Described Patents and Knowledge Diffusion between Firms

S. Katie Moon Paula Suh Guanqun Zhou*

Abstract

Surprisingly, only 7% of widely used front-page patent citations are explicitly described in the patent text (described patents), and 54% in-text described patents are not listed as citations on the front page of patents. Based on in-text described patents, we propose a new measure that captures firms' reliance on, as well as their intentions to acquire, key knowledge from other firms. Our measure of described patents is superior in embodying the true technology-based knowledge flows, as evidenced by citing firms' subsequent hiring of inventors. The widely used front-page patent citations are a noisier measure of knowledge flow due to selection biases introduced by patent lawyers and litigation liabilities. We consistently find that inventors who receive more in-text citations, particularly those described in a positive tone, are more likely to be hired by the citing firm. However, this prediction is significantly weaker using front-page citations. Furthermore, we find that trade secret protection through the Inevitable Disclosure Doctrine (IDD) affects the movement of inventors cited in the patent texts but not those cited on the front page.

Keywords: Knowledge Diffusion, Patent Citations, In-text Citations, Front-page Citations, Trade Secret, Inventor Hiring

JEL Classification: G14, E32, O3, O4

*Moon is at the University of Colorado at Boulder, Leeds School of Business, Campus Box 419, Boulder, CO 80309, USA, [katie.moon@colorado.edu]. Suh is at the University of Georgia, Terry College of Business, 620 S. Lumpkin St., Athens, GA, 30602, [paula.suh@uga.edu]. Zhou is at the University of Colorado at Boulder, Leeds School of Business, Campus Box 419, Boulder, CO 80309, USA, [guanqun.zhou@colorado.edu]

1 Introduction

Patents and patent citations have been widely used as important proxies for the value and quality of innovation.[1] Patent citations, in particular, have played an important role as the key measure of knowledge flow and helped overcome an empirical challenge as "knowledge flows...are invisible; they leave no paper trail by which they may be measured and tracked" (Krugman, 1991). A number of prominent papers have used patent citations to study localized knowledge spillovers and the speed of knowledge diffusion to highlight the agglomeration economies of Marshall (1920) and their impact on productivity (Hedge, Herkenhoff, and Zhu, 2023; Autor, Dorn, Katz, and Van Reenen, 2020; Jaffe, Trajtenberg, and Henderson, 1993).

The literature most commonly uses the citations to previous patents appearing on the *front page* of patent documents (henceforth termed "front-page citations"). However, recent studies raise concerns that front-page citations may not precisely measure the true diffusion of knowledge. The rise of aggressive patent protection strategies, which aims at elevating legal costs of commercializing innovation products (Hall, Helmers, and von Graevenitz, 2021; Shapiro, 2000), has dramatically increased patent filings with overlapping claims to form so-called "patent thickets" and created an overly dense web of citations as a result (Emery and Woeppel, 2024). Hence, strategic patenting and citation practices likely undermine the use of front-page citations as an accurate measure of knowledge diffusion, especially because the incentives to cite other patents on the front page are often driven by the duty of disclosure and shaped by legal strategies (Bryan, Ozcan, and Sampat, 2020). A few studies find that

[1] See Griliches (1981); Trajtenberg (1990); Hall, Jaffe, and Trajtenberg (2005) for the market value of patents. Also, see Acharya and Xu (2017); Howell (2017); Bernstein (2015); Aghion, Van Reenen, and Zingales (2013); Bloom, Schankerman, and Van Reenen (2013) for the use of citations as a proxy for innovation quality.

such motives behind front-page citations distort the relationship between citations and patent values at the high end (Abrams, Akcigit, and Grennan, 2019), and also ascribe the creation of patent thickets to a recent decline in the intensity of knowledge diffusion (Akcigit and Ates, 2021).

This paper proposes the use of described patents in the text of patent documents (termed "in-text citations") as a more accurate measure of knowledge flows, capturing firms' reliance on key knowledge from other firms and their intentions to further acquire that knowledge. By parsing about eight million patent documents from the U.S. Patent and Trademark Office (USPTO), we extract and analyze textual mentions of prior patents within the description texts of patent documents. There are several reasons why in-text citations would be a superior measure of knowledge flow. First, in-text citations come from the descriptive parts of patents, which are primarily written by inventors, in contrast to front-page citations, which are often added by patent lawyers or examiners. Second, while in-text citations are fewer in number, they represent the true knowledge flow more accurately, as they are embedded within the context of the patent descriptions and reflect the inventors' construct of the invention.

To substantiate the contextual usage of in-text citations, we conduct a textual analysis of the patent descriptions surrounding in-text citations and identify the reasons behind the textual mentions of specific patents using the Bayes Classifier machine-learning technique. The result consistently shows that the majority of in-text citations are used to describe the contribution of the underlying patent to the evolution of technological development, over other reasons and, hence, exhibit knowledge flow from the cited patent to the citing patent.

More importantly, unlike front-page citations, in-text citations do not seem to be influenced by firms' strategic behaviors. In Panel (a) of Figure 1, we compare the average numbers of in-text and front-page citations per patent over time. The average number of front-page citations shows exponential growth during our sample period of 1976-2020, accelerating particularly after the mid-2000s, likely associated with the rise of patent thickets. On the other hand, the average number of in-text citations shows only a slight increase over the same period, again suggesting that in-text citations are inherently different from front-page citations in nature, particularly in their susceptibility to strategic uses. In Panel (b), we examine the textual proximity between cited and citing patents over time by citation types. We find that the textual proximity between front-page cited and citing patents keeps decreasing over time, consistent with Kuhn, Younge, and Marco (2020) who suspect that this pattern reflects a violation of the assumption that front-page citations are made solely based on a review of the technological content of each patent. In contrast, the textual proximity between in-text cited and citing patents remain relatively stable over time at a persistently higher level than that of front-page citations.

[Insert Figure 1 Here]

Hence, we formally examine in regressions how in-text and front-page citations correlate with firms' patent litigation. That is, if the underlying incentives of making citations are in part shaped by legal strategies, we would expect firms to adjust their citing behavior in response to patent litigation. We find that in-text citations are unencumbered by patent litigation risks, while the patent litigation as a defendant in year t significantly increases the number of front-page citations in the subsequent year. This result is consistent with the underlying incentives of making front-page citations shaped by legal considerations, as a failure to disclose known prior art in citation can render the subject patent rights unenforceable.

Next, we explore the determinants of in-text citations in more detail. We find that patents that are geographically and technologically closer, as well as those granted more recently, are more likely to be cited as in-text citations. This result is consistent with Jaffe, Trajtenberg, and Fogarty (2000), who find from a survey of inventors that in-person interactions and "word of mouth" constitute one of the modes of knowledge flow, with more recent patents being perceived as more useful. We also find that patents with broader claims and higher stock market value are more likely to be cited in the text.

We then present our main empirical evidence showing that in-text citations represent direct knowledge flow between firms. When a firm seeks to acquire key technology from another firm, it often hires someone who possesses the foundational knowledge of that technology, thereby providing the hiring firm and its R&D with the greatest benefit from the inventor's expertise. There has been a large number of studies showing that inventors themselves are often vehicles for innovation diffusion and that hiring employers' productivity increases from the positive externalities of the moving inventors' human capital (Babina and Howell, 2024; Moretti, 2021; Serafinelli, 2019; Kerr and Lincoln, 2010; Kim and Marschke, 2005). Thus, if in-text citations indeed reflect knowledge diffusion between firms and a firm's intention to acquire key technology from another, they should be strongly correlated with inventor hiring.

Therefore, we examine whether inventors receiving more in-text citations are also more likely to be hired by the citing firms. By comparing inventors who receive citations from the same citing firm, we find that those who received more in-text citations in the past are substantially more likely to be hired by the citing firm. In-text citations have a significantly stronger impact on the likelihood of hiring compared to front-page citations, with the coefficient estimate for in-text citations being seven times greater than that for front-page citations (5% vs. 0.7%).

We next explore factors that either impede or facilitate knowledge flow and how they affect firms' inventor hiring decisions as predicted by in-text citations. We examine the Inevitable Disclosure Doctrine (IDD), which limits labor mobility by preventing employees from joining other firms, especially when doing so would inevitably divulge trade secrets (Dey and White, 2020; Klasa, Ortiz-Molina, and Serfling, 2018). This analysis is particularly relevant to our study, as we aim to show the realizations of technologically important and economically beneficial knowledge flow captured by in-text citations through inventor moves. We expect that the implementation of the IDD will significantly weaken the association between in-text citations and inventor hiring if the hiring of key inventors is inhibited by the concerns for the inevitable disclosure of key knowledge such as trade secrets.

We show that state adoptions of the IDD adversely affect the knowledge flow by impeding inventor moves, especially involving trade secrets, and reducing the effect of in-text citations on inventor hiring by 20%. In contrast, we find no evidence that the IDD affects the relation between front-page citations and inventor hiring, suggesting that front-page citations are less effective in predicting core knowledge flows between firms. Further, we find that both the geographic proximity of subject inventors to citing firms and the technological proximity between citing and cited patents substantially increase the likelihood that citing firms hire the in-text cited inventors by facilitating knowledge flows (Thompson and Fox-Kean, 2005; Jaffe et al., 2000, 1993). Taken together, these results indicate that in-text citations accurately capture the true diffusion of key knowledge between firms, as reflected in inventor hiring decisions.

We then exploit textual analysis of patent descriptions surrounding in-text citations to test whether the sentiment of neighboring words helps predict inventor moves. We employ a lexicon and rule-based sentiment analysis tool to assess the citing patent's tone when describing the cited patents. For example, one of Apple Inc's patents positively describes a patent held by Eastman Kodak in its text, stating: "... images with increased dynamic range would allow for more pleasing contrast improvements, such as described by Lee et al. in commonly assigned U.S. Pat. No. 5,012,333." Using only an in-text citation sample, we find that inventors whose cited patents have a higher sentiment score, computed as a positive minus negative sentiment scores, are incrementally more likely to be hired by the citing firm. That is, the citing firm is more likely to hire inventors whose patents are viewed as more constructive and useful for the firm's own invention. This result further corroborates that in-text citations convey the knowledge flow and subsequent inventor moves more accurately.

