SE features a Retina display, with thin borders and curved corners, that is 30% larger than Series 3. The interface allows for large and easy-to-read app icons and fonts, while complications are precise and informative. A variety of new watch faces are optimized for the display, so customers can easily view notifications, text messages, workout metrics, and more. With the S5 System in Package (SiP) and dual-core processor, Apple Watch SE delivers incredibly fast performance, up to two times faster than AppleWatch Series 3. The Digital Crown® with haptic feedback generates incremental clicks with an extraordinary mechanical feel as it is rotated. Apple Watch SE features the latest speaker and microphone, which are optimized for better sound quality for phone calls, Siri®, and Walkie-Talkie, along with Bluetooth® 5.0. The next-generation always-on altimeter provides real-time elevation all day long by using a new, more power-efficient barometric altimeter, along with GPS and nearby Wi-Fi networks. This feature allows for the detection of small elevation changes above ground level, up and down to the measurement of 1 foot, and can be shown as a new watch face complication or workout metric. The built-in compass provides users with better directions and compass headings, in addition to incline, elevation, and latitude and longitude. Users can add one of three new Compass complications to their watch face, and developers can take advantage of compass information in their apps to create rich experiences. With Emergency SOS on Apple Watch, customers can quickly and easily call for help and alert emergency services with just a push of a button. For added personal safety while traveling, users with cellular models of Apple Watch SE can complete international calls to emergency services, regardless of where the device was originally purchased or if the cellular plan has been activated. Fall detection uses a custom algorithm and the latest accelerometer and gyroscope in Apple Watch SE to detect when a user falls. By analyzing wrist trajectory and impact acceleration, Apple Watch sends the user an alert after a hard fall, which can be dismissed or used to initiate a call to emergency services. If the watch senses immobility for approximately 60 seconds after the fall, it will automatically call emergency services and play an audio message that provides the user's location as latitude and longitude coordinates, in addition to sending a message to emergency contacts. To provide enhanced insights into hearing health, Apple Watch SE utilizes the latest-generation microphone to measure ambient sound levels in a user's environment. Apple Watch SE sends a notification if the decibel level of surrounding sound has risen to a point that it could cause damage, and users can check noise levels at any time through the Noise app or Noise watch face complication.

Company Name	Apple Inc.
Event Date	16/09/2020
Announcement Date	15/09/2020
Development Type	Product-related Announcement
Source	Business Wire
Advisors	NA

Appendix B Technical Details

In Appendix B, we discuss in detail the preprocessing steps for patent and product texts, the advantages of the Word2vec model, the measurement of patent-product pair similarity, the decision of whether a patent is utilized in a product, and the measurement of corporate patent utilization rate.

B1 Text Data Sources and Sample Construction

We first obtain patent filing text data from PatentsView, which provides title, abstract, brief summary text, patent claims, and detailed description sections for each patent granted since 1976. Consistent with prior literature (e.g., Kogan et al., 2022; Kelly et al., 2021), we exploit the full text of patent filings (i.e., aggregate all the five sections of a patent document into a patent-level corpus) for textual analysis. To match patents with the U.S. publicly listed firms, we rely on the linking table developed by Kogan et al. (2017), which matches each patent assignee with a PERMNO ID from CRSP if available. Hence, our final patent text sample consists of 2,544,432 patents generated by the U.S. public firms from 1926 to 2022. Figure A2 illustrates an example of patent text filing from Google Patents website.

We further collect product-related text description data from the Capital IQ Key Development database. After merging each product-related text description with the U.S. publicly listed firms, we obtain 269,472 product-related announcements from 2002 to 2022. As suggested by Cao et al. (2018), there are generally four types of product-related announcements: R&D progress, new product introduction, product improvement, and product retirement. We follow prior studies to focus on the category of new product introduction. To select new product introduction-related announcements, we construct a list of keywords that are related to new product launches following Cao et al. (2018) and Mukherjee et al. (2017).[1] However, this keyword-discovery approach potentially suffers from two issues. First, if the keywords are of a narrow scope, we may not be able to fully capture all announcements that are related to new product launches (false negative). Second, it is also possible that some product-related announcements that we regard as new product launches may actually belong to other types of product announcements (false positive).

