Creative Construction: Knowledge Sharing and Cooperation Between Firms*

Evgenii Fadeev†

October 18, 2023
Click here for the latest version

Abstract

Knowledge spillovers are often measured using patent citations and are commonly assumed to diffuse across a diverse range of firms. I show that citations are highly concentrated and primarily come from business partners. I provide empirical evidence suggesting that, instead of spillovers, concentrated citations reflect intentional sharing of trade secrets between collaborating firms. The concentration of citations has increased since 2000, especially in technologies more exposed to import competition from China. This rise can be explained by a decrease in intentional knowledge sharing between partners, potentially in response to higher risks of trade secret misappropriation.

*The previous title of the paper was "Creative Construction: Knowledge Sharing in Production Networks." I am immensely grateful to my advisors Pol Antràs, Oliver Hart, Josh Lerner, and Andrei Shleifer for their guidance and support. I have been fortunate to benefit from insightful discussions with Ashish Arora, Sharon Belenzon, Jordi Blanes i Vidal, Maxim Boycko, Gabriel Chodorow-Reich, Wesley Cohen, Emmanuel Farhi, Edward Glaeser, Daniel Gross, Gordon Hanson, Elhanan Helpman, Sergei Izmalkov, Myrto Kaloutpsidi, William Kerr, Kristina Komissarova, Josh Krieger, Robin Lee, Marc Melitz, Tom Nicholas, Ariel Pakes, Gaétan de Rassenfosse, Jesse Shapiro, Ludwig Straub, Hosny Zoabi, and the participants in the NBER SI Innovation Session. I also thank John Squires and Christopher Tehrani for discussions about the legal issues surrounding patents and trade secrets.

†Duke University, Fuqua School of Business, email: evgenii.fadeev@duke.edu.

1 Introduction

Knowledge flows are key for economic growth (Romer 1990; Grossman & Helpman 1991; Aghion & Howitt 1992). They are often assumed to take a form of unintentional spillovers "in the air" (Marshall 1920) rather than intentional knowledge sharing. The presence of knowledge spillovers is a common justification for R&D subsidies (Bloom et al. 2019). The decline in knowledge diffusion is one of the explanations for the rising market concentration and slowdown in business dynamism in the U.S. (Akcigit & Ates 2022).

This paper shows evidence suggesting that knowledge does not flow through ether but through pipes of business relations. I argue that firms have significant control over the knowledge they generate, selectively sharing it with a limited set of business partners, such as input suppliers and customers. Intentional knowledge sharing has substantially declined over time due to changes in incentives for collaboration between partners.

Table 1: Example of a patent with a high concentration of citations

Patent Number	Assignee	Total Number of Citations	% of Citations from Amkor Technology Inc
5877043	IBM	218 top 0.005%	94%

A prevailing measure of knowledge spillovers is patent citations.[1] The existence of patents is often justified by the claim that they promote knowledge diffusion through the disclosure of inventions.[2] I show that the distribution of citations across firms is highly concentrated, raising a question about the role of patents in the diffusion of knowledge. For example, IBM's patent in Table 1 is heavily cited, but almost all of its citations come from IBM's *input supplier*, Amkor Technology Inc. Figure 1 shows that citations are highly concentrated in general: the most cited patents granted in the U.S. between 1980 and 2000 received around 50% of citations from one firm only, and this concentration increased to 77% in 2014. This pattern is not driven by citations from patent examiners or lawyers and is robust to various specifications.

The high concentration of citations is puzzling because valuable technologies disclosed in public patent files would be expected to generate spillovers across a broader set of firms (Romer 1990). Thus, interpreting patent citations as a measure of knowledge spillovers might

[1] For example, patent citations are used to measure knowledge spillovers in order to discipline growth models (Caballero & Jaffe 1993; Eeckhout & Jovanovic 2002; Akcigit & Kerr 2018), to evaluate the localization of spillovers in space (Jaffe et al. 1993; Thompson & Fox-Kean 2005; Singh & Marx 2013), and to identify high-quality technologies (Aghion et al. 2023a; Akcigit et al. 2021; Moretti 2021).

[2] "[T]he patent system represents a carefully crafted bargain that encourages both the creation and the public disclosure of new and useful advances in technology, in return for an exclusive monopoly for a limited period of time" (*Pfaff v. Wells Elecs., Inc.*, 525 U.S. 55, 63, 1998). See also Mazzoleni & Nelson (1998).

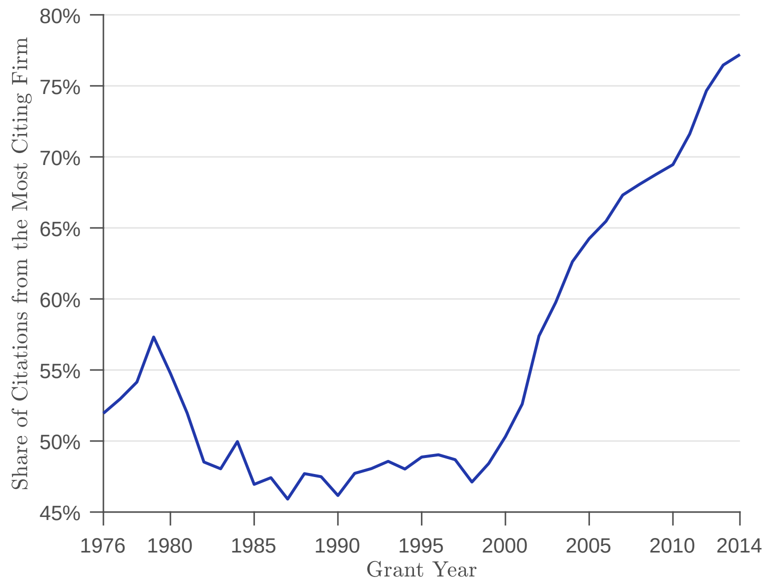


Figure 1: Concentration of Citations

This figure shows the average concentration of citations for the most cited patents between 1976 and 2014. In each grant year and technology class for the period of 1976–2014, I track citations within a five-year window for the top 1% of the most cited patents. For each cited patent, the concentration is defined as the share of citations coming from the most citing firm. The technological classes are defined at the group level in the cooperative patent classification system. To construct the aggregate measure, I take the average concentration across patents within each class and then the average across classes weighted by the number of patents.

be incorrect. Instead, I provide evidence supporting the view that citations reflect cooperation and intentional sharing of trade secrets between business partners.

I collect data on various types of inter-firm relations between publicly listed U.S. companies. I show that business partners — such as suppliers, customers, or firms with research collaboration — account for approximately 76% of inter-firm citations. Changes in citation patterns among business partners explain around 84% of the rise in the concentration since 2000. Specifically, the average number of partners citing a typical firm declined substantially, even though the overall number of partners for a firm increased. In addition, the distribution of citations within citing partners became much more skewed over time. These changes are explained by increasing differences across firms in their citation probabilities rather than by shifts in the distribution of patent counts. For instance, the increase in the concentration of citations is not driven by a rise in superstar firms in terms of the number of patents.

I argue that technologies often consist of multiple components of complementary knowledge (Anton et al. 2006). Some components, such as reverse-engineerable knowledge, are patented; other components, such as tacit knowledge, are kept secret (Hall et al. 2014). For instance, the debates on waiving intellectual property rights for COVID-19 vaccines emphasized that, in addition to the information in patents, a successful replication of the mRNA technology requires access to the trade secrets and technical know-how about it (Price II et al. 2020).

I argue that building new technologies based on a patent is easier with access to the trade secrets accompanying this patent. Firms with such access have an advantage relative to others in creating follow-on innovations. Therefore, patent citations might reflect the sharing of trade secrets between firms. As a result, citations might be concentrated because only a limited set of firms gets access to the private knowledge of a patent owner.

Testing the connection between citations and secrets is challenging because trade secrets are not observable. I use trade secret litigation data to find patents that were likely to be bundled with trade secrets. For example, the legal case *Waymo v. Uber* was about misappropriation of trade secrets related to the light detection and ranging (LiDAR) technology for self-driving cars.[3] However, the same lawsuit also had claims regarding patent infringement, and in the legal complaint, Waymo described the complementarity between their patents and secrets. I show that patents which were involved in both patent and trade secret litigation have more concentrated citations relative to similar patents within the same firm that were involved in patent litigation only. This evidence suggests that the complementarity between a patent and trade secrets might lead to a higher concentration of citations.

Bundling of patents with trade secrets does not imply that firms can control the diffusion of secrets. For example, they might be diffused through serendipitous interactions between inventors (Buera & Lucas 2018). I show that an inventor who heavily cites a particular patent in one company significantly decreases her citations to this patent once she moves to another firm. This evidence suggests that firms might have significant control over knowledge diffused through inventors, for example, through enforcement of non-disclosure agreements and other tools of trade secret laws.

I propose several alternative theories of knowledge flows and patent citations. The theories differ in the extent to which firms can control the disclosure and diffusion of their knowledge. For example, the theory of knowledge spillovers assumes full knowledge disclosure in patents, and firms cannot control its diffusion. This theory provides a reasonable approximation to citations until the 1990s. However, after 2000, the theory of intentional trade secret sharing provides the best explanation for patent citations.

I argue that an increase in the risk of trade secret misappropriation could be the reason behind the decline in knowledge sharing between business partners since 2000. Specifically, firms face the following trade-off. On the one hand, they have incentives to share secrets with certain partners, such as input suppliers. On the other hand, the more partners get access to these trade secrets, the harder it becomes to control their diffusion. As the risk of misappropriation

[3] *Waymo LLC v. Uber Technologies, Inc.*, No. 17-2235 (Fed. Cir. 2017).

rises, firms become more selective about which partners can access their trade secrets.

One of the factors influencing the risk of trade secret misappropriation could be an increase in trade with China. According to the U.S. Counterintelligence Office, "[t]he pace of foreign economic collection and industrial espionage activities against major US corporations and US Government agencies is accelerating."[4] In addition, the enforcement of U.S. trade secret laws tends to be less effective in cases involving international misappropriation (Almeling 2012). I show that the rise in the concentration of citations in Figure 1 is greater in the technologies more exposed to import competition from China, where U.S. imports are instrumented by Chinese exports to other high-income countries (Autor et al. 2020).

I also discuss additional factors that could increase the risks of trade secret misappropriation. For instance, advances in IT and the internet have reduced the costs of information storage and remote access, potentially making it easier to misappropriate trade secrets. In line with this argument, Hoberg et al. (2021) use staggered internet rollout in China to show that U.S. firms have increased their complaints about intellectual property theft as information access costs for Chinese firms have decreased.

Finally, I provide recommendations for the use of patent citations in the literature. I suggest taking into account the relationships between firms citing each other in empirical studies of knowledge spillovers.

Literature

Much of the literature in economics is based on the concept of unintended knowledge spillovers, assuming that a firm cannot control the diffusion of knowledge it has generated.[5] Knowledge spillovers are key in models of endogenous growth (Jones 2005; Aghion et al. 2014) and knowledge diffusion (Buera & Lucas 2018). The presence of spillovers is a common justification for government intervention in the economy (Bloom et al. 2019) and is used to explain the agglomeration of economic activity (Marshall 1920; Carlino & Kerr 2015).

Patent citations have been the prevailing measure of knowledge spillovers since the seminal work by Jaffe et al. (1993). Several surveys of inventors and firms confirm that citations are correlated with knowledge flows (e.g., Jaffe et al. 2000; Duguet & MacGarvie 2005). While these flows might represent intentional knowledge sharing, patent citations are commonly interpreted as unintended knowledge spillovers, for example, in the growth literature (e.g., Caballero & Jaffe 1993; Akcigit & Kerr 2018) and urban studies (e.g., Ellison et al. 2010; Singh & Marx 2013).

[4] Office of the National Counterintelligence Executive, "Foreign Spies Stealing US Economic Secrets in Cyberspace", 2011.

[5] "[I]nventors are free to spend time studying the patent application for the widget and learn knowledge that helps in the design of a wodget. The inventor of the widget has no ability to stop the inventor of a wodget from learning from the design of a widget" (Romer 1990, p. 84).

This paper highlights the importance of cooperation with intentional knowledge sharing. In contrast to unintended knowledge spillovers, intentional knowledge flows depend on the specifics of the economic environment, including aspects such as the legal protection of trade secrets, market structure, and the duration of relationships between partners. I argue that the increasing risks of trade secret misappropriation might explain the decline in knowledge flows between firms. This decline may be crucial for understanding recent trends in the U.S., such as the rising market concentration and declining business dynamism (Akcigit & Ates 2022).

I provide evidence suggesting that citations are unlikely to measure spillovers, but they still provide valuable information about cooperation between firms. These results give a new interpretation to some of the findings in the literature on knowledge spillovers. For instance, the spatial localization of citations (Jaffe et al. 1993) is commonly interpreted as evidence of the importance of proximity and serendipitous face-to-face interactions in knowledge diffusion (Carlino & Kerr 2015). However, the results in this paper suggest that such diffusion is either not serendipitous or that the localization of citations is a consequence of the co-location of business partners for reasons unrelated to spillovers (Ellison et al. 2010).