Lastly, we strengthen our results on inventor hiring by constructing a matched counterfactual set of potential employers to address concerns arising from using only observed job transitions. The counterfactual sample consists of the actual new firms that inventors join and their five nearest-neighbor firms matched on key characteristics, such as patent CPC classes, one-digit SIC code industry, the number of past front-page citations of the inventor's patents, and firm financial characteristics. The result consistently shows that the likelihood of a firm actually hiring an inventor is strongly associated with the intensity of in-text citations. This result corroborates our conclusion that in-text citations capture meaningful technological linkages that predict real inventor mobility beyond what can be explained by general front-page citation activity.

Overall, our large-sample analyses of in-text citations show that they are unencumbered by legal strategies and explicitly capture technologically driven knowledge flows, as evidenced by their ability to predict inventor hiring decisions. However, our findings do not necessarily suggest that front-page citations are inherently less useful. For example, front-page citations may serve as a better measure of patent value, because a large part of patent value derives not solely from the underlying technology but considerably from the commercialization of the end products embodying the patent and the dynamics of a firm's product markets.

Our paper contributes to several strands of the literature. First, it adds to the broader patent literature, specifically the use of patent citations as a key measure of innovation value and knowledge diffusion. Jaffe et al. (1993) first tackle the fundamental challenge of measuring and tracking the flow of knowledge by using patent citations, followed by a survey of inventors to elaborate on the mechanisms of knowledge spillovers. Hedge et al. (2023) and Autor et al. (2020) also use citation speed to infer the degree of technology diffusion. Thompson and Fox-Kean (2005) and Lerner and Seru (2022) discuss inherent biases in patent citation measures caused by geographic industry clusters and technology group ties. Bryan et al. (2020) first introduce in-text citation as a superior measure of knowledge flow and, importantly, highlight the fundamental problems with using front-page citations as a measure of knowledge flow, as they are often strategically placed to circumvent legal disputes and may not necessarily reflect the underlying technology.

Our paper is most closely related to Bryan et al. (2020), whose analysis focuses exclusively on academic in-text citations, whereas we extend the scope to include a much broader sample of patent-to-patent (or firm-to-firm) in-text citations. We empirically show that patent-to-patent in-text citations are more suited to capture firms' reliance on and intentions to further acquire key knowledge from other firms. We show the realization of firms' intentions to acquire key knowledge from other firms, particularly in the context of inventor hiring decisions, as captured by patent-to-patent in-text citations, compared to their counterparts, patent-to-patent front-page citations.

Second, our paper relates to the literature on knowledge diffusion through inventor moves. Prior work has mainly focused on providing evidence of knowledge diffusion when employees and inventors carry ideas, skills, or technologies with them as they move to new employers (Babina and Howell, 2024; Moretti, 2021; Serafinelli, 2019; Kaiser, Kongsted, and Ronde, 2015; Balsvik, 2011; Kerr and Lincoln, 2010; Kim and Marschke, 2005). Our paper emphasizes instead how firms' intentions to acquire key knowledge from other firms, predicted by in-text citations, are realized through inventor moves.

2 Citation Types

2.1 In-text Citations

We construct an in-text citation dataset by utilizing patent text data sourced from the PatentsView database, which is derived from public bulk data released by the U.S. Patent and Trademark Office (USPTO). The long text tables in the PatentsView database include brief summary text, claims, detailed description text, and drawing descriptions for all published patent applications (since 2001) and granted patents (since 1976). We focus on all utility patents granted between 1976 and 2021, encompassing a total of 7,991,363 patents with available text data.

As the first step in our filtering process, we retain observations that contain textual mentions of the citation prefix "No." Although the format of each in-text citation can vary dramatically due to different citing styles and patent systems, almost all in-text citations include the strings case-insensitive "No." or "Nos." (the plural form of "No."). Using a proper regular expression, we first filter all paragraphs that mention this citation prefix. This initial step serves to reduce the sample size and lessen the data processing burden. After this step, 4,989,161 patents remained that contained the citation prefix and its variations.

The second step involves identifying the various formats of textual mentions of in-text citations. This step presents two main challenges. First, an in-text citation can appear as a patent number, an application number, or a publication number. Second, each of these formats lacks standardized rules or codes, leading to variations in how patent writers document them. For example, the patent with number "7320315", the application with number "11/699572", and the publication with number "20070175455 A1" all refer to the same patent, "Fuel Vapor Treatment System for Internal Combustion Engine." Different patent writers may cite this patent using any of these formats, depending on the availability of citable numbers, which depend on the timing of application, publication, and grant, as well as individual writing preferences. Moreover, we observe further variations in the documentation of each of these three citation forms.

To address these challenges, we consider all possible citation formats comprehensively. This requires extensive manual verifications, through which we summarized 42 common regular expression patterns that patent writers follow when citing prior inventions. These regular expression patterns encompass various formats for patent numbers, publication numbers, and application numbers. We note that these patterns are not necessarily mutually exclusive as our goal is to include as many formats of in-text citations as possible, minimizing false negatives, while excluding obvious false positives. Our final dataset of in-text citations comprises observations of citing patents, their in-text cited patents, and the extracted text surrounding the cited patent numbers. Both citing and cited patents are identified by their patent numbers based on the crosswalk from PatentsView ("pg_granted_pgpubs_crosswalk"), which links patent numbers, application numbers, and publication numbers. We later collapse the dataset at the citing-cited patent pair level by aggregating the number of times each cited patent is mentioned in the citing patent's text.

2.1.1 Sentiment around In-text Citation

We also measure the sentiment of the textual content surrounding each in-text citation by analyzing 50 neighboring words before and after each in-text citation. We employed Valence Aware Dictionary and sEntiment Reasoner (VADER), an open-source lexicon and rule-based sentiment analysis tool. These sentiment scores are only available for in-text citations. Below are examples of the neighboring words surrounding an in-text citation with highly positive versus negative sentiment.

- Positive: *"Methods to increase the dynamic range of images acquired by an image sensing device would allow such images to be rebalanced to achieve a more pleasing rendition of the image. Also, images with increased dynamic range would allow for more pleasing contrast improvements, such as described by Lee et al. in commonly assigned U.S. Pat. No. 5,012,333."* (Citing Firm: Apple Inc, Cited Firm: Eastman Kodak)

- Negative: *"Processes which seek to optimize fuel composition by analysis of the amount* *of biodiesel in a biodiesel blend, such as the processes described in U.S. Pat. No. 7,404,411, fail to address the formulation problems mentioned above because they merely quantify the amount of biodiesel and do not provide any pre-blending qualitative analysis of biodiesel feedstock."* (Citing Firm: Exxon Mobil, Cited Firm: Marathon Petroleum)

2.1.2 Reasons for Making In-text Citation

We further analyze the textual content surrounding each in-text citation to identify its contextual usage. First, we instruct ChatGPT-4, an artificial intelligence (AI) language model developed by OpenAI, to list possible reasons why a patent writer might mention and describe another patent in the text. Below, we quote the exact responses generated by GPT to this instruction, which fall under the following three broad categories:

- Novelty and Advancement Over Existing Technologies: Highlights the invention's unique features and advancements over prior art, emphasizing its originality and superiority to justify patent protection. This reason underscores the importance of distinguishing the new invention from existing technologies.

- Building Upon and Evolving Existing Technology: Demonstrates how the invention builds on and refines existing technologies, showcasing an evolutionary approach to innovation. It emphasizes the invention's contribution to the ongoing progression of technological development.

- Legal and Procedural Context: Addresses the strategic use of citations to navigate patent law, emphasizing the importance of legal compliance and avoidance of infringement. It highlights the role of citations in establishing a legal foundation for the invention.

We then instruct GPT to process each text input we provide and select the contextual reason for citing a given patent from the three categories above. We allow GPT to list multiple reasons for a given text input, if needed, or categorize the text into an 'other' category if none of the three predefined reasons fit the context. For this process, we provide GPT with approximately 50,000 in-text citation instances, randomly selected from the neighboring words around each in-text citation appearing in about 8 million patent documents. Below, we provide examples of the neighboring words surrounding an in-text citation for cases where GPT selects each of the three reasons listed above.

- Novelty and Advancement Over Existing Technologies: *"U.S. Pat. No. 4,051,551 to Kuck et al describes a switch that accesses and aligns data that form multidimensional structures. The switch provides no means for combining requests."* (Citing Firm: IBM, Cited Firm: Unisys)

- Building Upon and Evolving Existing Technology: *"The amino and nucleic acid sequences of such antibodies as set forth in U.S. Pat. Nos. 6,682,736 and 6,984,720 are incorporated by reference herein in their entirety. Briefly, antibodies that can be used in the present invention include antibodies having amino acid sequences of the heavy and light chains of an antibody..."* (Citing Firm: Pfizer, Cited Firm: Medarex)

- Legal and Procedural Context: *"This application is a continuation-in-part of U.S. patent application Ser. No. 09/151,871, filed on Sep. 11, 1998, now U.S. Pat. 6,163,153 issued on Dec. 19, 2000."* (Citing Firm: Baker Hughes, Cited Firm: Western Atlas)

The selected reasons for approximately 50,000 random in-text citation instances are used as training samples in our Bayes classifier.[2] A Bayes classifier is a machine learning model based on Bayes' theorem, which calculates the probability of different categories (in this case, the reasons for citing a patent) given the observed data (the citation context). The classifier learns by analyzing patterns in the training samples and updates its probabilities for each category based on the relationship between the citation context and the predefined

[2] We exclude cases with multiple reasons and those categorized as 'other' to ensure clear training, enabling the classifier to make more accurate predictions for the remaining in-text citations.

reasons. Using these updated probabilities, the classifier can then predict the most likely reason for each in-text citation.

In Figure 2, we present bar graphs illustrating the average probabilities that a patent is mentioned in the text of another patent using the estimated probabilities for each category among the three reasons for all patents in our sample. The average probability that GPT selects the second reason—Building Upon and Evolving Existing Technology—is about 0.92, indicating that nearly all in-text mentions of other patents in a given patent's text are to describe how the subject invention builds upon prior art. The next common reason is Legal and Procedural Context, with an average probability of 0.06. The least common reason is Novelty and Advancement Over Existing Technologies, which asserts that the subject invention is superior to the referenced prior art.

[Insert Figure 2 Here]

This finding indicates that the majority of in-text citations highlight the contribution of the cited patent to the ongoing evolution of technological development. This pattern provides strong evidence that in-text citations directly reflect a clear, directional knowledge flow from the cited to citing patents, driving technological progress.