To improve the accuracy rate of our new product launch classification, we further employ an advanced natural language processing technique, *FinBert*, to help us automatically determine whether a product-related announcement is about new product introduction or not. Specifically, based on the new product launch keywords, we first construct a training sample that covers 3,000 randomly selected product announcement headlines, of which 1,500 headlines contain at

[1] Panel A of Table A2 lists the new product launches keywords.

least one of those keywords, and the other 1,500 headlines do not. We then manually read each of the 3,000 headlines to decide whether it is related to new product launch or not. With this training sample, we then fine-tune the *FinBert* model. Panel B of Table A2 tabulates the classification performance in the testing sample. Our fine-tuned *FinBert* model can accurately classify 93% of the headlines. Panel C further illustrates some (randomly) selected examples of new-product-introduction-related and non-related headlines predicted by our *FinBert* model. We then use this fine-tuned *FinBert* model to help us classify all the 269,472 product-related announcements. After requiring firms to have at least one patent granted throughout their histories, our final sample consists of 125,329 announcements related to new product launches. We also require our sample firms to have at least one new product launch in the key development database. Thus, our final sample contains 3,102 unique firms that have both produced patents and launched products. In Figure A3, we demonstrate an example of Apple Inc. announcing a new product in 2020.

B2 Preprocessing Text Data

We first remove all non-alphabetic characters, including numbers and punctuation marks, from both the patent filing text and the product announcement text. Next, we split the full text into lists of word tokens. Consistent with the natural language processing (NLP) literature, we further remove stop words from both the patent and product text documents. Stop words are those widely used in a language but contain little significant information. For example, some common stop words include articles (e.g., "the," "an"), prepositions (e.g., "in," "on") and conjunctions (e.g., "and," "but"). To construct the stop word list, we combine multiple sources that are commonly used in NLP: NLTK[2], Spacy[3], Scikit-learn[4], Bill Mcdonald Software Repository for Accounting and Finance[5], WebConfs[6], and MySQL[7]. The final list contains 938 unique stop words.[8]

After removing the stop words, we expect that a considerable proportion of the remaining words in the patent (product) text may provide little information for understanding the functions and characteristics of the patent (product). Thus, we follow Kogan et al. (2022) and Seegmiller et al. (2023) and retain only nouns and verbs, as these two syntactic terms likely contain more informative content. To identify the syntax of each word, we use the part-of-

[2] https://www.nltk.org/book/ch02.html

[3] https://github.com/explosion/spaCy/blob/master/spacy/lang/en/stop_words.py

[4] https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html

[5] https://sraf.nd.edu/textual-analysis/stopwords/

[6] https://www.webconfs.com/stop-words.php

[7] https://dev.mysql.com/doc/refman/8.0/en/fulltext-stopwords.html

[8] When cleaning the patent documents, we further filter out the following words that are commonly used in patent description: claim, present, invention, united, states, patent, description, background, and their variants.

speech tagger package from NLTK (Natural Language Toolkit) in Python. Finally, we convert all the remaining words to lowercase and use the NLTK Lemmatizer package to lemmatize them. Lemmatization is a natural language processing technique that aims to reduce inflected forms of a word to one single form. For example, "running" and "ran" will be lemmatized to "run." After completing the preprocessing steps, we construct a cleaned list of word tokens for each patent and product text.

B3 Measuring Patent Utilization Rate

In this study, we assume that a patent is utilized in a new product within a firm if the text description of the patent filing is abnormally similar to that of the product announcement. Therefore, we first need to compute the textual similarity score between a patent filing and a product description text.

*B3.1 Challenges in "Bag-of-Words" Approach To compute patent-product pair textual similarity score, a conventional approach in economics and finance literature is the "bag-of-words" approach (see Gentzkow et al., 2019). Consider two separate document D_i and D_j. In the "bag-of-words" approach, we convert each document into a one-hot vector V_i and V_j, with dimension equal to 1×N (N represents the number of unique words in these two documents). Each element of the vector, corresponding to a word, is set to zero if the word does not occur in the respective document, otherwise it is set to one. The two documents can thus be represented in two vectors, respectively. Next, we can compute the distance between the two documents using cosine similarity as follows:

$$Sim_{i,j} = \frac{V_i}{||V_i||} \cdot \frac{V_j}{||V_j||} \tag{1}$$

This traditional approach has been frequently employed in prior studies. For example, Hoberg and Phillips (2016) compute product similarity scores for the U.S. public firms by comparing the pairwise distance between their product description sections in 10-K filings. Chen and Srinivasan (2023) construct an industry-level AI technology exposure by comparing text similarity between AI patent abstract and the industry description from NAICS. Kelly et al. (2021) identify breakthrough patents by calculating their patent text similarities. However, the "bag-of-words" approach does not account for semantic similarities between words. That is, words could possess similar meanings even if they are in different forms. For instance, the word "big" is semantically similar to the word "large," but the "bag-of-words" approach will count as a zero match. Consider another extreme case: document i contains the phrase "one beautiful house," while document j contains the phrase "a lovely dwelling." As humans, we can discern the closeness of the two documents. However, when using the "bag-of-words" approach, we transform the two documents into two one-hot vectors, $V_i = [1, 1, 1, 0, 0, 0]$ and $V_j = [0, 0, 0,$

1, 1, 1]. Using Equation 1, we obtain a cosine similarity score of zero, indicating that the two documents are unrelated.

The underestimation bias could be even more pronounced when comparing two documents from different text sources that exhibit diverse language styles (Seegmiller et al., 2023). In this study, we aim to compare the formal, standardized, and legalistic language used in patent filing text descriptions with the more informal and less structured tone typically found in product announcement text descriptions. If we follow the "bag-of-words" approach, the contrasting language styles of the two corpora could lead to sparse one-hot vectors with many elements equal to zero. Consequently, this can result in an underestimated cosine similarity score that is close to zero. Moreover, as the corpus size increases (i.e., the number of unique words), the dimension of the one-hot vector also increases, significantly slowing down computational efficiency.

To summarize, the "bag-of-words" approach has two limitations: i) it fails to capture the semantics of words, and ii) it generates high-dimensional but sparse vectors that are computationally inefficient. To address these issues, we leverage on an advanced machine learning technique, *Word2vec* (Mikolov et al., 2013a), which can produce semantic, low-dimension, and dense word vectors via neural network. Seegmiller et al. (2023) have thoroughly discussed the advantage of *Word2vec* over the "bag-of-words" approach. They also replicate prior text-based measures, such as linking occupations with patents (Kogan et al., 2022), and find that *Word2vec* indeed outperforms the "bag-of-words" approach. We discuss more on the *Word2vec* model in the following subsection.

B3.2 *Word2vec* Model

The essence of the *Word2vec* model is based on the distributional hypothesis that "You should know a word by the company it keeps" (Firth, 1957), which suggests that the meaning of a word can be inferred from its neighboring words. For example, by comparing "I am majoring in Mathematics" and "I am majoring in Finance," we can easily understand that "Mathematics" and "Finance" both refer to specific subjects because they are surrounded by "I am majoring in." Recently, this linguistic concept has been incorporated into neural networks by Mikolov et al. (2013a), where a focal word is used to predict its neighboring words. The final product of *Word2vec* is a $N \times V$ parameter matrix: N denotes the dimension of a vector and V denotes the number of unique words in a corpus. This parameter matrix records the semantic vector representation of each word. Thus, Mikolov et al. (2013a) quantify words into dense and low-dimension vectors that also contain semantic information.[9]

[9] *Word2vec* has two different model architectures to produce semantic word vectors. The first one is Continuous Skip-gram (SG), which uses the focal (center) word to predict its neighboring words. The other is Continuous Bag of Words (CBOW), which instead uses neighboring words to predict the focal (center) words.

Figure B1 illustrates a simple neural network framework for the *Word2vec* model. Specifically, a focal word X_c is first initialized as a $1 \times V$ one-hot vector in the input layer of the neural network, where V represents the number of unique words in the corpus (e.g., patent text). Next, we multiply X_c by W, a $V \times N$ parameter matrix where N denotes the dimension of the final word vectors, which generally varies from 50 to 1,000 depending on research interest.[10] In this step, the initial one-hot vector X_c is projected, from the input layer, into a $1 \times N$ vector V_c in the hidden layer.