This paper is also related to the literature on market-mediated knowledge flows (e.g., Arora et al. 2001; Arqué-Castells & Spulber 2022). In line with this literature, I argue that a lot of knowledge flows between firms are intentional rather than the results of spillovers. In contrast to this literature, I argue that patent citations might be better explained by the presence of tacit knowledge rather than by formal licensing agreements.

Even without formal contracts on knowledge sharing, firms producing complementary products — such as partners in a supply chain — still have incentives to share knowledge with each other. The importance of vertical knowledge flows is highlighted in the literature on R&D cooperation (e.g., Cassiman & Veughelers 2002) and foreign direct investment (e.g., Alfaro-Ureña et al. 2022; Bai et al. 2022). Knowledge is often diffused through common business partners. Rosenberg (1963) emphasizes the importance of input suppliers in the diffusion of knowledge across industries. The trade-off between the benefits of knowledge sharing and costs of knowledge leakage is highlighted in the networks literature (e.g., Aghion et al. 2023b; Dasaratha 2023). I argue that firms reduced knowledge sharing with partners due to higher risks of leakage.

This paper is also related to the literature on intellectual property protection (e.g., Png 2017; Hall et al. 2014). Much of this literature treats patenting and secrecy as substitutes. I argue that the combination of patents and trade secrets might be used to protect the same technology. This idea is in line with surveys of firms (e.g., Cohen et al. 2000), the management literature (e.g., Amara et al. 2008), legal research (e.g., Jorda 2008), and case studies on intellectual property protection in the chemical and pharmaceutical industries (e.g., Arora 1997; Price II et al. 2020). Anton et al. (2006) argue that due to weak patent protection, "a combination of patenting and secrecy is common."

Finally, this paper contributes to the literature on trade with China. Autor et al. (2020) and Hoberg et al. (2021) show that trade with China reduced corporate patenting and R&D investments in the U.S. I complement their evidence by showing that trade with China also reduced knowledge sharing between firms, measured by the concentration of patent citations.

This paper is organized as follows. Section 2 documents that patent citations are highly concentrated and primarily come from business partners. Section 3 differentiates between possible explanations for this concentration. Section 4 discusses potential reasons behind the decline in cooperation between firms. Section 5 provides recommendations for the use of patent citations in the literature. Section 6 concludes the paper.

2 Concentration of Patent Citations and the Role of Business Partners

This section documents that patent citations are highly concentrated and primarily come from business partners. Section 2.1 provides background on the U.S. patent system and describes the data. Section 2.2 documents that, even for the most cited patents, the majority of citations come from one firm only, and this concentration has significantly increased since 2000. I also document multiple additional facts and robustness checks. In Section 2.3, I evaluate the role of business partners in the concentration of citations.

2.1 Data and Background

Patents are supposed to facilitate knowledge diffusion through the disclosure of information in the award. The U.S. Supreme Court has stated that patent disclosures "will stimulate ideas and the eventual development of further significant advances in the art" and that these "additions to the general store of knowledge are of such importance" that they are worth "the high price of . . . exclusive use."[6]

Patents consist of two parts: a written description of an invention, including citations to prior art (patents, publications, etc.), and claims defining the boundaries of intellectual property rights. To be patentable, an invention must be patent-eligible, useful, novel, and non-obvious. Additionally, the text of the application should satisfy the disclosure requirements.[7] Patent

[6] *Kewanee Oil Co. v. Bicron Corp.*, 416 U.S. 470, 481 (1974) and Ouellette (2012).

[7] These requirements are governed by the US Code, Title 35, sections 101, 102, 103, and 112. For a review, see Scotchmer (2004), ch. 3.

examiners use references to prior art to check whether the invention is novel and non-obvious. In the U.S., applicants have a "duty of candor" to disclose relevant prior art that they are aware of, and failure to do so can lead to patent invalidation. In general, prior art is used to strengthen, narrow, or reject certain claims. Therefore, citations serve the legal function of delimiting intellectual property rights on an invention.

I use the data on utility patents granted by the U.S. Patent and Trademark Office (USPTO) for the period from 1976–2019. Most granted patents contain information about assignees (patent owners). I clean assignee names to group patents by firms, individual inventors, universities, and other organizations. Autor et al. (2020) provide a matching of patent assignees to names of publicly traded firms in the Compustat data set. I extend their matching for the additional years of 2015–2019 and for private firms. Details are given in Appendix B.1.

In Section 2.3, I use three data sets to find the types of relationships between cited and citing firms. First, I use the FactSet Revere Supply Chain Relationships data set, which is based on public sources such as filings with the U.S. Securities and Exchange Commission (SEC), investor presentations, and press releases to collect data on business relations between firms. The data list partners such as suppliers and customers, firms with licensing agreements, research collaborations, joint product offerings, and firms with ownership stakes (e.g., joint ventures). Second, I use the Compustat Segments to collect data on supplier-customer relationships between firms. Finally, I use the USPTO data on patent re-assignment to find firms that are trading patents with each other. The data sets cover the period from 2003 to 2022.[8] I match all data sets with patents using company names. Details are given in Appendix B.1.

2.2 High Concentration of Patent Citations

For each year and technological class in the period from 1976–2014, I track citations within a five-year window for the top 1% of the most cited granted patents. Below I show that the results are robust to other thresholds, such as top 5% or top 10%. There are two reasons to focus on the set of the most cited patents. First, the value of patents is highly skewed, with many being of no value to the firm, so the empirical literature on innovation is often focused on the most cited patents (e.g., Aghion et al. 2023a). This focus is justified by the positive correlation between the number of patent citations and the firm's stock market valuation (Hall et al. 2005; Kogan et al. 2017). Second, by construction these patents are expected to generate most follow-on innovations and knowledge flows, providing a lower bound on the concentration

[8]FactSet covers most relationships between firms, though its coverage starts only from 2003. In general, Compustat Segments and USPTO re-assignment data provide coverage since 1976 and 1968, respectively.

measure. Taking a five-year window controls for the truncation bias that older patents have more time to accumulate citations. Comparing patents within each technological class controls for differences across classes in citation patterns (Lerner & Seru 2022). To classify technologies, I use the group level of the Cooperative Patent Classification (CPC) system, which has 672 groups.

For each patent among the most cited ones, I compute the distribution of citations across different organizations. The variable $n_{k,i}$ denotes the number of citations from organization i to patent k. Organizations are mostly firms but also include non-corporate entities, such as government agencies and universities. I exclude citations from individual inventors and patents with missing assignee information. The concentration measure for patent k is the share of citations coming from the most citing organization:[9]

$$\mathcal{C}_k = \max_i \left\{ \frac{n_{k,i}}{n_k} \right\} \tag{2.1}$$

where n_k is the total number of citations patent k receives. The most citing organizations are predominantly corporations, so I will use the terms "firms" and "organizations" interchangeably. To construct an aggregate measure, for each year I take the average of patents' concentration measures within each technological class and then the average across technological classes weighted by the number of patents in a class. Appendix B.2 provides more details.

Figure 1 on page 2 shows the resulting aggregate concentration measure. On average, a patent (among the most cited ones) granted between 1976 and 2000 received around 50% of citations from one firm only. This concentration has significantly increased since 2000: a patent granted in 2014 received around 77% of citations from one firm only.

Figures C1 and C2 in Appendix C provide additional results on the concentration. Figure C1 shows that the increase in the concentration is primarily driven by changes within technological classes rather than the rise of technologies with high concentrations of citations. An increase in the concentration after 2000 is observed in 87% of classes. Panel (a) of Figure C2 shows that the average number of citations has significantly increased over time: the most cited patents granted in 2014 received 11 times more citations within five years from the grant day than patents granted in 1976. Panel (b) shows that more cited patents have a higher concentration of citations, and that this relationship is driven by patents granted after 2000.[10] Therefore, the increase in the concentration of citations is not driven by the decline in the number of citations.

[9] In Section 2.3, I also consider the Herfindahl-Hirschman Index (HHI) for the distribution of citations across citing firms.

[10] The relationship holds for patents with more than seven citations. For patents with fewer than seven citations, the concentration is high due to a low number of citations. See Section 3.3 for more details.

Figures C3 – C5 provide multiple robustness checks. In particular, the results are robust to different thresholds for the most cited patents (top 5% and 10%). The concentration is not driven by superstar firms in patenting, and it is robust to the exclusion of firms' self-citations to themselves. The concentration is also robust when I group citations of patents from the same within-country family (continuations, continuations-in-part, and divisionals) as a single citation, indicating that its rise is not driven by increasing patent families. Citation patterns might be affected by patent examiners (Alcácer et al. 2009). I show that the concentration of citations from patent examiners is around two times lower than the concentration based on citations from non-examiners. Citations might also be affected by patent lawyers. In Section 3.4, I use the movement of inventors across companies to show that the concentration of citations is driven by firms rather than inventors. Using the same technique, I show that the concentration of citations is not driven by patent lawyers either. Appendix B.3 describes more robustness checks, including controls for outliers, different samples, and weighting schemes.

2.3 The Role of Business Partners

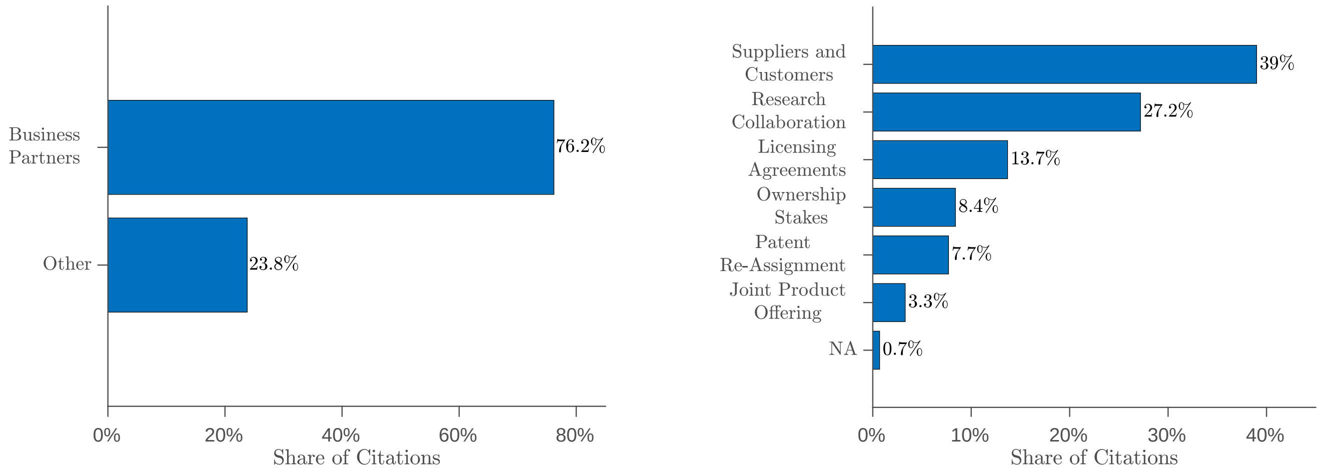
This section evaluates the role of business partners in the concentration of citations. I show that business partners account for the majority of citations between firms, and that the concentration has primarily increased due to changes within firms in how they receive citations from partners.

The data on relationships between firms are not available at the patent level, so it is preferable to redefine the concentration measure at the firm level. I propose two measures of concentration. First, for each firm in a year, I define the concentration measure as the Herfindahl-Hirschman Index (HHI) of all citations to the firm's patents granted in this year within a five-year period. The aggregate concentration in year t is defined as

$$\mathcal{H}_t = \sum_k s_{k,t} \mathcal{H}_{k,t} \tag{2.2}$$

where the summation goes across all firms receiving citations to their patents granted in year t, $s_{k,t}$ is the share of citations firm k receives in year t in the total number of citations firms receive in this year, and $\mathcal{H}_{k,t}$ is the HHI concentration of citations for firm k. Second, for each firm in a year, I define the concentration as the share of citations coming from the most citing firm, as in Section 2.2. The results are robust to both measures, but the additive structure of the HHI measure will be useful for evaluating the role of partners in the concentration, see equation (2.3).

Figure C6 in Appendix C shows that the firm-level concentration follows the same pattern as in Figure 1: a stable or slightly declining concentration of citations until 2000 and a significant rise after. This result is robust to different sample selections and concentration measures. Figure C7 decomposes the increase in the concentration from 2001 to 2014 based on



(a) Citations by Relationship Between Firms

(b) Citations Within Partners

Figure 2: Distribution of Citations Based on Relationship Between Firms

This figure shows the distribution of patent citations across different types of relationships between cited and citing firms. Panel (a) shows the share of citations coming from business partners. Panel (b) shows the distribution of citations coming from business partners across different types of partners.

the methodology of Melitz & Polanec (2015). The rise in the concentration is primarily explained by the increasing concentration of citations within firms, rather than the re-allocation of citations across firms or the exit and entry of new firms out of and into patenting.