2.2 Front-page Citations

Front-page citation has been the most commonly used proxy for patent value, quality, and knowledge flows in various literature, largely due to their accessibility and the ease of extraction (Narin and Noma, 1985). We follow the convention in the literature to collect front-page citations. To fully capture all front-page citations for each patent, we aggregate both citations made by U.S. patents to U.S. granted patents (g_us_patent_citation) and citations made by U.S. patents to U.S. patent applications (g_us_application_citation), using data sourced from PatentsView. Similar to our processing of in-text citation data, we apply the same crosswalk to convert publication numbers of front-page citations into their eventual patent numbers.

By combining both in-text citation data and front-page citation data, we construct our final citation dataset at the citing-cited patent pair level, setting the number of in-text citations to zero when a pair exhibits only front-page citation relationship. It is important to note that while front-page citations only provide information on whether a given patent is listed in the reference section of a citing patent, in-text citations offer a wealth of more nuanced information. This includes the occurrence of mentions of a given patent in the citing patent's text, the tone of the description, and the specific reasons for citing the patent.

3 Data and Sample Selection

As described in Section 2, we obtain our patent data from PatentsView and merge it with firm information from Compustat for the firm-level analyses. Our final citation data originally comprises 120 million pairs of in-text and front-page citations on U.S.-granted patents between 1976 and 2020. We are ultimately interested in the inventor moves within the United States. Hence, we only include citation pairs, where both the citing and cited patents are granted to U.S. corporate assignees (both public and private) located in the United States, and exclude self-citations. These filters leave about 53.8 million citation pairs in our final sample. Among them, 14% of the pairs are in-text citations, 46% (or 6.4% of all citation pairs) of which also appear in the front-page citations.

We begin our analysis by comparing patent characteristics by focal patents that receive in-text citations and front-page citations in Panel A of Table 1. There are about 2.9 million unique focal patents in the sample. The majority (92.6%) of citation relations are from receiving front-page citations, and 13.9% of focal patents receive at least one in-text citation. Generally, there are much fewer in-text citations because they are directly placed in the context of technology descriptions in patent documents. Patents can receive both in-text and front-page citations, accounting for about 6.4% of focal patents, whereas only about 7.4% of focal patents receive exclusively in-text citations without front-page listing. In this comparison analysis, we exclude focal patents that receive both in-text and front-page citations to provide a clean comparison of focal patent characteristics.

[Insert Table 1 Here]

The average number of in-text citations that focal patents receive is 2.29, while the average number of front-page citation is 13.56. Focal patents that receive in-text citations have more inventors, especially female inventors, and are more likely to be assigned to U.S.-based public companies compared to those that receive only front-page citations. We find that in-text citations have statistically significantly larger KPSS patent value based on the stock market. However, Bryan et al. (2020) uses only non-patent citations (e.g., scientific literature), whereas we focus on patent-to-patent citations. Moreover, we have a substantially larger sample of patents over a longer sample period. Lastly, focal patents that receive in-text citations tend to make fewer backward front-page citations to earlier inventions and have slightly fewer claims, suggesting that the subject innovation is more likely to be original and novel.

In Panel B of Table 1, we describe cited-citing patent pair-level characteristics. We measure technological proximity, the average geographic distance between cited and citing inventors, and the age gap between cited and citing patents. For the technological proximity measure of two patents, we use the cosine similarity between the 300-dimension vectors of the two patents using the Doc2Vec embedding model.[3] Again, in this analysis, we exclude the citation pairs with both in-text and front-page citation relationships to make a clean comparison. We find that in-text citations exhibit greater technological and geographic proximity. The age gap between the in-text cited and citing patents is also slightly smaller. These observations are consistent with Jaffe et al. (2000) survey responses, where direct communication is chosen as one of the modes of knowledge spillover by the inventors, and more recent patents are deemed more useful. All these findings are statistically significant at the 1% level.

4 Results

4.1 Determinants for Front-page Citations

Our first step in regression analyses is to examine the possible influence of selection biases introduced by patent lawyers or firms' litigation liabilities in making front-page citations. As we note in Panel (a) of Figure 1, firms list exponentially many patents in their reference section (front-page citations), particularly after the mid-2000 with the rise of patent tickets.

[3]We thank the authors of Whalen, Lungeanu, DeChurch, and Contractor (2020) for publicly sharing the embedded vectors of patent texts used in our analysis.

We conjecture that the increase in the number of listed patents is associated with firms' litigation liabilities and examine the correlation between front-page citations and firms' past litigation experiences. To do so, we supplement our sample with the litigation records of patent assignees obtained from Audit Analytics (category type 35 for patent law litigation cases) and regress the number of front-page citations (vs. in-text citations) on the indicator or number of patent litigation where the assignees are sued as defendants. We report the results of this test in Table 2.

[Insert Table 2 Here]

We find in the first two columns of Table 2 that a firm's past litigation as defendants is positively and significantly associated with an increase in the number of front-page citations in their subsequently granted patents. The economic significance of this effect is a 1.3% increase in the average number of front-page citations in the subsequent year if a firm has experienced at least one patent litigation as the defendant in a given year based on column 1. Alternatively, the economic magnitude is a 2.4% increase in the average number of front-page citations that the firm makes in the next year if the firm experiences a 10% increase in the number of patent litigation as the defendant in a given year based on column 2.

Conversely, we do not find any significant effect of the past litigation experience on the number of in-text citations in columns 3 and 4. These results are consistent with the interpretation that the underlying incentives for making front-page citations are related to assignees' legal liabilities and litigation strategies that are influenced by patent lawyers and law firms.

4.2 Determinants for In-text Citations

As a corresponding examination, we now explore the determinants of making in-text citations in this section. In this analysis, we focus on the characteristics of in-text cited patents, including distances in physical, technological, and time-series spaces, commercialization, process versus production innovation (Bena, Ortiz-Molina, and Simintzi, 2022), the breadth of patent applicability, and the patent value measured by the grant-day stock price reaction (Kogan, Papanikolaou, Seru, and Stoffman, 2017). We also examine the effect of financial characteristics of the assignees of the in-text cited patents. Table 3 presents the results of this analysis. Column 1 focuses on the cited-citing patent pair-level characteristics, particularly distances in physical, technological, and time-series spaces. We further add to the regression specification, the patenting characteristics of the cited patents, including the number of claims, the fraction of process claims (vs. product claims), the number of technology classes that the patent encompasses, and the stock price reaction for the assignee firm at the time of grant in column 2, and financial characteristics of the cited patent's assignee, including size, age, value, R&D investment, leverage, fixed asset ratio, product market competition (Hoberg and Phillips, 2014) in column 3.

[Insert Table 3 Here]

In all columns of Table 3, we find that the proximity of the cited and citing patents in terms of geography, technology, and patent granting time consistently increases the likelihood of citing patents' in-text mentions of cited patents. These results suggest that direct communications between inventors and firms (Jaffe et al., 2000) could be a key channel for knowledge spillover, with geographic proximity and overlapping patent timelines acting as facilitators.

In column 2, we also observe that patents with greater commercial potential, measured by the number of patent claims, and those with broader applicability, indicated by the number of reported technology classes, are more likely to be described in the text of citing patents. Also, process innovations, which can be applied to a variety of cost-saving methods in production (Bena et al., 2022), are more frequently cited than product innovations. This suggests that knowledge spillovers related to process innovations occur more often. We also find that patent value, as measured by Kogan et al. (2017), is a strong predictor of whether a patent is described in the text of citing patents. This suggests that in-text citations are indeed a reliable indicator of the technological value of the cited patents.

Finally, in column 3, we find that firm size is negatively associated with the likelihood of being in-text cited. This result is consistent with Akcigit and Goldschlag (2023) and Phillips and Zhdanov (2013), who find that more innovative R&D seems to be done by smaller firms, not large firms.

4.3 Predicting Inventor Moves with Citation Relations

In this section, we shift our focus to inventor moves, analyzing how in-text citation relationships between inventors and firms more effectively capture the positive realization of knowledge spillovers compared to front-page citation relationships. To do so, we specifically estimate the following regression model:

$$Y_{ijt} = \alpha + \beta_1 \mathit{Log(in\text{-}text)}_{ij\{t-2,t\}} + \beta_2 \mathit{Log(front\text{-}page)}_{ij\{t-2,t\}} + \gamma \Gamma_{ijt} + \lambda \Lambda_{jt} + \alpha_j + \delta_t + \epsilon_{ijt}, \quad (1)$$

where i is cited inventor, j is citing firm, and t is year. $\mathit{Log(in\text{-}text)}_{ij\{t-2,t\}}$ and *Log(front-*

page)$_{ij\{t-2,t\}}$ are the log of one plus the average number of in-text and front-page citations, respectively, received by inventor i from firm j in the past three years from $t-2$ to t. We use the previous three-year average number of citations because firms do not patent every year and, hence, do not make citations every year. For Y_{ijt}, we consider $Hired_{ijt}$, which is the hiring outcome in a given year, and $Eventually\ hired_{ij\{s>t\}}$, which accounts for any future hiring of the inventor i by firm j. The vector Γ_{ijt} is the set of control variables, which include an inventor-level variable (*Female inventor*) and firm-inventor-year level variables (*Log(total patents)*). The rest are firm-year level control variables (Λ_{jt}). For both specifications, the regression is at the citing firm-cited inventor-year level from an unbalanced panel and includes firm fixed effects (α_j) and year fixed effects (δ_t).

Table 4 presents the results, estimating Equation 1. The first two columns show the results for $Hired_{ijt}$, while the last two columns show those for $Eventually\ hired_{ij\{s>t\}}$. In columns 2 and 4, we alternatively use cleaner measures for the number of in-text versus front-page citations, excluding cases where a patent is both mentioned in the text (in-text citation) and listed in the reference section (front-page citation) of the same citing patent.

[Insert Table 4 Here]

We first find in column 1 that the average number of citations, in both forms of in-text and front-page citations, that a given inventor receives from a firm in the past three years predict the inventor's joining the citing firm. However, the economic magnitudes of the effects are significantly different between the two types of citations. An inventor with a 10% higher number of in-text citations by a firm in the past three years is more likely to join the citing firm by 5 percentage points. This effect is approximately seven times larger than that of the front-page citations (0.7%), and this difference is statistically significant with an *F-stat* exceeding 50. When we use the cleaner measures of the two types of citations by excluding cases where a patent is both mentioned in the text and listed in the reference section, in column 2, we find slightly stronger results for the difference between the effects of in-text and front-page citations.