[Please insert Figure B1 about here]

Next, V_c is further multiplied by the other $N \times V$ parameter matrix W', which produces the final $1 \times V$ vector, $Y_{c\text{-}m+j}$, where m denotes the window length of neighboring words.[11] We then use the Softmax function to transform $Y_{c\text{-}m+j}$, which is a vector of raw numbers, to a vector of probabilities that predicts the most likely neighboring word of the focal word X_c.[12] Let us call this vector of probabilities $S_{c\text{-}m+j}$.

To maximize the probability of predicting the correct neighboring word, we use the following likelihood function:

$$\begin{aligned} L(W, W') &= P(S_{c+j} | X_c) \\ &= \prod_{c=1}^{V} \prod_{\substack{-m<=j<=m \\ j \neq 0}} P(S_{c+j} | X_c; W, W') \end{aligned} \tag{2}$$

Note that W and W' are the two randomly initialized parameter matrices before the start of the model training process. When the training begins, each focal word in the corpus will be fed forward (i.e., from the input layer to the output layer) in the neural network, predicting its neighboring words. It is common that the model will make prediction errors, that is, the forecasted neighboring words are not the ground truth. To reduce the errors, the model will then feed backward (also called backpropagation in machine learning domain) to fine tune the parameter matrices W and W'. After rounds of iterations, the prediction errors converge and the two parameter matrices become stable. The best parameters should maximize the probability in Equation B2. When the training process is completed, the *Word2vec* model will regard V_c as the focal word X_c's numeric vector (also called word embeddings). Intuitively, V_c is one of the V embeddings in the parameter matrix W. Each embedding has a $1 \times N$ dimension, in which the numeric values indicate the semantic information of the word.

As the *Word2vec* model can produce semantic word embeddings, it significantly alleviates the underestimation issue inherent in the "bag-of-words" approach. The dense and low-

Please see Mikolov et al. (2013b) for more details.

[10] Intuitively, this $V \times N$ parameter matrix W, after model training, records the $1 \times N$ word vector for each of the V unique words. It hence reduces the sparse $1 \times V$ one-hot vector to a dense $1 \times N$ vector for each word.

[11] For example, when m equals to five, it means that this neural network will predict five words before and after the focal word X_c.

[12] The Softmax function restricts the vector of numbers to range from zero to one. The probability of each value in an element is proportional to the relative proportion of each value in the vector.

dimension word vectors also allow for more computationally efficient comparison between documents.

B3.3 *FastText*: An Improved Version of *Word2vec*

Despite the significant progress made by the *Word2vec* model in producing semantic vector representations for words in the vocabulary, there are still limitations: i) it does not provide vectors for words that are rare or out of the training corpus, and ii) it ignores the internal structure of words.[13] Since our text data originate from patent filings and product announcement texts, they likely contain extensive technological descriptions. However, many technical words are rarely seen or entirely absent in conventional training corpora. This can lead to no vector representations for those words when we use the pre-trained language model in later stages. The ignorance of technical words could potentially bias the patent-product pair similarity.

To overcome these challenges, we leverage *FastText* (Bojanowski et al., 2017), an extension of *Word2vec* model that takes into account subword information and also computes word vector representations for words that do not appear in the training corpus. Bojanowski et al. (2017) adopts a similar neural network structure and continuous skip-gram model as *Word2vec* to train *FastText*. But, instead of using a one-hot vector to represent each word in the input layer as outlined in Figure B1, *FastText* splits each word into n-grams (subword).[14] For example, the word "apple" can be split into 3-grams: "app," "ppl," and "ple." After neural network training, we obtain word embeddings not for the simple word "apple," but for each 3-gram "app," "ppl," and "ple." The final word embedding of "apple" will be represented as the sum of all these 3-gram word embeddings. Therefore, the advantage of *FastText* is that rare words or words that are out of the corpus can now be properly represented in semantic vectors by n-grams, as some of their n-grams are likely to appear in other words.

Bojanowski et al. (2017) and Mikolov et al. (2017) empirically examine the performance of the *FastText* model in different language tasks. They find that *FastText* outperforms other models such as the original *Word2vec* (Mikolov et al., 2013b) and *Glove* (Pennington et al., 2014). Thus, we leverage *FastText* to measure document similarity between patent filing text and product announcement text.