To document the role of partners in patent citations, I focus on inter-firm citations between U.S. publicly listed firms for patents granted after 2001. The sample selection is driven by data limitations.[11] I discuss how these limitations might affect the results at the end of this section. Following Section 2.2, I consider patents granted until 2014 and trace citations within a five-year window from a grant date.

I compute the distribution of citations across firms based on their relationship with a cited firm. I define firms to be business partners if they had at least one of the following relationships between 2003 and 2022:[12] suppliers or customers, research collaborations, licensing agreements, joint product offering, patent re-assignment, joint ownership stakes (e.g., a joint venture), and partners with an uncertain relationship.[13]

[11]The data on relationships cover the period from 2003 to 2022. I include the years 2001 and 2002 to analyze the rise in the concentration since 2001, but the results are similar without these years. The coverage of relationships is limited for private and foreign firms.

[12]Around 75% of citations between partners occur during the period of a reported relationship. The data are based on information disclosed by firms in public sources, but the disclosed dates often do not correspond to the actual dates of a contractual relationship between companies. Firms might delay the disclosure of a relationship relative to its actual start and might stop reporting it before it actually ends. Therefore, 75% provides a lower bound on the share of citations occurring during a business relationship.

[13]Table C1 in Appendix C provides the full list of all business relationships. FactSet defines some partners

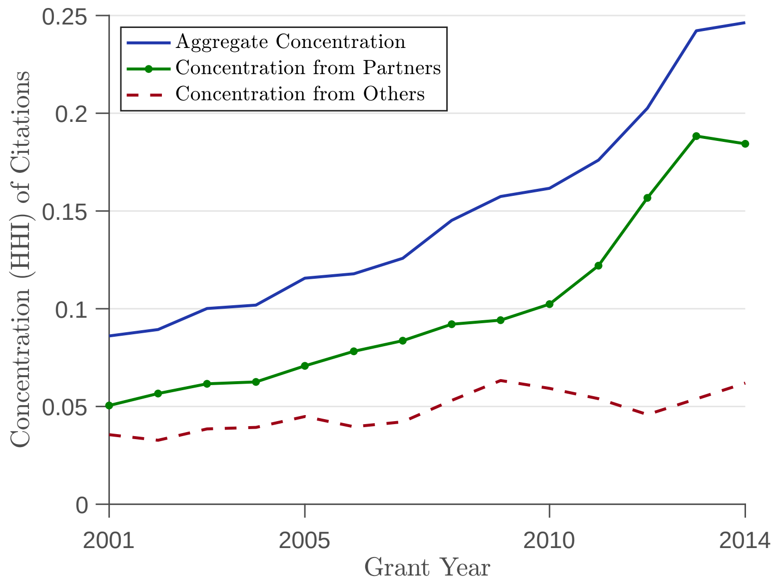


Figure 3: Decomposition of the Concentration Based on Relationship Between Firms

This figure decomposes the concentration into the roles of partners and other firms, see equation (2.3).

Panel (a) of Figure 2 shows that 76% of citations occur between business partners. Panel (b) shows the distribution of citations among partners based on a relationship between firms. If two firms have multiple relationships, I divide citations evenly across them. Firms with a supplier-customer relation and research collaboration account for the majority of citations: 39% and 27%, respectively. Other types of partners account for 34% of citations and include firms that signed licensing agreements, re-assigned patents, had joint ownership stakes, and had joint product offerings.

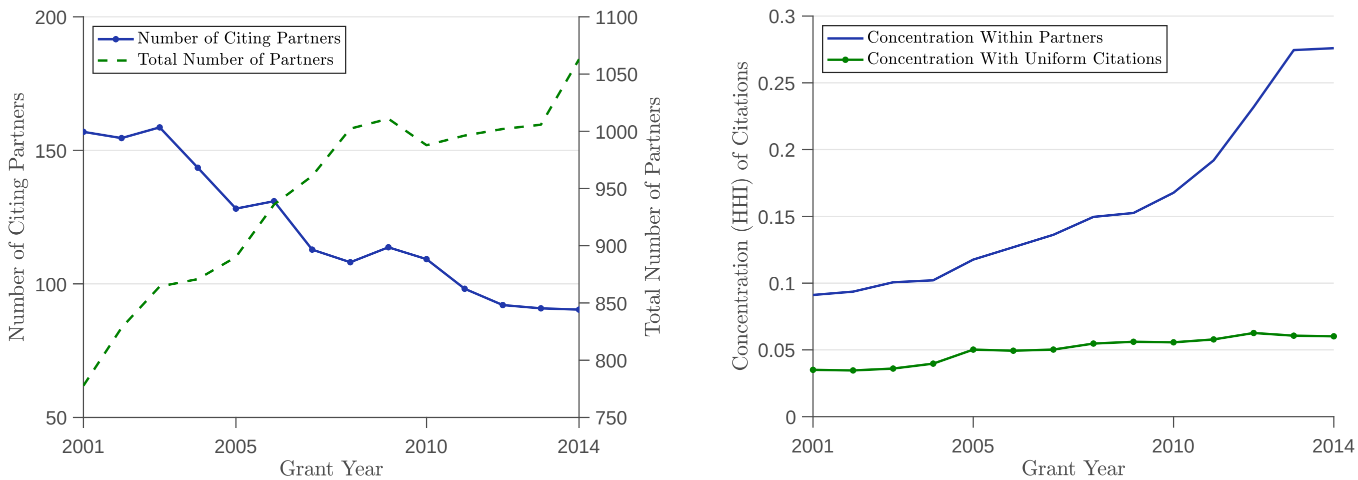
To evaluate the importance of partners in the concentration of citations, I decompose the HHI concentration within a firm into citations from partners and other firms.

$$\mathcal{H}_{k,t} = \sum_i \left(\frac{n_{k,i,t}}{n_{k,t}}\right)^2 = \mathcal{H}^{\mathcal{P}}_{k,t} + \mathcal{H}^{\mathcal{O}}_{k,t} = \underbrace{\sum_{i \in \mathcal{P}_k} \left(\frac{n_{k,i,t}}{n_{k,t}}\right)^2}_{\text{Partners },\mathcal{H}^{\mathcal{P}}_{k,t}} + \underbrace{\sum_{i \in \mathcal{O}_k} \left(\frac{n_{k,i,t}}{n_{k,t}}\right)^2}_{\text{Other },\mathcal{H}^{\mathcal{O}}_{k,t}} \tag{2.3}$$

where $n_{k,i,t}$ is the number of citations from firm i to firm k for patents granted in year t, $n_{k,t}$ is the total number of citations firm k receives to patents granted in year t, $i \in \mathcal{P}_k$ means that firm i was a business partner to firm k, and $i \in \mathcal{O}_k$ means that firm i was not a (revealed) business partner to firm k.

Figure 3 shows the decomposition into partners and other firms from equation (2.3). On

without clarifying the nature of the relationship. These account for only 0.7% of citations between partners.



(a) Extensive Margin

(b) Intensive Margin

Figure 4: The Concentration of Citations Within Partners: Extensive and Intensive Margins

This figure shows the roles of extensive and intensive margins in the concentration of citations within partners. Panel (a) compares the average number of citing partners relative to the total number of partners for a typical firm. For each firm, the number of partners is weighted by the number of citations this firm receives. The total number of partners of a firm in a given year is defined using partners reported within five years of this year. Panel (b) shows the average concentration of citations (HHI) from business partners. The dotted line shows the counterfactual concentration, in which citations are distributed across firms as uniformly as possible to minimize the HHI, given the existing number of citing firms. Both graphs are constructed based on a sample of patents with a unique assignee. Patents with more than one assignee account for less than 1% of citations.

average, partners explain around 66% of the concentration in citations: 59% in 2001 and 75% in 2014. Changes in citations among partners account for 84% of the rise in the concentration. Suppliers, customers, and firms with research collaboration account for 60% of the rise in the concentration. Figure C8 in Appendix C offers more details.

To provide further evidence, I also analyze citations by industries of the cited and citing firms. Figure C9 in Appendix C shows that firms within the same industry account for only 8% of the increase in the concentration of citations. Therefore, the rise in the concentration is largely driven by changes in citations among firms that are unlikely to be direct competitors.[14]

The concentration of citations among partners is determined by both the number of citing partners (an extensive margin) and the distribution of citations within them (an intensive margin). Panel (a) of Figure 4 shows that, from 2001 to 2014, the average number of citing partners for a typical firm declined by 42%, even though the number of (reported) partners increased over time.[15]

[14]In general, defining competitors requires a definition of the market, and industry affiliation might be an imperfect proxy for competitors. Nevertheless, firms from the same industry are more likely to compete with each other than with firms from different industries. Industries are defined at the four-digit level in Standard Industry Classification.

[15]This result is robust to restricting the sample to partners with patents.

Panel (b) evaluates the relative importance of the intensive and extensive margins in the concentration of citations. Specifically, I quantify the importance of the extensive margin by allocating citations uniformly among partners, subject to the integer constraint on the number of citations. For example, consider a firm receiving 10 citations from three partners with a citation distribution of $(8,1,1)$; the HHI is $\left(\frac{8}{10}\right)^2 + 2\cdot\left(\frac{1}{10}\right)^2 = 0.66$. The uniform distribution would be $(4,3,3)$ with $\text{HHI}_u = \left(\frac{4}{10}\right)^2 + 2\cdot\left(\frac{3}{10}\right)^2 = 0.34$. The difference $\text{HHI} - \text{HHI}_u$ represents the intensive margin. Panel (b) shows that on average the intensive margin explains 62% ($1 - \frac{\text{HHI}_u}{\text{HHI}} = 0.62$) of the concentration in 2001 and 78% in 2014. Changes in the extensive margin over time explain only 14% of the rise in the concentration of citations among business partners.

To sum up, the evidence in this section questions whether citations reflect unintentional knowledge spillovers. In Section 3, I provide additional evidence suggesting that citations are likely to reflect cooperation and intentional knowledge sharing between companies.

Discussion of Data Limitations

The data on inter-firm relations have some limitations because they are primarily based on the disclosure of relationships by publicly traded companies. Therefore, the data do not cover relations between all firms. As a robustness check, I focus on a sample of citations in which at least one of the firms is a U.S. publicly traded company, but not necessarily both. Figure C10 in Appendix C shows that partners explain 51% of citations, and the distribution within partners is similar to Panel (b) in Figure 2. The share of partners is lower in this sample, but it might reflect the incompleteness of the data for private or foreign firms, rather than the absence of a relationship. Moreover, Figure C6 in Appendix C shows that the concentration dynamics are consistent whether considering all firms or just U.S. publicly traded companies. Therefore, I expect a minimal bias regarding the role of partners in the rise of the concentration when focusing on the sample of U.S. publicly traded companies.

Another limitation of the data is related to firms' incentives to disclose relationships with other firms. Although regulation SFAS No. 131 requires publicly traded companies to report the identity of any customer representing more than 10% of their total sales, smaller customers or other types of relationships are self-reported. Firms' incentives to disclose a relationship with another firm might differ across types of relationships. For example, firms might have stronger incentives to conceal the identity of research collaborators relative to established input suppliers. In addition, the data lack details of contracting between firms, and certain types of relationships are not mutually exclusive. For instance, firms with a research collaboration might not report that they also have licensing agreements as part of the collaboration. Therefore, the comparison of citation patterns based on the type of partnership should be interpreted with care.

Table 2: Theories of Patent Citations and Concentration

		Presence of Trade Secrets (Tacit Knowledge)?	
		No	Yes
Intentional (Controlled) Knowledge Flows?	No	Spillovers and Specialization	Interactions Between Inventors
	Yes	Licensing	Intentional Sharing of Trade Secrets

Notes: This table classifies the four theories of patent citations described in Section 3.1.

3 What Do Patent Citations Mean?

In this section, I provide additional evidence on the concentration of citations and differentiate between possible explanations. Section 3.1 offers four potential theories of patent citations. Section 3.2 describes the data. Section 3.3 shows that the concentration of citations should be explained based on the difference across firms in their citation behavior, rather than on differences in the number of patents. In Section 3.4, I use the movement of inventors across firms to show that the concentration is primarily driven by firms rather than inventors. Section 3.5 shows that patents bundled with trade secrets have more concentrated citations. Finally, Section 3.6 differentiates between the theories of patent citations based on the documented evidence.

For all statistics on the concentration of citations, I also report the counterfactual statistics that would be observed if citations were random. Since I observe the universe of all patents and citations in the USPTO, there is no sampling uncertainty (Abadie et al. 2020). Instead, I assume that the randomness of citations provides a basis for inference. The details are given in Section 3.3.

3.1 Potential Explanations for the Concentration of Citations

In this section, I propose four potential explanations for the concentration of patent citations. All of these explanations assume that citations reflect knowledge flows, but they can be divided based on two criteria (see Table 2): First, do firms disclose all relevant knowledge in patent files, or do they leave some (tacit) knowledge private in the form of trade secrets? Second, to what extent can firms control the use of their knowledge by other firms?