Next, in column 3, where we examine the likelihood of the citing firm's hiring the inventor sometime in the future, we find that the firm's in-text citation of the inventor's patents significantly increases the future employment likelihood. The effect is two times larger than that we estimated in column 1 with a different specification. The results in column 4, after excluding cases where a patent is both mentioned in the text and listed in the reference section, are consistent with our previous findings.

Overall, our key empirical evidence in this section with investigating inventor moves clearly shows that in-text citations more effectively capture direct knowledge flow than front-page citations. This finding aligns with the previous literature (Babina and Howell, 2024; Moretti, 2021; Serafinelli, 2019; Kerr and Lincoln, 2010; Kim and Marschke, 2005) that highlights inventors as key vehicles for the innovation diffusion. To further strengthen this argument, in the next section, we examine how legal externalities, particularly the Inevitable Disclosure Doctrine (IDD), which limits labor mobility when the move has the potential for the transfer of core knowledge containing trade secrets of a firm, differentially moderate inventor moves predicted by in-text citations versus front-page citations.

4.4 Trade Secret and Labor Mobility

The IDD governs whether an employee leaving a company can be restricted from joining a competitor based on the argument that their new position would inevitably lead to the disclosure of trade secrets from their previous employer (Lowry, 1988; Klasa et al., 2018). As a result, state adoptions of the IDD limit employee mobility. Our hypothesis is that the likelihood of hiring inventors of cited patents will be affected by the IDD adoption in the states where these inventors reside. However, this effect will only manifest when the citations involve the transfer of core knowledge, which may contain trade secrets from the inventors' current employers. If, as we suggest in this paper, in-text citations more accurately capture these core knowledge transfers, we expect to observe the moderating effect of the IDD exclusively for in-text citations, not for front-page citations.

To test this hypothesis, we incorporate an indicator variable for whether the IDD has been adopted in an inventor's state in a given year. This indicator is based on the list of precedent-setting court decisions, drawn from Table 5 of (Sanati, 2025), that identify when the IDD was adopted or rejected. We then interact this IDD indicator with the number of in-text citations from the previous three years (*Log(in-text)*) in Equation 1. Our analysis focuses on the dependent variable, *Hired*, this time with the IDD indicator strictly reflecting whether the IDD was adopted in the year the inventor was hired. Table 5 present the results.

[Insert Table 5 Here]

In column 1, we find that the interaction term between *Log(in-text)* and IDD is significantly negative at the 1% level, while standalone *Log(in-text)* consistently predicts citing firms' hiring the inventors of in-text cited patents. Specifically, a 1% increase in the number of in-text citations of an inventor's patents by a citing firm is associated with a 56% increase in the likelihood of hiring that inventor by the firm. However, this effect is attenuated when the inventor's state has adopted the IDD. The magnitude of the IDD's mitigating effect is approximately one-fifth of the hiring effect driven by the standalone effect of in-text citations.

We run an analogous test in column 2 with the number of front-page citations. While the hiring effect from the number of front-page citations still exists, albeit much weaker compared to that of in-text citations, we do not observe any mitigating effect of the IDD on the hiring likelihood associated with front-page citations. This result is consistent in column 3, where we consider both in-text and front-page citations together in a single specification.

These findings provide strong evidence that in-text citations truly reflect the direct flow of "core" knowledge, such that hiring an inventor cited in the text of a patent would inevitably lead to the disclosure of trade secrets from the inventor's previous employer. The absence of this effect in the case of front-page citations further strengthens our argument that in-text citations are a superior measure for capturing real technology-based knowledge spillovers, compared to the more commonly used front-page citations.

4.5 Proximity and Sentiment Factors in Hiring Inventors

In this section, we explore the factors that may affect the hiring effect of in-text citations, focusing on two key dimensions: (i) the geographic or technological proximity between inventors and firms, which is likely to affect the cost of searching for and hiring talent, and (ii) the tone with which citing patents describe the cited patents, which reveals the citing firms' attitudes toward the referenced prior art. We aim to have a deeper understanding of how in-text citations affect hiring decisions by examining these additional factors.

In Table 6, we first consider the geographic and technological proximity. As we have done in the previous IDD analysis, we add an interaction term between the number of in-text citations and a measure of proximity to the specification in Equation 1. For the proximity measure, we consider *High geographic proximity* and *High tech proximity*, indicator variables equal to one if the geographic and technological proximity between the citing firm and cited inventor is above the median and zero otherwise, respectively. As in Table 4, the first two columns show the results from estimating Equation 1 for the outcome variable $Hired_{ijt}$, while the last two columns show those for the outcome variable $Eventually\ hired_{ij\{s>t\}}$. Also, in columns 2 and 4, we alternatively use cleaner measures for the number of in-text versus front-page citations, excluding cases where a patent is both mentioned in the text and listed in the reference section of the same citing patent.

[Insert Table 6 Here]

Across all columns in Table 6, the interaction terms are significantly positive at the 1% level. For example, in columns 1 and 2, the estimated incremental economic effects of geographic and technological proximity are approximately 40% and 43%, respectively, indicating that the proximity between inventors and firms significantly amplifies the hiring effects associated with in-text citations.

Next, we consider the tone of textual mentions of in-text cited patents. This analysis is redistricted to the sample of inventor-firm pairs involved in in-text citation relationships, as sentiment scores are only available for textually described citations. The focus of our interest is a measure of the average sentiment score of the text surrounding in-text citations of a given inventor's patents. We consider two sentiment measures including *Net sentiment*, which is the difference between the positive and negative sentiment scores of an inventor's described patents, and its discretized version, *High net sentiment*, which is an indicator variable equal to one if *Net sentiment* is above the median and zero otherwise. Table 7 shows the results.

[Insert Table 7 Here]

We find in Table 7 that the tone of textual descriptions also plays a significant role in predicting whether citing firms will hire the cited inventors. The effect is more pronounced when we estimate the outcome variable, *Eventually hired* that captures the inventor's future employment prospects. For example, in column 3, the estimated coefficient for *Net sentiment* suggests that one standard deviation (0.039) increase in the average sentiment score of the text surrounding the in-text citations increases the likelihood of the citing firm's hiring the inventor in the future by 12.9%. These results particularly underscore the value of described citations in the text of citing patents when compared to those listed in the front-page citations. The words surrounding in-text citations are informative on citing firms' perceptions of the referenced prior art, thereby providing a richer context for understanding inventor hiring decisions.

In Table 8, we reinforce the role of technology proximity inferred by in-text citation in the sample used in Table 7, which consists only of observations associated with in-text citations. We consistently find that technology proximity strongly predicts future employment at the citing firm, particularly for inventors that receive in-text citations.

[Insert Table 8 Here]

4.6 Counterfactual Employers

A potential concern with our earlier findings is that the predictive power of in-text citations for inventors' subsequent employment may be associated by the fact that we only observe actual job transitions. To address this concern, we construct a matched counterfactual set of potential employers. Following the nearest-neighbor approach used in the M&A target selection literature (e.g., Bena and Li, 2014; Erel, Jang, and Weisbach, 2015; Flannery, Hanousek, Shamshur, and Tresl, 2023), we identify, for each observed new employer, five closest firms matched on key characteristics including their number of past front-page citations of the inventor's patents, size, age, R&D intensity, asset tangibility, Tobin's Q, and leverage. We further restrict potential employers to operate in the same one-digit SIC code industry, hold patents in the same CPC classes as the first patent filed by the inventor at the new firm, and be exactly matched on whether the firm has ever cited the inventor on the front page. Using this counterfactual matched sample, we estimate regressions that examine the predictive power of in-text citations on actual hiring outcomes. Table 9 presents the results.

[Insert Table 9 Here]

Table 9 shows that the likelihood of a firm actually hiring an inventor is strongly associated with the intensity of in-text citations. In column 1, we find that whether a potential employer has ever cited the inventor's patents in the text significantly predicts the inventor's eventual move to that firm. In columns 2 to 4, where we use the number of in-text citations (*Log(in-text)*), the estimated coefficients remain positive and highly significant. Because both the actual and counterfactual employers are exactly matched on whether the inventor has ever received front-page citations, the effect of in-text citations reflects additional information beyond front-page citation activity. As expected, the effect of front-page citations is weaker in this analysis due to the matching design. However, the continuous measure of front-page citations (*Log(front-page)*) is insignificant and slightly negative across specifications, suggesting no incremental predictive power along the internal margin.

Columns 3 and 4 further consider the sentiment of text surrounding in-text citations. Consistent with our earlier results in Table 7, the tone of textual mentions provides additional explanatory power. Inventors associated with a higher overall sentiment score are more likely to be hired by potential employers, while negative tone particularly reduces the likelihood of subsequent employment.

Overall, the evidence in Table 9 shows that the predictive power of in-text citations for inventor hiring persists even when benchmarked against matched counterfactual employers. These results confirm our conclusion that in-text citations capture meaningful technological linkages that predict real inventor mobility beyond what can be explained by general front-page citation activity.

5 Conclusion

This paper introduces a novel approach for measuring knowledge flows in innovation by focusing on in-text citations in patent documents. Unlike front-page citations, which are often influenced by strategic legal considerations, in-text citations provide a more accurate and context-specific reflection of the underlying technological relation between citing and cited patents. Our analyses show that in-text citations are a superior measure of knowledge flow, as evidenced by their strong predictive power for inventor moves between firms. In particular, we show that inventors who are cited more frequently in the text of patent documents are significantly more likely to be hired by the citing firm, highlighting the direct link between knowledge flow and labor mobility.

Furthermore, our exploration of the Inevitable Disclosure Doctrine (IDD), which restricts labor mobility by preventing employees from joining rival firms if their work would inevitably disclose trade secrets, shows that in-text citations are indeed closely linked to the true transfer of knowledge, particularly core intellectual property that may involve proprietary or confidential trade secrets. Importantly, front-page citations do not exhibit the same sensitivity to IDD, suggesting that they are less directly linked to the flow of trade secrets or critical proprietary knowledge.

We also find that factors such as geographic and technological proximity, as well as the sentiment expressed in patent descriptions, further enhance the ability of in-text citations to predict inventor hiring. These results suggest that in-text citations are not only a more accurate reflection of knowledge flow but also provide insights into the different shades of inventors' potential contributions to a hiring employer's productivity gains through the transfer of their human capital.

By distinguishing between in-text and front-page citations, this study adds to the literature on patent citation analysis and knowledge diffusion. It provides a clearer understanding of the mechanisms driving inventor mobility and offers a more precise tool for future research on the dynamics of knowledge transfer and the value of innovation. Our findings also contribute to the broader discourse on the limitations of traditional citation measures, highlighting the need for a reevaluation of how we capture and interpret knowledge flow. While front-page citations may still be useful in measuring patent commercialization and legal strategies, our results emphasize the importance of in-text citations as a more reliable proxy for true technological knowledge diffusion.