[13] Many English word formations follow rules, so morphologically similar words could share similar meanings.For example, the adjective "happy" and the noun "happiness," which are close in meaning, share the same root "happ" and differ only in their suffix. In English, the suffix "ness" generally indicates a noun.

[14] N-grams are all the combinations of adjacent letters with length n in a word.

B3.4 Measuring Patent-Product Pair Textual Similarity using *FastText*

We download the pre-trained English word vectors using the *FastText* model.[15] These 300-dimensional vectors are estimated using skip-gram model with default parameters as introduced in Bojanowski et al. (2017), where the training corpus is sourced from Wikipedia. These pre-trained word embeddings are well recognized, publicly available, and frequently adopted in the computer science domain. Using the publicly available word embeddings also increase the replicability of our results in this paper.

There are alternative word vectors trained on general corpus such as Common Crawl.[16] We choose to use word embeddings that are pre-trained using Wikipedia text as the training corpus because our patent and product text data are more related to scientific fields. General training corpora may not work well in capturing the meanings of technical words. Additionally, Wikipedia generally includes substantial parts of technical descriptions. Therefore, pre-trained word vectors derived from Wikipedia are more likely to capture semantic information closely aligned with the technical context.

After obtaining the word vectors for each word in our corpus (i.e., all unique words in patent and product text), we next aggregate these word vectors to document level using the following equation:

$$D_i = \sum_{X_j \in Z_i} w_{i,j} x_j \tag{3}$$

where D is a vector for document i, measured as the weighted average of the word vectors x for each word j in the set of words Z in document i. Following prior textual analysis literature (see, e.g., Loughran and McDonald, 2011; Li et al., 2021; Hoberg and Phillips, 2016; Kelly et al., 2021), we give different weights w on word vectors based on the importance of the words in our corpus. Consistent with the "bag-of-words" approach, we use the term-frequency-inverse-document-frequency (TFIDF) as our weighting scheme. Specifically, the TFIDF is calculated as:

$$TFIDF_{i,j} \equiv w_{i,j} \equiv TF_{i,j} \times IDF_k \tag{4}$$

The first component of the weight, term frequency (TF), is defined as follows:

$$TF_{i,j} = \frac{c_{i,j}}{\sum_j c_{i,j}} \tag{5}$$

where it counts the number of times word j appears in the document i, further divided by the total number of words in document i. TF thus captures the relative importance of a word in a document. Similar to Loughran and McDonald (2011), the second component of the weight,

[15] The pre-trained word embeddings can be downloaded here: https://fasttext.cc/docs/en/pretrained-vectors.html.

[16] See https://commoncrawl.org/.

inverse-document-frequency (IDF), is measured as:

$$IDF_j = Log\left(\frac{\text{\# Documents in the sample}}{\text{\# Documents that include the word } c}\right) \tag{6}$$

Thus, if a word appears frequently across the set of documents, *IDF* will attenuate its impact using a log transformation. The product of *TF* and *IDF* in Equation B4 can then capture the importance of a given word (vector) j in a given document i. In addition, as we use two distinct sets of text data (patent and product), the corpus of each text source could be significantly different. In this case, we follow the suggestion of Kogan et al. (2022) and Seegmiller et al. (2023) to compute the *IDF* for the words in patent and product text separately. This approach assures that, for instance, the word "patent" will be assigned a much lower weight if it appears in patent documents due to its common occurrence.

After obtaining a dense semantic vector for each document, we use the following equation to measure the cosine similarity between a patent document vector D_p and a product description text D_t within a firm f:

$$Sim_{p,t,f} = \frac{D_{p,f}}{||D_{p,f}||} \cdot \frac{D_{t,f}}{||D_{t,f}||} \tag{7}$$

Unlike Equation B1, Equation B7 emphasizes within-firm patent-product pair similarity because we want to measure a firm's self-invented patent utilization in its new product development. It is worth noting that for a firm's self-invented patents, we only focus on the firm's five-year patent application (later granted) portfolio before the launching date of a new product, since patents may become obsolescent as other technologies evolve (Ma, 2021).[17] The calculation process of within-firm patent-product pair similarity is illustrated in Figure 1. Suppose that firm A launches two products in 2015, NP1 and NP2. We then source the three patents (PAT1, PAT2, and PAT3) that firm A applied (and later granted) in the five years before 2015. For each patent-product pair, we compute its text similarity score using Equation B7.