Explanation 1. ***Spillovers and Specialization.*** *Knowledge in patents is available to everybody, and patent citations reflect knowledge spillovers. Citations are concentrated because* *knowledge spillovers occur within narrow technologies, and only a few firms benefit from a particular patent.*

Patent citations as a measure of knowledge spillovers are commonly used in the growth literature (e.g., Akcigit & Kerr 2018). Explanation 1 is based on the assumptions that firms fully disclose their knowledge in patent applications and that it is available to everybody for follow-on inventions (e.g., Romer 1990, p. 84). The concentration of citations could be explained by the specialization in patenting. Suppose that all firms have equal access to the knowledge disclosed in the cited patent. However, this knowledge is valuable only within a narrow technological space, and only a few firms specialize (patent) in this space. The rise in the concentration could be explained by the increasing specialization within technologies.

In Section 3.3, I show that the concentration of citations is observed even within narrow technological classes. For instance, Amkor is responsible for the majority of citations to IBM's patent in Table 1 not because Amkor has more patents in its technological space than other companies (specialization), but because a large share of its patents makes citations to IBM, while other companies with similar patents make no citations to IBM's patent. Moreover, I show that the explanatory power of the "specialization theory" has significantly declined over time, and that this theory cannot explain the rise in the concentration of citations.

Explanation 2. *Interactions Between Inventors.* *Citations reflect communication between inventors. They are concentrated because of limited interactions between inventors from different firms.*

Patent citations as a measure of communication between inventors are commonly used in the urban literature (e.g., Jaffe et al. 1993). According to this theory, patents are surrounded by tacit knowledge, and interactions between inventors facilitate the diffusion of knowledge. The presence of tacit knowledge and the need for the geographical co-location of inventors is one of the theories for the agglomeration of economic activity (Marshall 1920; Carlino & Kerr 2015).

The key assumption of Explanation 2 is that firms have limited control over knowledge diffused through employees. In Section 3.4, I trace citations of inventors who filed similar patents in multiple companies to separate the roles of firms and inventors in citation patterns. I show that inventors significantly change their citation probabilities to a particular patent once they move to another company, and that the concentration of citations is primarily explained by firm-specific factors. This evidence contradicts Explanation 2 and suggests that firms might have significant control over knowledge diffused through employees, at least for knowledge flows measured by patent citations. For example, firms can control employees' communication with others through the enforcement of non-disclosure agreements and other tools of trade secret laws.

The last two explanations are based on cooperation between firms.

Explanation 3. ***Licensing.*** *A firm is more likely to cite a patent if it has a licensing agreement with the patent owner. Citations are concentrated because the patent owner licenses its patent to a small number of firms.*[16]

Patent citations as a proxy for licensing agreements are used in the literature on patent thickets (e.g., Ziedonis 2004). Explanation 3 has the following interpretation: Patents that are close in a technological space are more likely to cite each other, but they are also more likely to have overlapping claims. Given that patents have significant "uncertainty about the validity and scope of the legal rights being granted" (Lemley & Shapiro 2005), firms might strategically avoid technological areas already crowded by other companies unless they have licensing agreements with these companies. Therefore, firms with licensing agreements are more likely to file patents with overlapping claims and make citations to each other.

In Section 2.3, I show that inter-firm citations primarily occur between business partners, and the rise in the concentration is driven by changes in citation patterns among partners. The significant role of partners in citations supports the view that citations reflect *intentional* knowledge flows between firms rather than unintentional spillovers. However, among business partners, firms with formal licensing agreements can explain only around 14% of citations and around 11% of the rise in the concentration. In other words, a lot of citations occur between business partners without (observable) explicit contracts governing knowledge sharing. In the last explanation, I propose a theory of citations that does rely on formal licensing contracts.

Explanation 4. ***Intentional Sharing of Trade Secrets.*** *Firms do not disclose all knowledge relevant to a technology in patent files. Instead, they keep it secret. Citations are correlated with the sharing of trade secrets accompanying patents. Furthermore, citations are concentrated because only a limited set of firms gets access to private knowledge of a patent owner.*

Much of the literature on intellectual property (IP) protection treats patenting and secrecy as substitutes (e.g., Hall et al. 2014). I argue that technologies often consist of multiple pieces of complementary knowledge. Within each piece, the choice between patenting and secrecy might be mutually exclusive, but firms can choose to patent only some parts of the knowledge and keep the rest secret. For example, firms prefer patenting for knowledge that is codified and can be reverse-engineered, and secrecy for knowledge that is tacit and easier to hide (Hall

[16] In the context of this paper, "licensing" refers to "pure" patent licensing, meaning the legal capability of firms to prohibit the use of their technologies without a licensing agreement. In practice, patent licensing is often accompanied by the sharing of trade secrets (Arora 1995; Zuniga & Guellec 2009). The role of trade secret sharing is discussed in Explanation 4.

et al. 2014). The complementarity between different parts of knowledge makes it difficult to replicate and build on a certain technology without access to the knowledge that is kept secret.[17] Therefore, firms can use the benefits of patents without full disclosure of the technology to the public, which is consistent with the arguments of legal scholars that the patent system fails in its disclosure function (e.g., Roin 2005). The idea of complementarity between patenting and secrecy is in line with surveys of firms (e.g., Cohen et al. 2000), the management literature (e.g., Amara et al. 2008), legal research (e.g., Jorda 2008), and case studies on IP protection in the chemical and pharmaceutical industries (e.g., Arora 1997; Price II et al. 2020).

To the best of my knowledge, this paper is the first one to make the explicit connection between patent–secrecy complementarity and citations. However, testing this connection is challenging because trade secrets are not observable. In Section 3.5, I use patents involved in trade secret litigation to provide suggestive evidence for the connection between secrets and citations. I show that patents involved in trade secret litigation have more concentrated citations relative to similar patents within the same firm that were involved in patent litigation only.

3.2 Data

Sections 3.3 and 3.4 rely on the USPTO patent data only. In Section 3.5, I use Lex Machina data on patent litigation. Lex Machina complements the USPTO data on patent litigation with information on whether the patents were involved in trade secret litigation. Details are given in Appendix B.1.

3.3 Narrow Technological Paths

The "Specialization" theory (Explanation 1) argues that the concentration of citations is driven by the concentration in patenting. According to this theory, Amkor is responsible for 94% of citations to IBM's patent in Table 1 because Amkor has more patents than other companies in a technology that benefits from IBM's patent. Formally, consider patent k that can receive citations from M firms. Firm i has $N_i > 0$ patents and makes $n_{k,i} \geqslant 0$ citations. The number of citations can be decomposed into intensive and extensive margins

$$n_{k,i} = p_{k,i} \cdot N_i \tag{3.1}$$

[17] Appendix A provides legal background and case studies on how firms combine patenting and secrecy.

where $p_{k,i} = {n_{k,i}}/{N_i}$ is the share of patents in firm i that make citations to patent k. The concentration measure for patent k is defined as

$$\mathcal{C}_k = \max_i \left\{ \frac{n_{k,i}}{\sum_{j=1}^{M} n_{k,j}} \right\} = \max_i \left\{ \frac{p_{k,i} \cdot N_i}{\sum_{j=1}^{M} p_{k,j} \cdot N_j} \right\}$$

The "Specialization" theory assumes that firms do not differ in their citation behavior ($p_{k,i} = p_{k,j}$ for $i \neq j$), but one firm dominates others in terms of the number of patents ($N_i \gg N_j$ for $j \neq i$).

To evaluate this theory, one needs to find all patents that could have potentially made citations to patent k. I divide all granted patents into disjoint groups based on common characteristics. Then, I find all patents that share similar characteristics with patents that *actually* make citations to patent k. For characteristics of patents, I choose an application year and a detailed technological class (main subgroup level in CPC). Below I also describe a robustness check based on textual similarity of patents using BERT model.[18]

For example, patents making citations to IBM's patent in Table 1 ($k = 5877043$) are divided in 68 disjoint groups based on the application year and technological class. Most citations (16 out of 218, all from Amkor) come from patents in technological class $H01L23$ and with application year 2003. Overall, there are 1465 patents with such characteristics, and only 20 of them are assigned to Amkor. Many companies have more patents than Amkor in this class and year, for instance, Intel and Micron Technology have 122 and 101 patents, respectively.

To evaluate the role of "Specialization" theory in the concentration of citations, for each patent I do the decomposition from (3.1) within all possible groups of citing patents:

$$n_{k,i}(g) = p_{k,i}(g) \cdot N_i(g)$$

where $n_{k,i}(g)$ is the number of citations from firm i to patent k from patents within group g. For example, for IBM's patent the number of citations from patents in technological class $H01L23$, with application year 2003, and assigned to Amkor is $n_{k,i}(g) = 16$, where $k = 5877043$, $i =$ Amkor, and $g = \{H01L23, 2003\}$. In this example, $N_i(g) = 20$ and $p_{k,i}(g) = {16}/{20}$ for $i =$ Amkor, and $p_{k,j}(g) = 0$ for $j \neq$ Amkor.

Next, I do two types of Monte-Carlo simulations. First, I randomize $n_{k,i}(g)$ citations across all patents with the same characteristics g. This exercise equates citation rates across firms that have patents with characteristics g ($p_{k,i}(g) = p_{k,j}(g)$ for $i \neq j$). Second, I allocate $n_{k,i}(g)$ citations randomly across firms assuming that all firms with patents in g have the same number of patents. This exercise equates both the citation rates and the number of patents across firms ($p_{k,i}(g) = p_{k,j}(g)$ and $N_i(g) = N_j(g)$ for $i \neq j$). I do these Monte-Carlo simulations for the most

[18]Bidirectional Encoder Representations from Transformers (BERT) is a family of language models developed by Google in 2018 (Devlin et al. 2019).

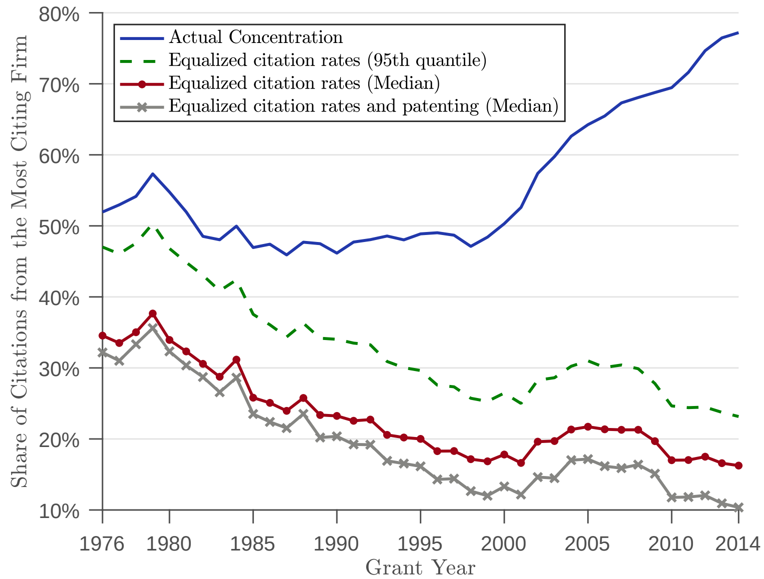


Figure 5: Decomposition of the Concentration of Citations

This figure compares the actual concentration of citations (the upper solid line) with the counterfactual ones in which citation rates are equalized across firms (the dashed and the dotted lines in the middle), and in which both citation rates and the number of patents are equalized across firms (the crossed line at the bottom). These counterfactual concentration measures are constructed using Monte-Carlo simulations in which citations are allocated randomly across observationally similar patents (equalized citation rates) and across firms with observationally similar patents (equalized citation rates and patenting). The details are given in Appendix B.4.

cited patents, and then I recompute the aggregate concentration measure from Section 2.

Figure 5 shows the results of Monte-Carlo simulations. The upper solid line shows the actual aggregate concentration ($\mathcal{AC}_t$). The dashed and the dotted lines in the middle show the 95th quantile and the median of the concentration in which citation probabilities are equalized across firms (denote by $\mathcal{RC}_t$ the median). Finally, the crossed line at the bottom shows the median concentration with equalized citation rates and equalized patenting across firms ($\mathcal{PC}_t$).[19]

The variable $\mathcal{PC}_t$ shows a part of the concentration that is driven by the number of firms. For example, with M firms the concentration cannot be lower than $^1/_M$. The measure $\mathcal{PC}_t$ has declined over time, indicating an increase in the number of firms filing patents within the same technological class over the same time period.

The difference $\mathcal{RC}_t - \mathcal{PC}_t$ quantifies the impact of variations in the number of patents among firms, conditional that each firm has at least one patent in a given (application year, technological class) pair. These variations in patenting rates account for only a minor portion of the overall concentration in citations.

Finally, the difference $\mathcal{AC}_t - \mathcal{RC}_t$ shows the role of variations in citation intensities across

[19]Figure C11 in Appendix C shows a similar decomposition for the firm-level concentration measure defined in Section 2.3.

firms. This difference has significantly increased over time, indicating growing disparities in citation behavior among firms. The difference in citation intensities across firms explains around 38% of the concentration in 1980, 50% in 1990, and 79% in 2014. Although the "Specialization" theory provides a reasonable approximation to citations until the 1990s, it is inconsistent with the observed changes in citation patterns.