References

Abrams, David, Ufuk Akcigit, and Jillian Grennan, 2019, Patent value and citations: Creative destruction or strategic disruption?, *Working paper* 1–78.

Acharya, Viral, and Zhaoxia Xu, 2017, Financial dependence and innovation: The case of public versus private firms, *Journal of Financial Economics* 124, 223–243.

Aghion, Philippe, John Van Reenen, and Luigi Zingales, 2013, Innovation and institutional ownership, *American Economic Review* 103, 277–304.

Akcigit, Ufuk, and Sina Ates, 2021, Ten facts on declining business dynamism and lessons from endogenous growth theory, *American Economic Journal: Macroeconomics* 13, 257–298.

Akcigit, Ufuk, and Nathan Goldschlag, 2023, Measuring the characteristics and employment dynamics of u.s. inventors, *NBER Working Paper* 31086.

Autor, David, David Dorn, Lawrence Katz, and John Van Reenen, 2020, The fall of the labor share and the rise of superstart firms, *Quarterly Journal of Economics* 135, 645–709.

Babina, Tania, and Sabrina Howell, 2024, Entrepreneurial spillovers from corporate r&d, *Journal of Labor Economics* 42, 469–509.

Balsvik, Ragnhild, 2011, Is labor mobility a channel for spillovers from multinationals? evidence from norwegian manufacturing, *Review of Economics and Statistics* 93, 285–297.

Bena, Jan, and Kai Li, 2014, Corporate innovations and mergers and acquisitions, *The Journal of Finance* 69, 1923–1960.

Bena, Jan, Hernan Ortiz-Molina, and Elena Simintzi, 2022, Shielding firm value: Employment protection and process innovation, *Journal of Financial Economics* 146, 637–664.

Bernstein, Shai, 2015, Does going public affect innovation?, *Journal of Finance* 70, 1365–1403.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen, 2013, Identifying technology spillovers and product market rivalry, *Econometrica* 81, 1347–1393.

Bryan, Kevin, Yasin Ozcan, and Bhaven Sampat, 2020, In-text patent citations: A user's guide, *Research Policy* 49, 103946.

Dey, Aiyesha, and Joshua White, 2020, Labor mobility and antitakeover provisions, *Harvard Business School Working Paper* 19-102.

Emery, Logan, and Michael Woeppel, 2024, Patent-right uncerntainty and mergers and acquisitions, *Working paper* 1–70.

Erel, Isil, Yeejin Jang, and Michael S Weisbach, 2015, Do acquisitions relieve target firms' financial constraints?, *The journal of finance* 70, 289–328.

Flannery, Mark J, Jan Hanousek, Anastasiya Shamshur, and Jiri Tresl, 2023, M&a activity and the capital structure of target firms, *Journal of Financial and Quantitative Analysis* 58, 2064–2095.

Griliches, Zvi, 1981, Market value, r&d, and patents, *Economics Letters* 7, 183–187.

Hall, Bronwyn, Christian Helmers, and Georg von Graevenitz, 2021, Technology entry in the presence of patent thickets, *Oxford Economic Papers* 73, 903–926.

Hall, Bronwyn, Adam Jaffe, and Manuel Trajtenberg, 2005, Market value and patent citations, *RAND Journal of Economics* 36, 16–38.

Hedge, Deepak, Kyle Herkenhoff, and Chenqi Zhu, 2023, Patent publication and innovation, *Journal of Political Economy* 131, 1845–1903.

Hoberg, Gerald, and Gordon Phillips, 2014, Product market threats, payouts, and financial flexibility, *Journal of Finance* 69, 293–324.

Howell, Sabrina, 2017, Financing innovation: Evidence from r&d grants, *American Economic Review* 107, 1136–1164.

Jaffe, Adam, Manuel Trajtenberg, and Michael Fogarty, 2000, Knowledge spillovers and patent citations: Evidence from a survey of inventors, *American Economic Review* 90, 215–218.

Jaffe, Adam, Manuel Trajtenberg, and Rebecca Henderson, 1993, Geographic localization of knowledge spillovers as evidenced by patent citations, *Quarterly Journal of Economics* 108, 577–598.

Kaiser, Ulrich, Hans Christian Kongsted, and Thomas Ronde, 2015, Does the mobility of rḷabor increase innovation?, *Journal of Economic Behavior & Organization* 110, 91–105.

Kerr, William, and William Lincoln, 2010, The supply side of innovation: H-1b visa reforms and u.s. ethnic invention, *Journal of Labor Economics* 28, 473–508.

Kim, Jinyoung, and Gerald Marschke, 2005, Labor mobility of scientists, technological diffusion, an the firm's patenting decision, *RAND Journal of Economics* 36, 298–317.

Klasa, Sandy, Hernan Ortiz-Molina, and Matthew Serfling, 2018, Protection of trade secrets and capital structure decisions, *Journal of Financial Economics* 128, 266–286.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, 2017, Technological innovation, resource allocation, and growth, *Quarterly Journal of Economics* 132, 665–712.

Krugman, Paul, 1991, in *Geography and trade. Gaston Eyskens Lecture Series.* (MIT Press).

Kuhn, Jeffrey, Kenneth Younge, and Alan Marco, 2020, Patent citations reexamined, *RAND Journal of Economics* 51, 109–132.

Lerner, Josh, and Amit Seru, 2022, The use and misuse of patent data: Issues for corporate finance and beyond, *Review of Financial Studies* 35, 1–61.

Lowry, Suellen, 1988, Inevitable disclosure trade secret disputes: dissolutions of concurrent property interests, *Stanford Law Review* 519–544.

Marshall, Alfred, 1920, Principles of economics, *London Mcmillan* .

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

Narin, F., and E. Noma, 1985, Is technology becoming science?, *Scientometrics* 7, 369–381.

Phillips, Gordon M, and Alexei Zhdanov, 2013, R&d and the incentives from merger and acquisition activity, *The Review of Financial Studies* 26, 34–78.

Sanati, Ali, 2025, How does labor mobility affect corporate leverage and investment?, *Journal of Financial and Quantitative Analysis* 60, 1146–1184.

Serafinelli, Michel, 2019, "good" firms, worker flows, and local productivity, *Journal of Labor Economics* 37, 747–791.

Shapiro, Carl, 2000, Navigating the patent thicket: Cross licenses, patent pools, and standard setting, in *Innovation Policy and the Economy*, volume 1, 119–150 (The University of Chicago Press).

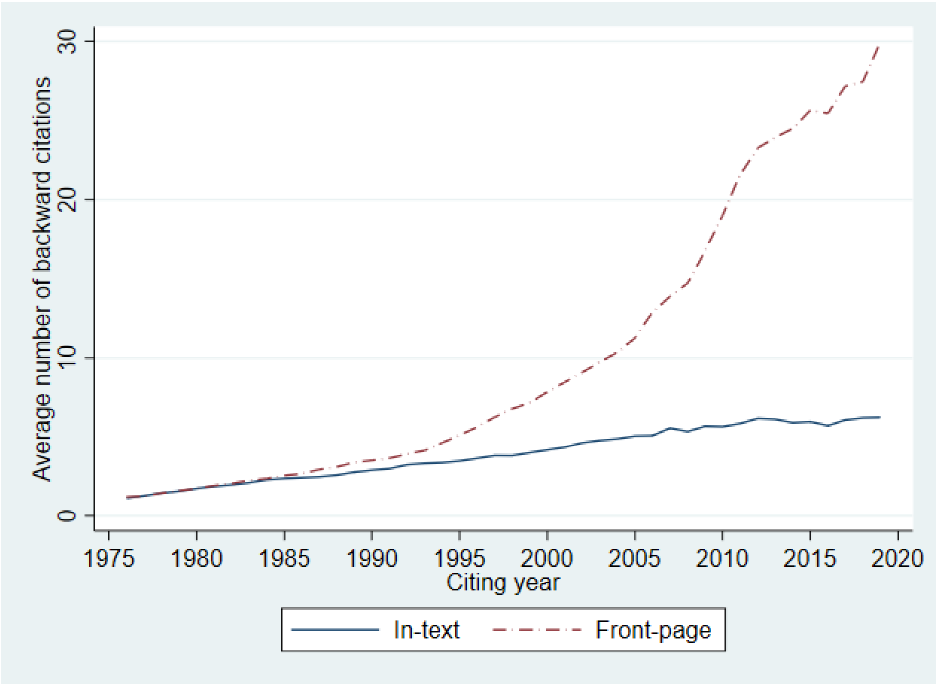
Thompson, Peter, and Melanie Fox-Kean, 2005, Patent citations and the geography of knowledge spillovers: A reassessment, *American Economic Review* 95, 450–459.

Trajtenberg, Manuel, 1990, A penny for your quotes: Patent citations and the value of innovations, *RAND Journal of Economics* 172–187.

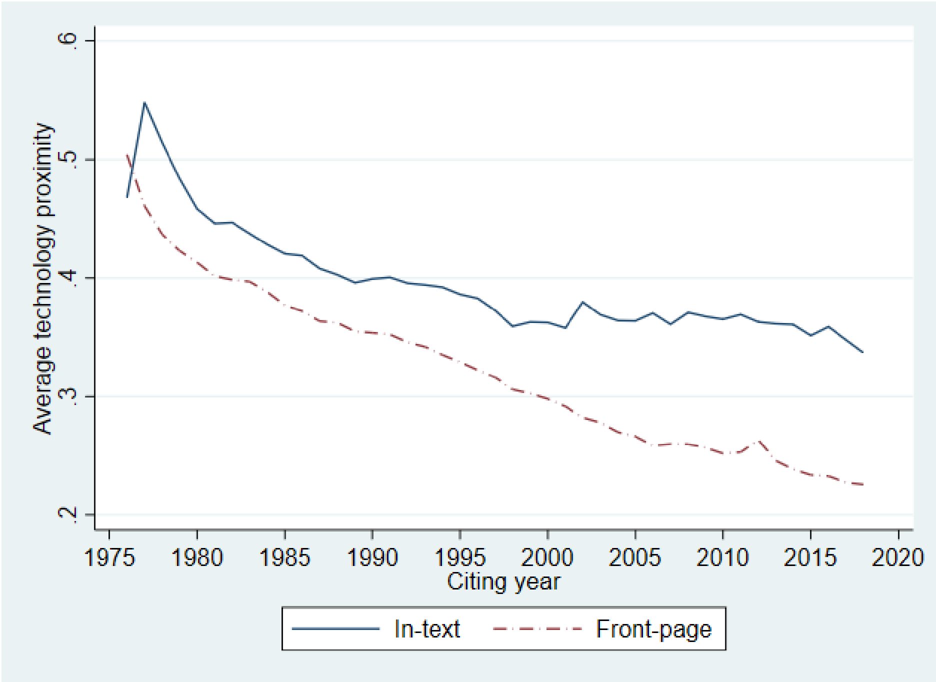
Whalen, Ryan, Alina Lungeanu, Leslie DeChurch, and Noshir Contractor, 2020, Patent similarity data and innovation metrics, *Journal of Empirical Legal Studies* 17, 615–639.