B3.5 Determining Whether a Patent is Utilized in a Product

Critical to our study is the assumption that a patent is utilized in a new product if the patent-product pair similarity is abnormally high. We acknowledge that this assumption is strong, as high similarity may not definitively indicate utilization in the product. However, it does suggest that the product is very likely to have been heavily influenced by or derived from

[17] The USPTO requires that for patent applications filed after June 8, 1995, the terms of patents will end 20 years after the patent application date. In robustness tests, we also consider the 10-year patent application (later granted) portfolio of a firm and obtain qualitatively similar results.

the patented technology.[18] In this regard, we are similar in spirit to the innovation literature that investigates knowledge diffusion across firms. Prior studies typically use patent citations to determine whether knowledge is diffused across firms (see., e.g., Jaffe et al., 1993; Thompson and Fox-Kean, 2005; Singh and Marx, 2013; Arora et al., 2021; Fadeev, 2023).[19]

Moreover, the recent literature on innovation has categorized patents into process patents, which are inventions of new methods or processes that could improve firms' production efficiency, and non-process (product) patents, which generally refer to inventions of new or improved products (see, e.g., Bena and Simintzi, 2022; Bena et al., 2022). Process-related innovations are of less interest in our study, as these patents primarily focus on improving production processes, while our focus is on whether product patents are utilized in new product development. Therefore, we follow the classification algorithm by Bena et al. (2022) to differentiate process and non-process patents.

Specifically, we define a patent as a process patent if the first patent claim is a process claim, and as non-process patent if the first patent claim is a non-process claim.[20] A patent claim is defined as a process claim if it contains words such as "*A method for . . .*" or "*A process for . . .*", followed by a verb.[21] For example, General Motor's patent "Method for automatic wireless replenishment using DTMF" (US7313382B2) is a process patent as its first claim "*A method for replenishing call-use authorization to a mobile vehicle ...*" is a process claim. After the classification, we retain patent-product pairs where the patents are non-process patents.

Furthermore, since a majority of patent–product pairs within a firm have low textual similarity scores and are considered unrelated to one another, we follow the prior literature (e.g., Kogan et al., 2022; Hoberg and Phillips, 2016) to impose a stringent criteria: we only regard a patent as being utilized in a product if the textual similarity score is above 80th percentile of our sample patent-product pair scores.[22] Table A3 demonstrates some (randomly) selected examples of within-firm patent-product pair linkage.

Finally, we measure a firm's patent utilization rate as the number of granted patents applied forby a firm in the past five years and utilized in the new products launched by the same firm in the current year, scaled by the total number of granted patents applied forby that firm in

[18] Confirming whether a patented technology is indeed being utilized in a product poses a challenge, as it requires seeking advice from technical experts to confirm the usage of the patented technology.

[19] The literature assumes that if a patent of firm A cites a patent of firm B, knowledge is then spillovered from firm B to firm A. In a similar vein, Cohen et al. (2023) assume the utilization of a patent by a firm if the firm has cited the patent previously. Our criteria for patent utilization in products are more stringent, as we require extremely high text similarity between patents and products.

[20] As suggested by Bena and Simintzi (2022), the first claim listed on a patent generally indicates the primary invention.

[21] We use the following keywords to identify process patent claims: "a method,""the method," "method for," "method of," "method in," "method, comprising, "method comprising," "method to," "method applicable to," "a process," "the process," "process for," "process according," "process in," and "process of."

[22] We also consider alternative percentile cutoffs such as 70th and 90th, and obtain qualitatively similar results.

the past five years.

Figure B1. Neural Network Framework for *Word2vec*

This figure presents a simple neural network for the *Word2vec* model. X_c is the focal word which is initialized as a $1\times V$ one-hot vector in the input layer. V represents the number of unique words in the corpus. $W_{V\times N}$ is a $V\times N$ parameter matrix, where N is a dimension of interest which generally varies from 50 to 1000. V_c is a projected vector with the size of $1\times N$ in the hidden layer. $W'_{N\times V}$ is the other parameter matrix with the size of $N\times V$. $Y_{c\text{-}m}$, $Y_{c\text{-}m+1}$, and $Y_{c\text{-}m+2}$ are $1\times V$ prediction vector in the output layer.