The decomposition in Figure 5 might underestimate the role of specialization in patenting if technological classes are not granular enough to capture this specialization. I use the technological classification which is more granular than the one commonly used in the literature.[20]

As a robustness check, I also use a natural language processing model called BERT to find textual similarity between patents, in addition to considering application years and technological classes. For each cited patent, I measure a similarity between patents citing it. I then select all patents that exhibit at least as much similarity to the citing patents as the citing ones do among themselves. The details are given in Appendix B.4. Figure C12 in Appendix C shows that the results are robust with this more restrictive specification.

Another important patent characteristic that might affect citations is a location of inventors. However, according to Explanation 1 knowledge disclosed in patents should be available to everyone regardless of their geographical location. A location of inventors is more relevant to the theory on interactions between inventors (Jaffe et al. 1993; Carlino & Kerr 2015). In Section 3.4, I show that the concentration is high even within the same inventor who is located in the same geographical area and works across multiple companies. In all subsequent sections, I control for locations of inventors to account for both Explanations 1 and 2.

The decomposition in Figure 5 might also overestimate the role of specialization in patenting because it excludes citations from never-citing technologies. For example, IBM's patent in Table 1 receives most citations from IBM's supplier, Amkor Technology. Therefore, the citations are allocated randomly across the patents that are similar to Amkor's patents and are likely to represent non-competing technologies to IBM. This randomization exercise does not take into account many patents from IBM's competitors that could have made citations to it.

3.4 Movement of Inventors

The "Inventors' Interactions" theory (Explanation 2) argues that the concentration of citations is driven by limited communication between inventors from different firms. According to this theory, Amkor is responsible for 94% of citations to IBM's patent in Table 1 because inventors

[20] The main subgroup level in CPC has 7137 detailed categories while the literature (e.g., Jaffe et al. 1993 and Bell et al. 2019) often considers technologies to be similar if they come from the same 3-digit USPC or NBER sub-class classifications, which have 876 and 445 categories, respectively.

from IBM exclusively communicate with inventors at Amkor. In this section, I ask the following question: if an inventor citing a particular patent in one firm moves to another company, does she continue to cite it? According to Explanation 2, inventors should transfer tacit knowledge across firms and continue citing the same patents. However, I find that inventors significantly change their citation patterns when they move to a different company.

To separate whether citations are firm- or inventor-specific, I find inventors who filed similar patents in multiple companies. I define patents to be similar if they have close application years, a narrow technological class, and the same geographical location of an inventor-mover.

Formally, consider the following statistical framework. Suppose inventor ℓ worked in two companies, i and j, and created $N_i^\ell(g)$ and $N_j^\ell(g)$ patents with characteristics g, respectively. Assume that each patent in firm i (j) makes an independent citation to patent k with probability $p_{k,i}^\ell(g)$ ($p_{k,j}^\ell(g)$). Then the expected number of citations from inventor ℓ in firm i to patent k is

$$n_{k,i}^\ell(g) = p_{k,i}^\ell(g) \cdot N_i^\ell(g),$$

and the goal is to test whether $p_{k,i}^\ell(g) = p_{k,j}^\ell(g)$. To do this, I compare the actual concentration of citations within an inventor with the counterfactual one where citation probabilities are equalized across companies ($p_{k,i}^\ell(g) = p_{k,j}^\ell(g)$).

For example, during 2008 to 2017 inventor Stefan G. Schreck from California created 16 patents in the technological class $A61F2$ while working in Endologix Inc. In 15 out of 16 patents, he made a citation to patent 5690642 assigned to Cook Incorporated. He also applied for 9 patents with similar characteristics in another company, Edwards Lifesciences Corporation, but made zero citations to patent 5690642. If 15 citations were allocated randomly across $16 + 9 = 25$ patents, the expected share of citations from Endologix Inc. would be $^{9.5}/_{15} = 0.63$ and the 95th quantile would be $^{12}/_{15} = 0.8$. However, the actual share ($^{15}/_{15} = 1$) is significantly higher. Notice that for an inventor who worked in two companies the concentration (the share of citations from the most citing firm) cannot be less than 50%.

I compute the average concentration of citations across firms within all inventors who moved between companies, and do the decomposition from Section 3.3. The details are given in Appendix B.5. Figure 6 shows the actual average concentration within an inventor ($\mathcal{AC}_t^w$), the 95th quantile and the median of the concentration with equalized citation rates across firms ($\mathcal{RC}_t^w(q95)$ and $\mathcal{RC}_t^w$), and the median concentration in which both citation rates and patenting are equalized across firms ($\mathcal{PC}_t^w$).[21] The actual average concentration within an

[21]The variable $\mathcal{PC}_t^w$ is greater than 50% because the majority of inventor-movers worked in two firms only. For example, with two citations randomly allocated across two firms the expected concentration measure is

$$\frac{1}{4} \cdot 1 + \frac{1}{4} \cdot 0.5 + \frac{1}{4} \cdot 0.5 + \frac{1}{4} \cdot 1 = 0.75$$

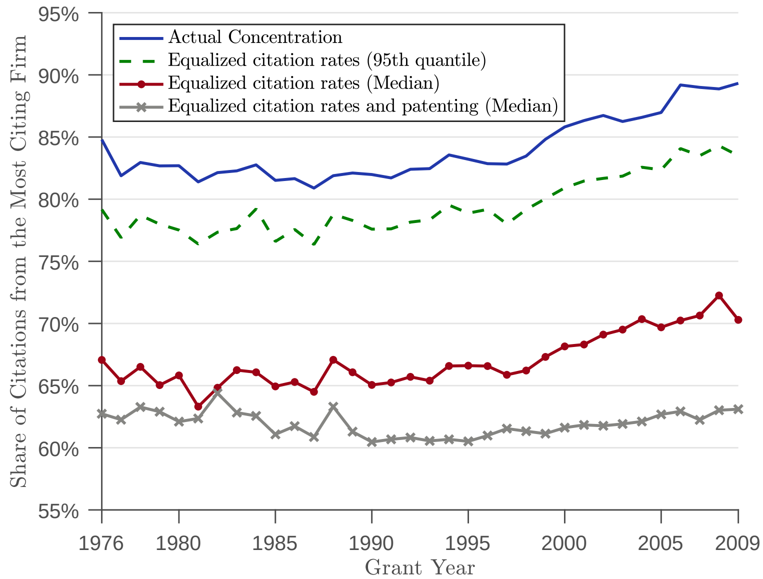


Figure 6: Decomposition of the Concentration Within Inventors Who Moved Across Firms

This figure shows the same decomposition from Figure 5 for the concentration of citations across firms within inventors who patented in multiple companies. The solid line shows the actual aggregate within-inventor concentration of citations across firms. The dashed and the dotted lines show the 95th quantile and the median of the same measure in the Monte-Carlo simulations where citations rate are equalized across firms within an inventor. The crossed line at the bottom shows the concentration (median) where both the citation rates and the number of patents are equalized across firms within an inventor. The averages are almost the same as the medians. To increase the sample size I consider citations from all years, not only 5-year window. The graph is taken until 2009 to ensure that patents have enough time to accumulate citations from inventors-movers. More details are given in Appendix B.5.

inventor is significantly higher relative to the what we would expect if citation probabilities were equalized across firms ($\mathcal{AC}_t^w > \mathcal{RC}_t^w(q95)$). The average decomposition over all years

$$\underbrace{\overline{\mathcal{AC}}^w - \overline{\mathcal{PC}}^w}_{22.1\%} = \underbrace{\overline{\mathcal{AC}}^w - \overline{\mathcal{RC}}^w}_{17.0\%} + \underbrace{\overline{\mathcal{RC}}^w - \overline{\mathcal{PC}}^w}_{5.1\%}$$

shows that the concentration is primarily explained by the differences across firms in citation probabilities rather than by the variance in the number of patents. The difference $\mathcal{AC}_t^w - \mathcal{RC}_t^w$ is stable over time, and there was a slight increase in the difference $\mathcal{RC}_t^w - \mathcal{PC}_t^w$, meaning that the dispersion in the number of patents across firms within an inventor has slightly increased by the end of the period.

This evidence should be interpreted with caution because inventors-movers might differ in their citation rates from inventors who always work in one company. My conjecture is that non-movers would have higher concentration of citations across firms if they were randomly moved to another company. Below I argue that citations are correlated with access to trade secrets.

Based on this interpretation, the conjecture is that inventors who do move between companies are less bound by contractual obligations, such as confidentiality agreements and non-compete clauses, resulting in a less concentrated distribution of citations. An interesting area for future research is to study the movement of inventors caused by exogenous shocks to firms, for example, natural disasters (Barrot & Sauvagnat 2016) or financial constraints (Chodorow-Reich 2014).

3.5 Trade Secret Litigation and Concentration of Citations

The theory of intentional sharing of trade secrets (Explanation 4) argues that firms combine patenting and secrecy. Citations are concentrated because only a limited set of firms gets access to trade secrets of a patent owner.

Testing the connection between secrecy and patent citations is challenging at least for two reasons. First, trade secrets are not observable. I suggest using trade secret litigation to make progress in this measurement problem. Specifically, I find patents involved in federal trade secret litigation.[22] These patents are likely to be a part of a broader technology that also involves trade secrets. For example, the legal case *Waymo LLC v. Uber Technologies, Inc.* was about misappropriation of the trade secrets related to the LiDAR technology for self-driving cars.[23] However, the same lawsuit also had claims regarding patent infringement for three patents.[24] In the complaint, Waymo describes how these patents and trade secrets are complementary to each other:

> "The Replicated Board reflects Waymo's highly confidential proprietary LiDAR technology and Waymo trade secrets. Moreover, the Replicated Board is specifically designed to be used in conjunction with many other Waymo trade secrets and in the context of overall LiDAR systems covered by Waymo patents."

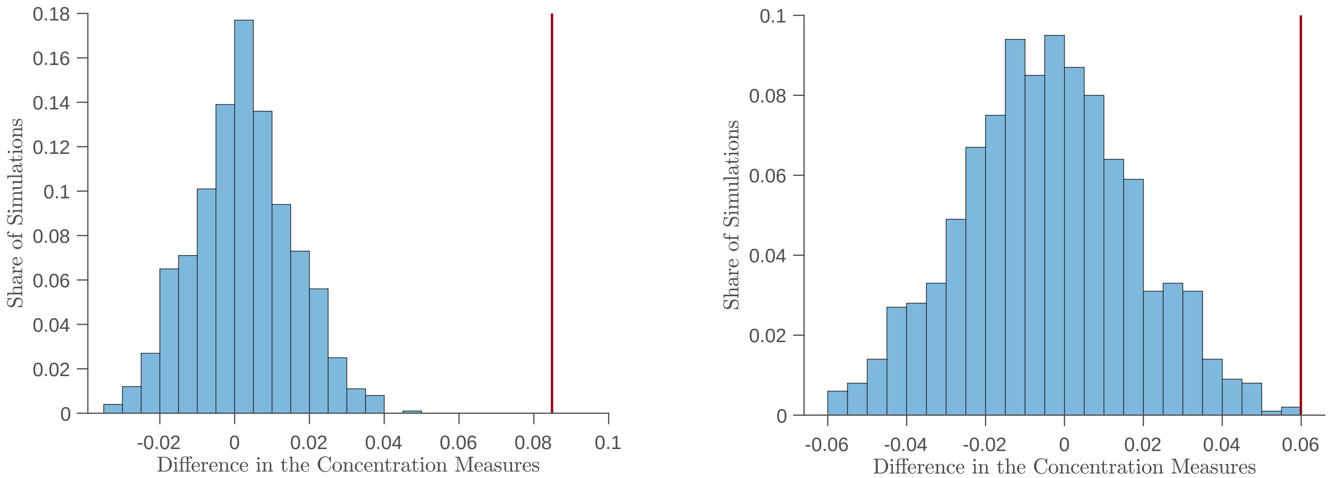
This example highlights that many technologies consist of multiple pieces of knowledge, some of which are kept secret. To replicate and build on a technology, a firm needs access not only to patents, but also to trade secrets. The presence of patents in trade secret litigation is consistent with the "Trade Secrets" explanation for the concentration of citations (Explanation 4). Appendix A provides legal background and more case studies on how firms combine patenting and secrecy.

The second challenge in testing the connection between secrecy and patent citations is identification. Patents bundled with secrets and involved in trade secret litigation are not

[22] The misappropriation of trade secrets can be litigated in both state and federal courts. However, the infringement of patents is litigated in federal courts. Therefore, there is no loss of generality in a focus on federal litigation.

[23] *Waymo LLC v. Uber Technologies, Inc.*, No. 17-2235 (Fed. Cir. 2017).

[24] The patent numbers are 8836922, 9368936, and 9086273.