Figure 1: Citation Surge

The figures presents the average number of backward citations per patent over time in Panel (a) and the average technology proximity between the cited and citing patents over time in Panel (b), categorized by citation type. There are two types of citations: in-text citations (blue, solid line) and front-page citations (red, dash-dot line). The sample consists of all U.S. patents granted between 1976 and 2020 for Panel (a) and 1976-2018 for Panel (b) due to data limitation.



(a) Number of backward citations



(b) Average technology proximity

Figure 2: Reasons for Textual Mentions of Citations

This figure presents bar graphs illustrating the average probabilities that a patent is mentioned in the text of another patent. Using GPT-4, we classified the reasons for in-text citations by analyzing approximately 50,000 in-text citation instances, focusing on the following three categories:

- Novelty and Advancement Over Existing Technologies: Emphasizes the invention's unique features and advancements compared to prior art, highlighting its originality and superiority to justify patent protection. This category underscores the importance of distinguishing the new invention from existing technologies.
- Building Upon and Evolving Existing Technology: Demonstrates how the invention builds upon and refines existing technologies, showcasing its contribution to the continuous evolution of technological development. This category emphasizes the invention's role in advancing the state of the art.
- Legal and Procedural Context: Focuses on the strategic use of citations to navigate patent law, with an emphasis on legal compliance and the avoidance of infringement. It highlights the role of citations in establishing a legal foundation for the invention.

We note that the three categories were also derived from GPT-4's responses to our query about the possible reasons why a patent is mentioned in the text of another patent. Using the reasons identified by GPT-4 for about 50,000 in-text citations, we apply a machine learning technique—Bayes Classifier—to classify the remaining 9 million textual mentions based on their contextual usage.

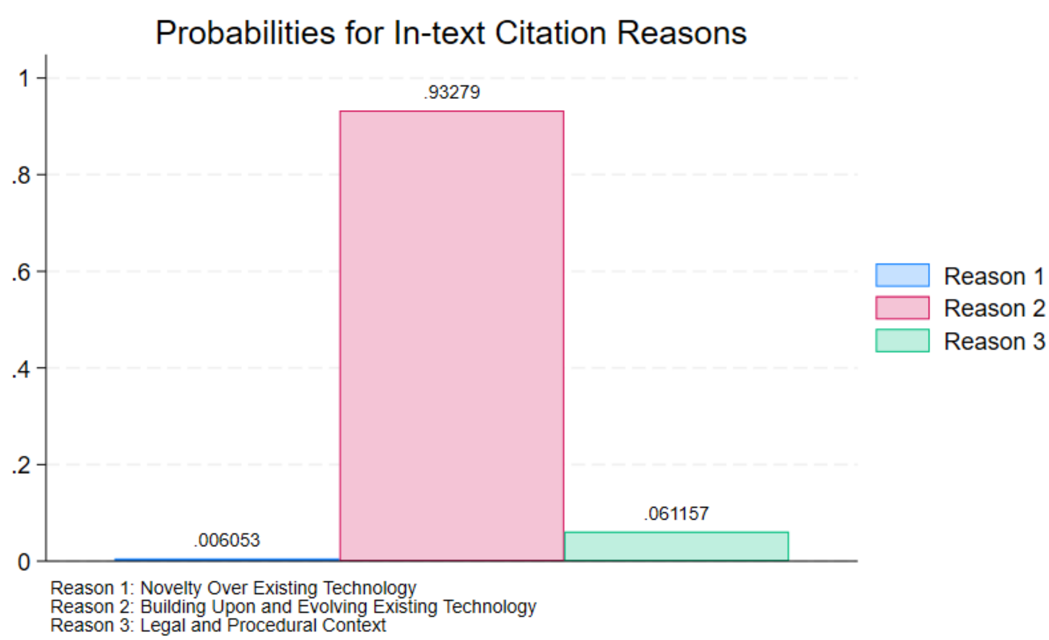


Table 1: Descriptive Statistics of Cited Patents

The table presents summary statistics for the characteristics of cited patents. The sample consists of U.S. patents granted to U.S. corporations between 1976 and 2020, with observations at the level of the cited patent and its citation type. There are two types of citations: in-text citations and front-page citations. This sample excludes non-mutually exclusive citation types. The variables, *Technology proximity*, *Geographic distance*, and *Age gap* are pairwise values between the cited and citing patents. All other variables are patenting characteristics at the time of the patent grant. For variable definitions and further details of their construction, see Appendix A.

Panel A: Patent-level focal patent characteristics

	In-text			Front-page			
	mean	median	sd	mean	median	sd	Difference
Number of citations	2.29	1	24.10	13.56	5	37.80	-11.28***
Number of inventors	2.81	2.00	1.98	2.56	2.00	1.77	0.26***
Fraction of female inventors	0.09	0.00	0.21	0.07	0.00	0.19	0.02***
Fraction of US inventors	0.88	1.00	0.30	0.86	1.00	0.33	0.02***
Public assignee	0.52	1.00	0.50	0.43	0.00	0.50	0.09***
KPSS value	16.08	5.72	36.11	14.23	5.94	35.59	1.85***
Number of backward citations	2.56	1.00	4.57	13.60	5.00	32.25	-11.04***
Number of claims	16.83	17.00	10.76	17.45	17.00	11.41	-0.62***
Observations		85,059			1,272,581		

Panel B: Citation pair-level characteristics

	Cited-citing pair relation						
	In-text			Front-page			
	mean	median	sd	mean	median	sd	Difference
Technology proximity	0.37	0.32	0.19	0.25	0.23	0.12	0.12***
Geographic distance (U.S. inventors only)	553.13	239.13	715.46	948.18	734.20	824.28	-395.05***
Geographic distance	972.99	428.05	1332.37	1509.39	1034.60	1638.95	-536.40***
Age gap	9.84	8.00	8.31	10.18	8.00	7.96	-0.34***
Observations		3,995,907			46,311,720		

Table 2: Patent Litigation and Front-page Citations

The table examines the effect of firms' previous patent litigation experiences on the subsequent number of front-page and in-text citations they make. The sample consists of U.S. public firms with patents granted between 1984 and 2020. The sample period starts in 1984 because the patent litigation data is available from 1984. The observations are at the firm-year level. *Log(front-page made)* and *Log(in-text made)* are the firm-year level natural log of one plus the number of front-page and in-text citations made by the firm in $t+1$. *Defendant litigation* is the indicator variable equal to one if the firm has been a defendant and zero otherwise. *Log(num defendant litigation)* is the log of one plus the number of defendant cases. Standard errors in parentheses are clustered at the firm level. ***p<0.01, **p<0.05, *p<0.1. For variable definitions and further details of their construction, see Appendix A.

	Log(front-page made)$_{t+1}$		Log(in-text made)$_{t+1}$	
	(1)	(2)	(3)	(4)
Defendant litigation	0.013***		0.003	
	(0.004)		(0.004)	
Log(num defendant litigation)		0.024***		0.006
		(0.006)		(0.005)
Log(total patents)	-0.006***	-0.007***	-0.005***	-0.005***
	(0.002)	(0.002)	(0.002)	(0.002)
Log(total citations)	0.889***	0.887***	0.239***	0.239***
	(0.014)	(0.014)	(0.017)	(0.017)
Log(asset)	0.009***	0.008***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Log(age)	-0.005***	-0.005***	0.002	0.002
	(0.001)	(0.001)	(0.001)	(0.001)
R&D_asset	0.014**	0.014**	0.010	0.009
	(0.007)	(0.007)	(0.007)	(0.007)
PPE_asset	0.001	0.001	0.005*	0.005*
	(0.003)	(0.002)	(0.002)	(0.002)
ROA	-0.002	-0.002	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Tobin's Q	0.000	0.000	0.000*	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	-0.007**	-0.007**	-0.000	-0.000
	(0.003)	(0.003)	(0.003)	(0.003)
KPSS value	0.000*	0.000*	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Fluidity	-0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Citing Firm FE	Y	Y	Y	Y
Citing Year FE	Y	Y	Y	Y
Observations	84698	84698	84698	84698
Adjusted R^2	0.928	0.928	0.846	0.846

Table 3: Determinants for In-text Citations

The table examines the determinants of being in-text cited. The sample consists of U.S. patents granted to U.S. corporations between 1976 and 2020, with observations at the level of the cited patent and its citation type. There are two types of citations: in-text citations and front-page citations. The variables, *Geographic distance*, *Technology proximity*, and *Age gap* are pairwise values between the cited and citing patents. All other variables are patenting or financial characteristics at the time of the cited patent grant. ***p<0.01, **p<0.05, *p<0.1. For variable definitions and further details of their construction, see Appendix A.