(a) Herfindahl-Hirschman Index (HHI)

(b) Share From the Most Citing Firm

Figure 7: The Difference in the Concentration of Citations Between Patents With and Without Trade Secrets

These figures show the average difference in the concentration of citations between patents involved in litigation with and without trade secret claims. The vertical lines show the actual difference in the concentration. The histograms show the distribution of the difference when citations are random within (time, location, technology) triple (see Section 3.3 for the details). Panel (a) shows the results when the concentration is measured as the Herfindahl-Hirschman Index, and Panel (b) shows the results when the concentration is measured as the share of citations from the most citing firm.

random. For instance, citations to these patents might differ from citations to other patents due to a publicity effect of litigation. Furthermore, the intellectual property strategies of firms engaged in litigation could differ from those of other companies. To partially address this concern, I find control patents involved in patent infringement litigation but *without* trade secret claims, and I require both treatment and control patents to be from the same firm. In various specifications, I also require patents to share similar characteristics, such as grant years, technological classes, and a number of citations. Nevertheless, the comparison of citation patterns to these patents should be interpreted with caution.

For each patent involved in trade secret litigation, I find control patents which were involved in patent infringement litigation but without trade secret claims. I compute the difference in the concentration of citations between treatment and control patents using two measures: Herfindahl-Hirschman Index and the share of citations from the most citing firm ("Top Share"). Then, I take the average of this difference across patents. I test whether the average difference in the concentration between patents with and without trade secret claims significantly deviates from the difference we would expect if citations were random, controlling for application years, locations of inventors, and technological classes of patents (see Section 3.3). The details are given in Appendix B.6.

Figure 7 gives a visual test against the null hypothesis. The vertical line shows the actual difference in the concentration, and the histogram shows the distribution of this difference if citations were random within the same application years, locations of inventors, and technological classes. For the HHI measure, the difference in concentration is 0.085 which corresponds to approximately 16% higher concentration of citations for patents bundled with trade secrets. For the "Top Share" measure, the difference in concentration is 0.06 which corresponds to approximately 9.2% higher concentration of citations. Table C2 in Appendix C shows the results based on different criteria for selecting control patents, that is whether treatment and control patents have the same grant year, receive a similar number of citations, belong to the same technological class, or are assigned to plaintiffs rather than defendants.[25] Figure 7 shows the results for the most restrictive set of controls (columns 4 and 8 in Table C2). All specifications show a positive and significant difference in concentration between patents involved in litigation with and without trade secret claims.

3.6 Summary

I use the evidence from the previous sections to differentiate between the theories of patent citations proposed in Section 3.1. The results are summarized in Table 3: rows and columns correspond to the theories and evidence, respectively. Green check-marks (✓) indicate that evidence in a column is consistent with a theory in a row, while red cross-marks (✗) mean that the evidence is inconsistent with the theory.

The proposed explanations for patent citations in Section 3.1 are distinct in theory, but they are not mutually exclusive in practice. For example, patent licensing often involves the sharing of tacit knowledge (Arora 1995; Zuniga & Guellec 2009). Therefore, it is not possible to fully differentiate these theories. Nevertheless, all theories should be reconciled with the fact that citations are highly concentrated and primarily come from business partners. For instance, if citations represent knowledge spillovers, the nature of these spillovers differs from the spillovers "in the air" usually assumed in the literature. Although all theories can account for some evidence and rationalize citation patterns under certain assumptions, the theory of intentional trade secret sharing offers the most consistent explanation across all empirical facts.

[25] Controls for grant years and technological classes ensure that patents represent similar technologies. A control for the number of citations ensures that there are no mechanical differences in the concentration. Requiring patents to be assigned to a plaintiff increases a probability that patents are bundled with trade secrets involved in litigation. For example, if firm A shares secrets with firm B under some contractual arrangement (e.g., an acquisition), and firm B patents these secrets, then firm A might sue firm B for misappropriation of the trade secrets. However, in this situation, patents are not bundled with secrets. Requiring patents to be assigned to a plaintiff eliminates such cases.

Table 3: Theories and Evidence of Patent Citations

Theory \ Evidence	Business Partners (Section 2.3)	Control for Narrow Tech (Section 3.3)	Movement of Inventors (Section 3.4)	Trade Secret Litigation (Section 3.5)
Specialization	✓	✗	✗	✗
Inventor Interactions	✗	✓	✗	✓
Licensing	✗	✓	✓	✗
Sharing of Trade Secrets	✓	✓	✓	✓

Notes: The rows list Explanations 1–4 from Section 3.1. The columns list evidence from Sections 2.3 and 3.3–3.5. Green check-marks (✓) indicate that evidence in a column is consistent with a theory in a row, while red cross-marks (✗) mean that the evidence is inconsistent with the theory.

Spillovers and Specialization

The specialization theory (Explanation 1) is based on the assumption that firms fully disclose their knowledge in patent files, and that citations reflect knowledge spillovers. The concentration of citations is a consequence of firm specialization in patents. According to this theory, Amkor is responsible for 94% of citations to IBM's patent in Table 1 because Amkor specializes in patents that can build on IBM's patent. Other firms do not patent in this technology, and other technologies do not benefit from IBM's patent.

The main challenge in testing this theory is finding a measure of technological heterogeneity that is granular enough to capture the specialization. In Sections 3.3 and 3.4, I assume that patents sharing the same application year, detailed patent classes, the textual similarity of abstracts, the location of inventors and those filed by the same inventor should be similar enough to benefit from the same pool of knowledge. Yet, I show that the concentration of citations across firms is observed even for patents sharing these characteristics. Moreover, the presence of patents bundled with trade secrets (Section 3.5) contradicts the assumption of full knowledge disclosure in patent files. The correlation between trade secret bundling and the concentration of citations contradicts the assumption that citations reflect knowledge spillovers.

In general, the specialization theory is consistent with the significant role of business partners in the concentration of citations (Section 2.3) if the knowledge disclosed in patents is highly customized to the patent owner. For example, suppose that Amkor is the only producer of a *customized* input to IBM for the product defined in IBM's patent in Table 1. Firms without an input contract with IBM do not find it profitable to build technologies based on IBM's patent, so Amkor is responsible for the majority of citations. The key assumption here is that Amkor's patents are so customized to IBM that the observable patent characteristics cannot capture this knowledge specificity. However, such spillovers differ from knowledge "in the air" commonly assumed in the literature. The realization of these spillovers requires an input contract with the patent owner.

Overall, Figure 5 shows that the specialization theory provides a reasonable approximation to patent citations until the 1990s. After 2000, to explain the concentration of citations through the specialization theory one needs to assume that knowledge disclosed in patents is so specific that it cannot be measured by observable patent characteristics. Otherwise, it is inconsistent with the evidence in Sections 3.3 ("Control for Narrow Tech") and 3.4 ("Movement of Inventors"). Moreover, the specialization theory is inconsistent with the bundling of patents with trade secrets, documented in Section 3.5 and suggested by legal scholars (Roin 2005).

Interactions Between Inventors

The inventor interactions theory (Explanation 2) is based on the assumption that patents are surrounded by tacit knowledge. This knowledge is diffused through interactions between inventors, which can be captured by patent citations. An important assumption of this theory is that firms have limited control over knowledge diffusion through employees.

The main evidence against this theory is based on citations of inventors-movers (Section 3.4): citations made by the same inventor significantly differ across firms, even when this inventor files observationally similar patents in different companies. Therefore, the concentration of citations is driven by factors specific to firms and not inventors. This evidence supports the view that firms can control the use of their knowledge by former employees in other companies, for example, though non-disclosure agreements and trade secret litigation. In general, the movement of inventors might diffuse knowledge across companies, but these knowledge flows are unlikely to be captured by patent citations.

This theory is also inconsistent with the large role of business partners in citations. While it is expected that inventors employed by business partners communicate more with each other than inventors from two random firms, these inventor interactions between partners are likely to be intentional and controlled by firms, not serendipitous.

This theory is consistent with the high concentration of citations within narrowly defined technologies (Section 3.3). If multiple firms patent in Amkor's technological field, but IBM's inventors only communicate with Amkor's inventors, then it is expected that most citations to IBM will come from Amkor. This theory is also consistent with the evidence on bundling of patents with trade secrets (Section 3.5) because it is based on the presence of tacit knowledge.

Overall, the theory of interactions between inventors is consistent with the presence of tacit knowledge around patents, but the diffusion of this tacit knowledge is likely to be controlled by the owners of the patents.

Licensing

The licensing theory (Explanation 3) is based on two assumptions: first, that firms fully disclose their knowledge in patent applications; second, that firms can control the use of their knowledge though licensing, and patent citations mostly occur between firms with such agreements.

The significant role of business partners in patent citations is consistent with firms' control over their knowledge flows. However, firms with observable licensing contracts account for a small portion of citations among partners and cannot explain the increased concentration. In addition, the presence of patents bundled with trade secrets and its positive correlation with the concentration of citations contradict the assumption of full knowledge disclosure in patent files.

The licensing theory is consistent with the evidence in Section 3.3. If IBM licenses its patent to Amkor only, then we would expect to see the concentration of citations even within narrowly defined technologies. It is also consistent with the evidence on the movement of inventors. If an inventor moves from Amkor to a firm without a licensing agreement with IBM, then this inventor stops citing IBM in the new firm.

Overall, the licensing theory provides a reasonable explanation for citations between companies with licensing agreements, but the share of such citations is small.

Intentional Sharing of Trade Secrets

The theory of intentional trade secret sharing is consistent with all empirical facts. First, firms prefer to keep their trade secrets confidential. However, they might have incentives to share them with certain partners — such as the producers of complementary products, including input suppliers and customers. These incentives explain the significant share of partners in citations. Furthermore, sharing knowledge with some partners does not necessarily require formal licensing contracts. For example, a firm might share knowledge with its supplier to enable the production of higher-quality inputs, and both parties can then use input pricing to divide the benefits of this knowledge sharing.

Second, firms prefer to limit the number of partners with whom they share their secrets, due to the risk of both accidental and intentional leakages. The greater the number of partners who know the secrets, the higher the probability of a leakage to competitors. The incentive to keep secrets within a narrow circle of firms can explain the observed concentration of patent citations, even within narrowly defined technologies.

Third, firms use non-disclosure agreements and threats of trade secret litigation to prevent knowledge leakages through employees moving to other companies. The threats of litigation might explain the evidence on the movement of inventors.

Finally, the trade secret explanation is consistent with the evidence on trade secret litigation.

4 The Decline in Cooperation Between Firms

In this section, I argue that a rise in the risk of trade secret misappropriation might explain the decline in knowledge sharing between business partners, as evidenced from the rise in the concentration of citations since 2000. Section 4.1 describes how risks of trade secret misappropriation affect incentives for knowledge sharing. I propose two factors that could have increased these risks: a rise in trade with China and advances in IT. Section 4.2 shows that technologies more exposed to trade with China experienced a higher growth in the concentration of citations. Section 4.3 discusses the role of IT in the risks of trade secret misappropriation. Section 4.4 discusses policy implications.

4.1 Knowledge Sharing and Risks of Trade Secret Misappropriation

The growth literature is often centered on the process of "creative destruction" — a competitive environment with unintentional knowledge spillovers (Akcigit & Van Reenen 2023).[26] In contrast, the evidence in this paper highlights the importance of cooperation with *intentional* knowledge sharing ("creative construction"). Moreover, the rise in the concentration of citations since 2000 suggests a decline in knowledge flows between firms. Traditional models that assume exogenous knowledge diffusion cannot explain this decline. However, the changes in knowledge flows between firms might be crucial for understanding recent trends in the U.S., such as the rising market concentration and declining business dynamism (Akcigit & Ates 2022).

I argue that a rise in the risks of trade secret misappropriation might be a potential reason behind the decline in cooperation between firms. Firms have incentives to share secrets with certain partners, for example, with input suppliers. However, the more partners know the secrets, the higher the probability that secrets might be leaked to competitors. Therefore, firms face a trade-off: the gains from knowledge sharing, like improved input quality, are weighted against less control over knowledge diffusion. With higher risks of trade secret misappropriation, firms become more selective about which partners can access their trade secrets.

The concentration of citations has started to increase around the year 2000. At least two macro-level trends — potentially influencing the risk of trade secret misappropriation — also

[26] Knowledge spillovers are usually assumed to occur through the disclosure of inventions in patents (Romer 1990) or via serendipitous interactions of inventors (Buera & Lucas 2018). Patent citations are a widely-used measure of knowledge spillovers (Caballero & Jaffe 1993; Akcigit & Kerr 2018; Liu & Ma 2023).

emerged around that time: the rise in trade with China and the increase in the use of IT. I discuss both of these trends in Sections 4.2 and 4.3 below.

To formalize these ideas, consider the following stylized framework. Suppose a firm decides on the number of partners with whom to share secrets, $n \geqslant 0$. For simplicity, assume that partners are input suppliers. If a supplier knows the firm's trade secrets, it can build an input of higher quality. Therefore, the firm's profits are non-decreasing in n, $\pi'(n) \geqslant 0$. Suppose that with probability $\mathbb{P}(q, n)$ competitors acquire the firm's trade secrets and the firm loses its profits. This probability depends on protection against trade secret misappropriation, q. The parameter q represents both ex-ante protection, a prevention of the misappropriation in the first place, and ex-post legal protection, which is invoked if the misappropriation occurs. The probability also depends on the number of partners with access to the trade secrets, n.