	(1)	(2)	(3)
	In-text cited (×100)		
Log(Geographic distance)	-0.037***	-0.037***	-0.036***
	(0.002)	(0.002)	(0.002)
Technology proximity	0.805***	0.803***	0.816***
	(0.044)	(0.044)	(0.044)
Age gap	-0.060***	-0.061***	-0.052***
	(0.010)	(0.010)	(0.013)
Log(Number of claims)		0.017***	0.017***
		(0.001)	(0.002)
Fraction of process claims		0.003	0.003
		(0.002)	(0.002)
Number of tech classes		0.016***	0.015***
		(0.001)	(0.001)
KPSS value		0.0006**	0.0004**
		(0.0002)	(0.0002)
Log(asset)			-0.016***
			(0.004)
Log(age)			0.002
			(0.007)
R&D_asset			-0.002
			(0.029)
PPE_asset			-0.017
			(0.011)
Tobin's Q			0.002
			(0.001)
Leverage			0.019**
			(0.009)
Fluidity			0.002
			(0.002)
Cited Firm FE	Y	Y	Y
Cited Patent Grant Year FE	Y	Y	Y
Observations	1306640	1304459	1108405
Adjusted R^2	0.170	0.172	0.168

Table 4: In-text Citations and Inventor Hiring

The table examines the cross-sectional effects of the number of front-page and in-text citations made to an inventor and her subsequent employment at the citing firm. The sample consists of patenting U.S. public firms between 1976 and 2020. The observations are at the citing firm-cited inventor-year level. *Hired (×100)* is an indicator variable equal to one if the cited inventor is hired in the year by the citing firm and zero if otherwise, which is then multiplied by 100. *Eventually hired (×100)* is an indicator variable equal to one if the cited inventor is eventually hired by the citing firm at any future time and zero if otherwise, which is then multiplied by 100. *Log(in-text)* and *Log(front-page)* are the natural log of one plus the average inventor-level number of in-text or front-page citations received from the citing firm in the past 3 years $(t-2, t)$. *Log(clean in-text)* and *Log(clean front-page)* exclude non-mutually exclusive citation types. Standard errors in parentheses are clustered at the inventor and citing firm level. ***p<0.01, **p<0.05, *p<0.1. For variable definitions and further details of their construction, see Appendix A.

	Hired (×100)		Eventually hired (×100)	
	(1)	(2)	(3)	(4)
Log(in-text)	0.498***		0.814***	
	(0.059)		(0.093)	
Log(front-page)	0.071***		0.219***	
	(0.011)		(0.032)	
Log(clean in-text)		0.520***		0.874***
		(0.074)		(0.120)
Log(clean front-page)		0.076***		0.222***
		(0.010)		(0.031)
Female inventor	0.013**	0.013**	0.001	0.001
	(0.006)	(0.006)	(0.024)	(0.024)
Log(total patents)	0.282***	0.286***	0.413***	0.420***
	(0.020)	(0.020)	(0.079)	(0.079)
Log(asset)	0.001	0.001	-0.182***	-0.182***
	(0.014)	(0.014)	(0.062)	(0.063)
Log(age)	-0.045**	-0.044**	-0.186***	-0.185***
	(0.018)	(0.018)	(0.064)	(0.064)
R&D_asset	0.479	0.485	-0.069	-0.059
	(0.346)	(0.349)	(0.475)	(0.478)
PPE_asset	-0.067	-0.069	0.052	0.048
	(0.121)	(0.122)	(0.226)	(0.224)
Tobin's Q	0.001	0.001	0.048***	0.047***
	(0.004)	(0.004)	(0.016)	(0.016)
Leverage	0.111	0.116	0.050	0.059
	(0.084)	(0.084)	(0.230)	(0.231)
Fluidity	-0.003	-0.003*	0.003	0.003
	(0.002)	(0.002)	(0.008)	(0.008)
H_0: $\beta_{in-text} = \beta_{front-page}$				
F-stat	56.59	38.78	47.98	34.38
p-value	0.00	0.00	0.00	0.00
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Observations	15010455	15010455	15010455	15010455
Adjusted R^2	0.005	0.004	0.010	0.010

Table 5: In-text Citations, Labor Mobility, and Inventor Hiring

The table examines the cross-sectional effects of the number of front-page and in-text citations made to an inventor and her subsequent employment at the citing firm. The sample consists of patenting U.S. public firms between 1976 and 2020. The observations are at the firm-cited inventor-year level. *Hired (×100)* is an indicator variable equal to one if the cited inventor is hired in the year by the citing firm and zero if otherwise, which is then multiplied by 100. *Log(in-text)* and *Log(front-page)* are the natural log of one plus the average inventor-level number of in-text or front-page citations received from the citing firm in the past 3 years $(t-2, t)$. IDD is an indicator for whether the Inevitable Disclosure Doctrine (IDD) has been adopted in the inventor's residing state in a given year. Standard errors in parentheses are clustered at the inventor and citing firm level. ***p<0.01, **p<0.05, *p<0.1. For variable definitions and further details of their construction, see Appendix A.

	Hired (×100)		
	(1)	(2)	(3)
Log(in-text) × IDD	-0.110***		-0.110***
	(0.042)		(0.042)
Log(front-page)× IDD		-0.009	-0.016
		(0.011)	(0.011)
Log(in-text)	0.561***		0.547***
	(0.060)		(0.059)
Log(front-page)		0.104***	0.077***
		(0.013)	(0.012)
IDD	-0.009	-0.009	0.008
	(0.008)	(0.010)	(0.012)
Female inventor	0.012**	0.014**	0.013**
	(0.006)	(0.006)	(0.006)
Log(total patents)	0.288***	0.291***	0.283***
	(0.020)	(0.020)	(0.020)
Log(asset)	0.004	-0.001	0.001
	(0.014)	(0.015)	(0.014)
Log(age)	-0.047***	-0.040**	-0.045**
	(0.018)	(0.018)	(0.018)
R&D_asset	0.476	0.507	0.478
	(0.344)	(0.351)	(0.347)
PPE_asset	-0.063	-0.073	-0.067
	(0.120)	(0.123)	(0.121)
Tobin's Q	0.002	0.001	0.001
	(0.004)	(0.004)	(0.004)
Leverage	0.108	0.128	0.112
	(0.083)	(0.085)	(0.083)
Fluidity	-0.003	-0.003	-0.003
	(0.002)	(0.002)	(0.002)
Firm FE	Y	Y	Y
Year FE	Y	Y	Y
Observations	15010455	15010455	15010455
Adjusted R-squared	0.005	0.003	0.005

Table 6: Geographic and Technological Proximity and Inventor Hiring

The table examines the cross-sectional effects of the number of front-page and in-text citations made to an inventor and her subsequent employment at the citing firm. The sample consists of patenting U.S. public firms between 1976 and 2020. The observations are at the firm-cited inventor-year level. *Hired (×100)* is an indicator variable equal to one if the cited inventor is hired in the year by the citing firm and zero if otherwise, which is then multiplied by 100. *Eventually hired (×100)* is an indicator variable equal to one if the cited inventor is eventually hired by the citing firm at any future time and zero if otherwise, which is then multiplied by 100. *High geographic proximity* and *High tech proximity* are indicator variables equal to one if the average geographic or technological distance between the citing firm and cited inventor patent citation pairs is below the median and zero otherwise.*Log(in-text)* and *Log(front-page)* are the natural log of one plus the average inventor-level number of in-text or front-page citations received from the citing firm in the past 3 years $(t-2, t)$. Standard errors in parentheses are clustered at the inventor and citing firm level. ***p<0.01, **p<0.05, *p<0.1. For variable definitions and further details of their construction, see Appendix A.

	Hired (×100)		Eventually hired (×100)	
	(1)	(2)	(3)	(4)
Log(in-text)×High geographic proximity	0.395***		0.604***	
	(0.064)		(0.110)	
High geographic proximity	0.090***		0.277***	
	(0.009)		(0.025)	
Log(in-text)×High tech proximity		0.431***		0.531***
		(0.062)		(0.097)
High tech proximity		0.106***		0.458***
		(0.010)		(0.042)
Log(in-text)	0.239***	0.203***	0.410***	0.425***
	(0.028)	(0.027)	(0.054)	(0.061)
Log(front-page)	0.066***	0.061***	0.211***	0.202***
	(0.011)	(0.009)	(0.033)	(0.030)
Female inventor	0.012**	0.007	-0.001	-0.017
	(0.006)	(0.006)	(0.024)	(0.024)
Log(total patents)	0.280***	0.275***	0.409***	0.390***
	(0.020)	(0.019)	(0.079)	(0.078)
Log(asset)	0.001	0.001	-0.185***	-0.185***
	(0.014)	(0.014)	(0.062)	(0.062)
Log(age)	-0.045**	-0.047**	-0.187***	-0.194***
	(0.018)	(0.018)	(0.065)	(0.065)
R&D_asset	0.488	0.465	-0.050	-0.107
	(0.342)	(0.344)	(0.471)	(0.474)
PPE_asset	-0.071	-0.065	0.042	0.059
	(0.120)	(0.121)	(0.226)	(0.228)
Tobin's Q	0.001	0.001	0.048***	0.048***
	(0.004)	(0.004)	(0.016)	(0.016)
Leverage	0.110	0.107	0.049	0.034
	(0.083)	(0.083)	(0.230)	(0.230)
Fluidity	-0.002	-0.002	0.004	0.005
	(0.002)	(0.002)	(0.008)	(0.008)
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Observations	15010455	15010455	15010455	15010455
Adjusted R^2	0.006	0.006	0.011	0.011

Table 7: Textual Sentiment and Inventor Hiring

The table examines the cross-sectional effects of the sentiment of in-text citation made to an inventor and her subsequent employment at the citing firm. The sample consists of patenting U.S. public firms between 1976 and 2020, excluding observations that do not have any in-text citations, thus do not have the sentiment score. The observations are at the firm-cited inventor-year level. *Hired (×100)* is an indicator variable equal to one if the cited inventor is hired in the year by the citing firm and zero if otherwise, which is then multiplied by 100. *Eventually hired (×100)* is an indicator variable equal to one if the cited inventor is eventually hired by the citing firm at any future time and zero if otherwise, which is then multiplied by 100. *Net sentiment* is the difference between the positive and negative sentiment scores of an inventor's cited patents. *High net sentiment*, is an indicator variable equal to one if the average sentiment score of the cited inventors' patents is below the median and zero otherwise. Standard errors in parentheses are clustered at the inventor and citing firm level. ***p<0.01, **p<0.05, *p<0.1. For variable definitions and further details of their construction, see Appendix A.