I assume that the more partners know the secrets the higher the probability that secrets might be leaked to a competitor, $\frac{\partial \mathbb{P}}{\partial n} \geqslant 0$. The rationale for this assumption is the following. To protect their trade secrets, firms employ a variety of measures, such as ensuring data security, signing non-disclosure agreements, and implementing other contracts that regulate employees' behavior. When a firm shares its secrets with a partner, the protection of these trade secrets becomes dependent not only on the firm's own actions but also on the protective measures and incentives of the partner. The more partners know the secrets, the harder to ensure their security and controlled diffusion.

I also assume that stronger protection against the misappropriation leads to a decrease in misappropriation, $\frac{\partial \mathbb{P}}{\partial q} \leqslant 0$. For example, better legal protection might discourage competitors and partners to misappropriate trade secrets.[27,28] Better enforcement of trade secret laws might also increase incentives of partners to invest more into protection of trade secrets.

The firm decides on the optimal number of suppliers with access to its trade secrets

$$[1 - \mathbb{P}(q, n)] \cdot \pi(n) \to \max_{n} \tag{4.1}$$

[27]For instance, Fadeev (2022) studies a framework where an input supplier has incentives to share trade secrets from one customer with its other customers. These incentives arise from the timing of contractual agreements between firms. Better legal protection is modeled as the probability of wining the litigation against the supplier in court. An increase in this probability leads to less knowledge leakages.

[28]For example, $\mathbb{P}(q, n)$ might have the following structure. Suppose that if there is misappropriation, the firm has the option to sue another firm responsible for it. In such a case, the likelihood of winning the lawsuit and being compensated for the loss of profits is given by the probability q. Also, suppose that the secrets can be independently leaked from each partner with probability p. Then $\mathbb{P}(q, n) = (1 - q) \cdot (1 - p)^n$. In this example, the probability q corresponds to the legal protection against trade secret misappropriation, and the misappropriation is independent from q. In general, $\mathbb{P}(q, n)$ depends on the incentives of firms, so the probability of the misappropriation, p, might be a function of the legal protection, q. For more details, see Fadeev (2022).

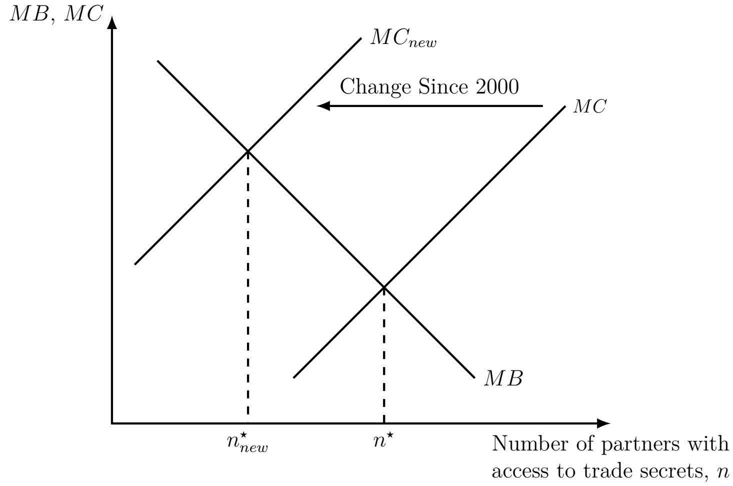


Figure 8: Optimal Number of Partners with Access to Trade Secrets

A firm solves the problem in (4.1). The MB and MC curves represent marginal benefits and costs of sharing trade secrets with more partners, $\frac{\pi'(n)}{\pi(n)}$ and $\frac{1}{1-\mathbb{P}(q,n)} \cdot \frac{\partial P(q,n)}{\partial n}$, respectively.

which is the solution to the following first-order condition[29]

$$\frac{\pi'(n)}{\pi(n)} = \frac{1}{1-\mathbb{P}(q,n)} \cdot \frac{\partial P(q,n)}{\partial n}$$

Figure 8 shows the optimal solution to (4.1). The shift in the marginal costs of trade secret misappropriation, $\frac{1}{1-\mathbb{P}(q,n)} \cdot \frac{\partial P(q,n)}{\partial n}$, to the left caused a decline in the number of partners with access to the firm's trade secrets.

4.2 Trade with China and Concentration of Citations

One of the factors influencing the rise in risks of trade secret misappropriation could be increasing trade with China. According to the U.S. Counterintelligence Office, "[t]he pace of foreign economic collection and industrial espionage activities against major US corporations and US Government agencies is accelerating."[30] Moreover, trade with China could weaken trade secret protection because the enforcement of U.S. trade secret laws tends to be less effective in cases involving international trade secret misappropriation (Almeling 2012).[31] For

[29] I also assume that $\pi''(n) \leqslant 0$ and $\frac{\partial^2 P}{\partial n^2} \geqslant 0$.

[30] Office of the National Counterintelligence Executive, "Foreign Spies Stealing US Economic Secrets in Cyberspace", 2011.

[31] In the context of the example from footnote 28, the probability of legal protection, q, can be decomposed into protection against international and domestic misappropriation: $q = \alpha q_c + (1-\alpha) q_d$, where α is the share of competitors from China, q_c is the legal protection against misappropriation from China, and q_d is the protection in the U.S. Assume that $q_c < q_d$, and the increasing entry of Chinese competitors corresponds to the increase in α. In this case, the overall protection goes down.

instance, U.S. courts may not have jurisdiction to hear certain international cases.[32]

This section shows that technologies more exposed to import competition from China experienced higher growth in the concentration of citations. To isolate changes in the import competition level unrelated to U.S. demand and technological shocks, I instrument U.S. imports from China by Chinese exports to other high-income countries (Autor et al. 2013). Autor et al. (2020) used this methodology to show that imports from China led to the decline in corporate patenting in the manufacturing sector. I complement their evidence by studying how imports from China affected citation patterns between firms.

Following Autor et al. (2020), I define the measure of trade exposure at the four-digit Standard Industry Classification (SIC) over the two subperiods, 1991 to 1999 and 1999 to 2007,

$$\Delta IP_{i1} = \frac{M_{i,1999} - M_{i,1991}}{Y_{i,91} + M_{i,91} - E_{i,91}} \text{ and } \Delta IP_{i2} = \frac{M_{i,2007} - M_{i,1999}}{Y_{i,91} + M_{i,91} - E_{i,91}} \tag{4.2}$$

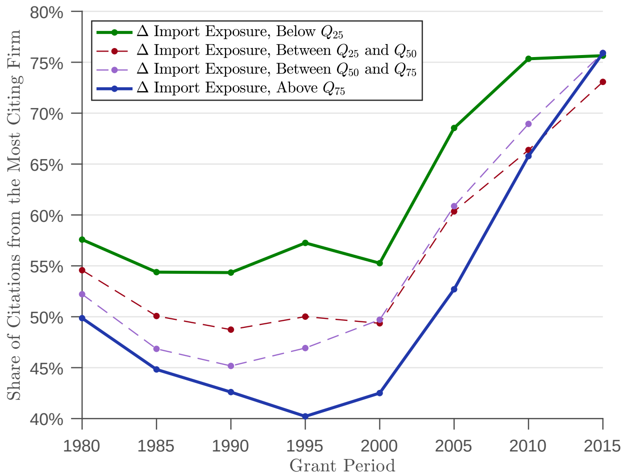
where $M_{i,t}$ is the U.S. imports from China for industry i and year $t \in \{1991, 1999, 2007\}$, and $Y_{i,91} + M_{i,91} - E_{i,91}$ is the absorption at the start of the period (industry shipments plus imports minus exports). For each patent, I calculate the import penetration for its technological class using the mapping of four-digit SIC industries to technological classes implied from patents owned by publicly traded firms as in Autor et al. (2020).

Panel (a) of Figure 9 shows the concentration measure from Figure 1 for different technological classes based on their exposure to import competition from China, ΔIP_{i2}. All classes experienced an increase in the concentration of citations after 2000. However, the dynamics are different based on the exposure to import competition. For classes with the least exposure (below first quartile Q_{25}), the concentration of citations was initially high (around 55%) and stable prior to 2000, and then it increased up to 75%. For classes with the most exposure (above the third quartile Q_{75}), the concentration was initially lower (around 50%), decreased to around 40% near 2000, and then it also increased up to 75%. Thus, technological classes with the most exposure to China shock experienced faster growth in the concentration.

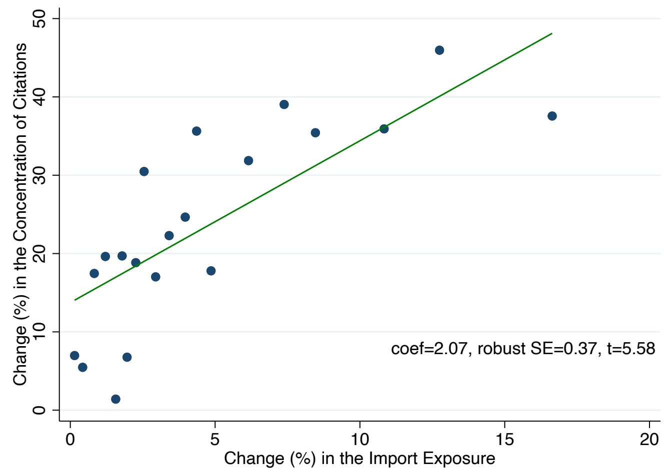
I study the change in the concentration of citations around 2000 in the regression specifications. Specifically, I define for each technological class the average concentration measure among patents granted in the seven-year period starting from 1977, 1984, 1991, 1998, 2005.[33] Appendix B.7 provides more details. The concentration measure for technological class

[32] See *TianRui Group Co. Ltd. v. Int'l Trade Comm'n*, 661 F.3d 1322 (Fed. Cir. 2011).

[33] Periods are 1977–1983, 1984–1990, 1991–1997, 1998–2004, 2005–2011. The sample construction for patents is different from Autor et al. (2020) due to the nature of the outcome variable. Autor et al. (2020) consider patents applied in the years 1975, 1983, 1991, 1999, 2007. Their main outcome variable is the number of patents while in this paper it is the concentration of citations. The concentration measure has a meaningful interpretation only for highly-cited patents. For example, the patent with one citation always has concentration equal to one. Therefore, I focus on the set of the top 1% of the most cited patents. Since each technological class has a small number of highly-cited patents in each year, I take the average of the concentration measure over a seven-year



(a) The concentration for technological classes with different exposure to import competition.



(b) Changes in the concentration and in the import penetration from China.

Figure 9: Trade with China and Concentration of Citations

These figures show the relationship between the concentration of citations and the exposure to import competition from China. Panel (a) shows the concentration measure from Section 2 (Figure 1) for different technological classes divided into quartiles based on their exposure to import competition from China, ΔIP_{i2} in (4.2). The concentration measure for year t shows the average concentration in years $[t-4, t]$. Panel (b) shows the binned scatter plot of the change in the concentration of citations and the change in the import exposure from China. The specification is weighted by the number of Compustat-matched U.S.-inventor patents in a technology class.

j and for the period starting from t is denoted by $\mathcal{C}_{j,t}$. I define the following growth measures

$$\Delta y_{j1} = 100 \cdot \ln\left(\mathcal{C}_{j,1998}/\mathcal{C}_{j,1991}\right) \text{ and } \Delta y_{j2} = 100 \cdot \ln\left(\mathcal{C}_{j,2005}/\mathcal{C}_{j,1998}\right)$$

Panel (b) of Figure 9 shows that technological classes more exposed to trade with China ($\Delta IP_{j\tau}$) experienced a higher growth in the concentration of citations ($y_{j\tau}$).

I estimate the following specification

$$\Delta y_{j\tau} = \beta \Delta IP_{j\tau} + \gamma X_{j0} + \varepsilon_{j\tau} \tag{4.3}$$

where $\tau \in \{1, 2\}$ and X_{j0} is the set of controls. To control for the aggregate trend in the concentration of citations, I include time fixed effects. Since the concentration measure depends on the total number of citations, I also include the change in the average number of citations for each technological class. Moreover, I include two lags of the outcome variable to control for technology-specific trends prior to China shock. I also include fixed effects for 11 manufacturing

period. I also do the analysis at the technology class level rather than at the firm level because most firms have a small number of highly-cited patents.