	Hired (×100)		Eventually hired (×100)	
	(1)	(2)	(3)	(4)
Net sentiment	0.811*		2.311**	
	(0.472)		(0.966)	
High net sentiment		0.030		0.143**
		(0.035)		(0.061)
Female inventor	0.118***	0.118***	0.131	0.132
	(0.040)	(0.040)	(0.090)	(0.090)
Log(total patents)	1.459***	1.459***	1.115***	1.116***
	(0.119)	(0.119)	(0.109)	(0.109)
Log(front-page)	0.281***	0.281***	0.499***	0.498***
	(0.046)	(0.046)	(0.074)	(0.074)
Log(asset)	0.014	0.014	-0.302	-0.302
	(0.117)	(0.117)	(0.199)	(0.199)
Log(age)	-0.262**	-0.262**	-0.702***	-0.702***
	(0.111)	(0.111)	(0.240)	(0.240)
R&D_asset	0.862	0.865	0.122	0.133
	(1.038)	(1.039)	(1.489)	(1.490)
PPE_asset	-0.393	-0.393	-0.961	-0.961
	(0.693)	(0.693)	(0.957)	(0.957)
Tobin's Q	-0.004	-0.004	0.019	0.019
	(0.024)	(0.024)	(0.045)	(0.045)
Leverage	0.662	0.662	0.893	0.891
	(0.406)	(0.406)	(0.553)	(0.553)
Fluidity	-0.020*	-0.020*	-0.011	-0.011
	(0.012)	(0.012)	(0.022)	(0.022)
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Observations	1147701	1147701	1147701	1147701
Adjusted R^2	0.034	0.034	0.036	0.036

Table 8: Technology Proximity and Inventor Hiring

The table examines the cross-sectional effects of the technology proximity of in-text citation made to an inventor and her subsequent employment at the citing firm. The sample consists of patenting U.S. public firms between 1976 and 2020, excluding observations that do not have any in-text citations, thus do not have the sentiment score. The observations are at the firm-cited inventor-year level. *Hired (×100)* is an indicator variable equal to one if the cited inventor is hired in the year by the citing firm and zero if otherwise, which is then multiplied by 100. *Eventually hired (×100)* is an indicator variable equal to one if the cited inventor is eventually hired by the citing firm at any future time and zero if otherwise, which is then multiplied by 100. *Net sentiment* is the difference between the positive and negative sentiment scores of an inventor's cited patents. *High net sentiment*, is an indicator variable equal to one if the average sentiment score of the cited inventors' patents is below the median and zero otherwise. Standard errors in parentheses are clustered at the inventor and citing firm level. ***p<0.01, **p<0.05, *p<0.1. For variable definitions and further details of their construction, see Appendix A.

	Hired (×100)		Eventually hired (×100)	
	(1)	(2)	(3)	(4)
Technology proximity	3.952***		6.979***	
	(0.339)		(0.506)	
High technology proximity		0.694***		1.268***
		(0.075)		(0.113)
Female inventor	0.046	0.081**	0.002	0.064
	(0.041)	(0.039)	(0.095)	(0.091)
Log(total patents)	1.303***	1.414***	0.839***	1.034***
	(0.112)	(0.116)	(0.106)	(0.106)
Log(front-page)	0.233***	0.224***	0.422***	0.394***
	(0.043)	(0.041)	(0.069)	(0.065)
Log(asset)	0.029	0.020	-0.294	-0.292
	(0.121)	(0.116)	(0.203)	(0.197)
Log(age)	-0.292**	-0.272**	-0.746***	-0.722***
	(0.113)	(0.111)	(0.242)	(0.240)
R&D_asset	0.899	0.853	0.134	0.110
	(1.058)	(1.035)	(1.513)	(1.484)
PPE_asset	-0.393	-0.367	-0.948	-0.914
	(0.712)	(0.691)	(0.978)	(0.955)
Tobin's Q	-0.002	-0.006	0.024	0.015
	(0.023)	(0.023)	(0.044)	(0.044)
Leverage	0.700*	0.654	0.887	0.877
	(0.413)	(0.404)	(0.556)	(0.550)
Fluidity	-0.015	-0.018	-0.005	-0.007
	(0.012)	(0.012)	(0.022)	(0.022)
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Observations	1096698	1147701	1096698	1147701
Adjusted R^2	0.037	0.035	0.040	0.037

Table 9: In-text Citations and Inventor Hiring with Counterfactual Employers

The table examines the effect of in-text citations on actual inventor moves relative to randomly drawn matched employers. The sample consists of observed new firms that inventors actually join and their five nearest-neighbor firms matched on key characteristics, including the potential employer's number of past front-page citations of the inventor's patents, firm size, age, R&D, asset tangibility, Tobin's Q, and leverage. All potential employers operate in the same 1-digit SIC industry, hold patents in the same CPC classes as the first patent that the inventor file at the new firm, and are exactly matched on whether the firm has ever cited the inventor on the patent front page. *Actually hired* is an indicator variable equal to one if the inventor joins the firm and zero if the firm is one of the randomly drawn matched employers. *Log(in-text)* is the natural log one plus the total number of in-text citations of the inventor's patents by the firm. *Log(front-page)* is the natural log one plus the total number of front-page citations of the inventor's patents by the firm. *Sentiment* is the average compound sentiment score across the inventor's in-text paragraphs cited previously by the firm. *Positive sentiment* is the average positive sentiment across the inventor's in-text paragraphs cited previously by the firm. *Negative sentiment* is the average negative sentiment across the inventor's in-text paragraphs cited previously by the firm. All firm control variables are in the year of the inventor's actual move. Standard errors in parentheses are clustered at the inventor level. ***p<0.01, **p<0.05, *p<0.1. For variable definitions and further details of their construction, see Appendix A.

	Actually hired			
	(1)	(2)	(3)	(4)
$\mathbb{1}$(in-text)	0.442***			
	(10.18)			
Log(in-text)		0.479***	0.419***	0.528***
		(8.02)	(6.31)	(5.91)
Log(front-page)		-0.100	-0.0779	-0.0787
		(-1.03)	(-0.84)	(-0.86)
Sentiment			0.178**	
			(2.02)	
Positive sentiment				0.764
				(0.60)
Negative sentiment				-4.030***
				(-3.15)
Log(asset)	0.0660***	0.0660***	0.0660***	0.0660***
	(15.77)	(15.77)	(15.78)	(15.77)
Log(age)	-0.0636***	-0.0637***	-0.0639***	-0.0639***
	(-9.88)	(-9.89)	(-9.92)	(-9.92)
R&D_asset	0.598***	0.599***	0.598***	0.599***
	(12.58)	(12.59)	(12.58)	(12.59)
PPE_asset	0.242***	0.242***	0.241***	0.242***
	(14.19)	(14.18)	(14.15)	(14.18)
Tobin's Q	0.0307***	0.0307***	0.0307***	0.0307***
	(18.19)	(18.19)	(18.20)	(18.21)
Leverage	0.265***	0.265***	0.265***	0.266***
	(13.51)	(13.51)	(13.53)	(13.55)
Cohort FE	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y
Observations	91933	91933	91933	91933
Adjusted R^2	0.121	0.120	0.121	0.121

Appendix to:

Described Patents and Knowledge Diffusion between Firms

Appendix A Variable Definition

Variable Name	Definition
Log(in-text made)	The natural log of one plus the number of in-text citations a firm makes.
Log(front-page made)	The natural log of one plus the number of front-page citations a firm makes.
Defendant litigation	An indicator variable equal to one if a firm experiences patent litigation as a defendant and zero otherwise. The data is obtained from Audit Analytics for the 1984-2020 period, and we only use category type 35, which pertains to patent law litigation cases.
Log(num defendant litigation)	The natural log of one plus the number of patent litigation a firm has as a defendant. The data is obtained from Audit Analytics for the 1984-2020 period, and we only use category type 35, which pertains to patent law litigation cases.
Log(total patents)	The natural log of one plus the number of total patents a firm is granted in a given year.
Log(total citations)	The natural log of one plus the sum of in-text and front-page citations a firm makes in a given year.
Log(asset)	The natural log of total assets.
Log(age)	The natural log of firm age.
Cash_asset	Cash and cash equivalents scaled by the total assets.
R&D_asset	R&D expenditure scaled by the total assets.
PPE_asset	Property, plant, and equipment scaled by the total assets.
ROA	Return on assets, calculated as the net income divided by the total assets.
Leverage	The debt-to-assets ratio, calculated as total debt divided by the total assets.
Tobin's Q	The Tobin's Q ratio, calculated as the market value of a company divided by the total assets.
KPSS value	Kogan et al. (2017) value of patent.
Fluidity	Product market fluidity measures the degree of competitive threat and product market change surrounding a firm (Hoberg and Phillips, 2014). The data is obtained from the Hoberg-Phillips Data Library.
In-text cited	An indicator variable equal to one if a patent receives at least one in-text citation and zero otherwise.
Geographic distance	The average distance (miles) among all combinations of cited-citing patent inventors.
Geographic distance (U.S. inventors only)	The average distance (miles) among all combinations of cited-citing patent inventors, restricting to U.S.-based inventors only.
Technology proximity	The cosine similarity calculated between the CPC class vectors of how many patents with other CPC classes cite the patents.
Age gap	The age gap between the cited and citing patent, calculated as the grant year of the citing patent minus the grant year of the cited patent.
Number of claims	The number of claims in a given patent.
Fraction of process claims	The number of process claim patents filed as a fraction of all patents filed by the firm in a given year. The process patent data is obtained from Bena et al. (2022).
Number of tech classes	The number of CPC technology classes.
Hired	An indicator variable equal to one if an inventor is hired by the citing firm at time t and zero otherwise.
Eventually hired	An indicator variable equal to one if an inventor is hired by the citing firm anytime in the future after the citing year ($s > t$) and zero otherwise.
Log(in-text)	The natural log of the average number of in-text citations an inventor receives from the citing firm over $t-2$ to t.
Log(front-page)	The natural log of the average number of front-page citations an inventor receives from the citing firm over $t-2$ to t.
Log(clean in-text)	The natural log of the average number of in-text citations an inventor receives from the citing firm, excluding in-text and front-page overlapping citations over $t-2$ to t.
Log(clean front-page)	The natural log of the average number of front-page citations an inventor receives from the citing firm, excluding in-text and front-page overlapping citations over $t-2$ to t.
Female inventor	An indicator variable equal to one if an inventor is female and zero otherwise.
IDD	An indicator variable equal to one if the citing firm resides in the state that has adopted the Inevitable Disclosure Doctrine (IDD) and zero otherwise. The IDD data is obtained from Sanati (2025).
High geographic proximity	An indicator variable equal to one if the geographic distance between the citing and cited patents are below the sample median and zero otherwise.
High tech proximity	An indicator variable equal to one if the technology distance between the citing and cited patents are below the sample median and zero otherwise.

Net sentiment	The average differences between the positive and negative sentiment scores across an inventor's in-text paragraphs of cited patents by the citing firm over $t-2$ to t. The sentiment score is calculated as described in Section 2 and only for in-text citations.
High net sentiment	An indicator variable equal to one if the average net sentiment over $t-2$ to t is above the sample median and zero otherwise.