Table 4: Trade with China and Increase in the Concentration of Citations

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: OLS			
Δ Tech Class Exposure	2.06	1.52	1.51	1.77	1.41	1.41
to Chinese Imports	(0.44)	(0.42)	(0.42)	(0.37)	(0.42)	(0.42)
			Panel B: 2SLS			
Δ Tech Class Exposure	2.31	1.57	1.55	1.92	1.71	1.70
to Chinese Imports	(0.49)	(0.50)	(0.50)	(0.44)	(0.61)	(0.62)
Time FE		Yes	Yes	Yes	Yes	Yes
Δ Citations			Yes	Yes	Yes	Yes
2 Lags of outcomes				Yes	Yes	Yes
11 sectors, 6 Tech					Yes	Yes
Software Patents						Yes

Notes: This table shows the estimated coefficient β for the specification in (4.3). Panel A shows the results for simple OLS regressions. Panel B shows the results for the specification in which the import penetration from China is instrumented with Chinese exports to non-U.S. high-income markets (Autor et al. 2020). Regressions consider the effect of higher growth in import penetration from China on the increase in the concentration of citations at the technology class level. Industry exposure to Chinese competition is mapped to technology class exposure using the mapping implied by the U.S. publicly traded firms in Compustat as in Autor et al. (2020). Controls include time fixed effects, a change in the average number of citations for a technology class, 2 lags of the outcome variable, fixed effects for 11 manufacturing sectors and for 6 main NBER technology categories, and a dummy for software technological classes. I define the software classes as classes where more than 50% of software subclaclasses according to Graham & Vishnubhakat (2013). All specifications are weighted by the number of Compustat-matched U.S.-inventor patents in a technology class. Standard errors are clustered at the technology class level.

sectors and for 6 main NBER technological categories. Finally, I control for the rising importance of software inventions (Chattergoon & Kerr 2021). Specifically, for each technological class I include a dummy variable indicating whether it has more than 50% of software subclasses (Graham & Vishnubhakat 2013).

Panel A in Table 4 shows the results of simple OLS regressions, and Panel B shows the results for the specification in which changes in US import exposure ($\Delta IP_{j\tau}$) are instrumented by changes in Chinese exports to non-U.S. high-income countries. All specifications show a positive and significant relationship between the changes in the import competition from China and the growth in the concentration of citations.

Discussion and Robustness Checks

Trade with China has had a profound effect on various aspects of the U.S. economy, from innovation (Autor et al. 2020; Hoberg et al. 2021) and labor markets (Autor et al. 2013) to the structure of supply chains (Antràs et al. 2017). The evidence in this section suggests that it might have also affected incentives for knowledge sharing among business partners. I provide additional discussion and robustness checks for this result below.

The instrument from Autor et al. (2013) is designed to isolate changes in the trade with China unrelated to U.S. demand and technological shocks. Table C3 in Appendix C shows the results from two additional placebo exercises. First, I show that the relationship between China shock and the rise in the concentration is insignificant for patents assigned to non-corporate entities (e.g., universities and government agencies). Therefore, the effect of trade competition with China is specific to the corporate sector, and the results are unlikely to be driven by the correlation between general technological changes and globalization. Second, I regress lag outcome variables (pre 1991) on future changes in imports from China. The coefficients are insignificant, so the main results are unlikely to be driven by contemporaneous changes in the technological opportunities and trade.

Using the methodology outlined in Section 3.3, I also study whether China shock influenced the rise in the concentration of citations through changes in firm patenting behavior (N) or in firm citation rates (p).[34] Figure C13 in Appendix C shows that 84% of the rise in the concentration of citations in the technologies most exposed to trade with China is attributed to changes in the citation rates. In contrast, for the least exposed technologies this share is 45%. Overall, China shock changed the way firms cite each other. This effect is distinct from the decline in patenting documented in Autor et al. (2020). For instance, firms' exit from patenting cannot fully explain the rise in the concentration of citations.

The evidence in this section suggests that trade with China led to the decline in knowledge sharing between firms. Identifying the precise mechanism of how trade with China led to this decline is challenging because of multiple, simultaneous changes in firms' incentives. I consider the increase in the risks of trade secret misappropriation as the primary mechanism through which trade with China has influenced knowledge-sharing incentives among business partners. This increase corresponds to a leftward shift of the marginal costs curve of knowledge sharing in Figure 8. However, trade with China could have also affected the marginal benefits curve. For instance, it has changed the level of competition (Autor et al. 2020) and the structure of supply chains (Antràs et al. 2017).[35] I leave the analysis of various mechanisms for future research.

[34] In addition to application years and technological classes, I also control for the location of inventors to take into account the geographical specialization of firms (see Appendix B.4).

[35] For example, trade with China has improved U.S. firms' access to more efficient suppliers (Antràs et al.

4.3 Advances in IT and Risks of Trade Secret Theft

Advances in IT and the internet have reduced the costs of information storage and remote access, potentially making it easier to misappropriate trade secrets. Legal practitioners argue that "[t]he digital world is no friend to trade secrets" (Candiff 2009). In the IT era, sensitive information like blueprints are stored in a digital form, and the ease with which digital files can be downloaded, emailed, or saved to a flash drive makes them more susceptible to theft, even with multiple layers of security. The remote access to information also makes companies more susceptible to espionage. In line with this argument, Hoberg et al. (2021) use staggered internet rollout in China to show that U.S. firms have increased their complaints about intellectual property theft as information access costs for Chinese firms have decreased.

Figure C14 in Appendix C shows that the rise in the concentration of citations is more pronounced in the technologies with software patents, where I measure software patents using the methodology from Graham & Vishnubhakat (2013).[36] However, this methodology identifies technologies *developing* software and not necessarily the ones that rely on IT. The pervasive use of IT across all industries might explain why the rise in the concentration of citations is observed in almost all technological classes. Finding the appropriate measure of the IT use and testing its connection with knowledge sharing would be a promising area for future research.

4.4 Policy Implications

Akcigit & Ates (2022) argue that a decline in knowledge flows from frontier firms to lagging competitors might be responsible for the recent macro trends in the U.S. economy, such as the rising market concentration and declining business dynamism. They argue that this decrease could be explained by the anticompetitive use of patent protection: frontier firms accumulate large portfolios of patents ("patent thickets") and pursue legal actions to prevent patent infringement, making it harder for other firms to build on the existing technologies.

The evidence in this paper is consistent with the decline in knowledge diffusion, but it points to the alternative mechanism behind this decline: a decrease in knowledge sharing with business partners, such as input suppliers. Consider the framework from Section 4.1. If the risks of trade secret misappropriation go up (a decline in q), then all else being equal, the probability of knowledge flows to a competitor, $\mathbb{P}(q, n)$, goes up. However, firms respond to these risks

2017). Sharing secrets with these more efficient suppliers could be more beneficial for U.S. firms compared to knowledge sharing with less efficient domestic suppliers. However, the location of these suppliers in China could be associated with higher risks of trade secret misappropriation.

[36]Specifically, Graham & Vishnubhakat (2013) define subsclasses in the US Patent Classification (USPC) associated with software technologies. I separate technological classes (USPC main class) based on whether they have more or less than 50% of software subclasses. Figure C14 shows the dynamics of the concentration of citations from Figure 1 based on this separation of technologies.

by decreasing knowledge sharing with partners, n. This endogenous response might lead to a decline in knowledge flows to competitors, $\mathbb{P}(q, n)$.

An increase in the anticompetitive practices would primarily affect citations among competitors. In contrast, the increase in the concentration of citations is driven by changes in citation patterns among business partners. Based on industry affiliation and contractual relationships, these partners are unlikely to represent direct competitors that could replace cited firms (see details in Section 2.3).

The traditional innovation policy is usually focused on the problem of R&D underinvestment due to knowledge spillovers. If firms can control the diffusion of knowledge, then R&D subsidies might not be the right policy tool to increase economic growth. Instead, the innovation policy should aim to increase knowledge sharing among firms. Given that the decline in knowledge sharing could be caused by increasing risks of trade secret misappropriation, a potential policy response might be to increase legal protection against the misappropriation.[37] Such policies are regulated by trade secret laws.

The U.S. government recognizes problems of trade secret misappropriation. Over the last decades, several laws were enacted to expand the set of legal tools for protection of trade secrets. One of the recent laws was the Defend Trade Secret Act (DTSA) of 2016, which created for the first time a federal civil cause of action for misappropriation of trade secrets. One of the primary motives behind this act was to provide stronger protection of American firms against foreign trade secret misappropriation.[38] The effect of these changes as well as the design of the optimal trade secret policy would be a promising area for future research.

5 The Use of Patent Citations in the Literature

Patent citations are a common measure of knowledge spillovers. They are used to estimate growth models (Caballero & Jaffe 1993; Akcigit & Kerr 2018), to test theories in economic geography (Ellison et al. 2010), and to provide policy recommendations (Liu & Ma 2023).

This paper shows that patent citations are unlikely to measure spillovers, but they still provide valuable information on collaboration between firms. Therefore, patent citations might be more useful in testing theories of intentional knowledge sharing. For instance, Gomes-

[37] The USPTO can also make stronger disclosure requirements for patent applications. However, this policy is hard to implement because patent examiners already struggle to enforce existing disclosure rules (see Appendix A). Moreover, stronger disclosure rules might push firms toward greater secrecy.

[38] President Obama at the signing ceremony of the DTSA in 2016: "One of the biggest advantages that we've got in this global economy is that we innovate. We come up with new services, new goods, new products, new technologies. Unfortunately, all too often, some of our competitors, instead of competing with us fairly, are trying to steal these trade secrets from American companies, and that means a loss of American jobs, a loss of American markets, a loss of American leadership."

Casseres et al. (2006) use patent citations to study how firm characteristics affect knowledge flows in strategic alliances. Fadeev (2022) uses patent citations to study conditions under which suppliers transfer knowledge from one customer to another. Self-citations might also be useful in studies of knowledge flows within organizations.

In general, researchers should make adjustments to the use of patent citations depending on a question they study. For instance, the studies on localization of knowledge spillovers usually exclude self-citations because it is a common assumption that such citations do not represent knowledge spillovers (Jaffe et al. 1993). A similar argument can be made regarding citations between business partners. To the best of my knowledge, the only paper on knowledge spillovers that adjusts citations for business relations between firms is Atkin et al. (2022).[39]

I propose two possible adjustments to the use of patent citations. First, researchers can separate inter-firm citations into citations between business partners and unrelated parties. Citations between unrelated parties are more likely to represent knowledge spillovers commonly assumed in the literature. However, this approach might be too restrictive in a sense that some citations between partners can still reflect spillovers. The second adjustment divides all citations into *concentrated* citations between partners and all other citations. For instance, IBM's patent from Table 1 receives 94% of citations from its input supplier, Amkor Technology. This share is so high that citations from Amkor are unlikely to represent unintentional spillovers. Formally, for each cited patent the concentration of citations should be compared relative to a benchmark of random citations, as in Section 3.3. If the concentration of citations for a patent exceeds the 95^{th} quantile of the random concentration, then citations from the most citing firm should be interpreted as *intentional* knowledge flows.

6 Conclusion

This paper provides evidence that firms have significant control over the diffusion of knowledge they generate. Therefore, knowledge flows are determined by a firm's incentives for cooperation with other companies. I argue that firms decreased knowledge sharing with business partners due to an increased risk of trade secret misappropriation. In addition, the incentives for knowledge sharing might also depend on market structure, the duration of a relationship, available contracts, and other factors. More research on the management of knowledge flows might provide deeper insights about recent changes in the economic growth.

[39] Atkin et al. (2022) study knowledge spillovers coming from serendipitous face-to-face interactions between inventors. They instrument face-to-face meetings between workers from two establishments with meetings between workers from similar establishments whose industries neither cite nor supply each other. They also exclude citations based on additional restrictions regarding the geography of meetings and establishments.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2020). Sampling-Based Versus Design-Based Uncertainty in Regression Analysis. *Econometrica*, 88(1), 265–296.

Aghion, P., Akcigit, U., & Howitt, P. (2014). What Do We Learn from Schumpeterian Growth Theory,. *Handbook of Economic Growth*, 2B, 515–563.

Aghion, P., Bergeaud, A., & Van Reenen, J. (2023a). The Impact of Regulation on Innovation. *American Economic Review*, Fortcoming.

Aghion, P. & Howitt, P. (1992). A Model of Growth Through Creative Destruction. *Econometrica*, 60(2), 323–51.

Aghion, P., Jackson, M. O., Mayerowitz, A., & Tagade, A. V. (2023b). Innovation Networks and Business-Stealing. *The Economics of Creative Destruction, edited by Ufuk Akcigit and Jon Van Reenen*.

Akcigit, U. & Ates, S. T. (2022). What Happened to U.S. Business Dynamism? *Journal of Political Economy*, 131(8), 2059–2124.

Akcigit, U., Hanley, D., & Stantcheva, S. (2021). Optimal Taxation and R&D Policies. *Econometrica*, Fortcoming.

Akcigit, U. & Kerr, W. (2018). Growth Through Heterogenous Innovations. *Journal of Political Economy*, 126(4), 1374–1443.

Akcigit, U. & Van Reenen, J., Eds. (2023). *The Economics of Creative Destruction*. Harvard University Press.

Alcácer, J., Gittelman, M., & Sampat, B. (2009). Applicant and Examier Citations in U.S. Patents: An Overview and Analysis. *Research Policy*, 38, 415–427.

Alfaro-Ureña, A., Manelici, I., & Vasquez, J. P. (2022). The Effects of Joining Multinational Supply Chains: New Evidence from Firm-to-Firm Linkages. *Quarterly Journal of Economics*, 137(3), 1495–1552.

Almeling, D. S. (2012). Seven Reasons Why Trade Secrets Are Increasingly Important. *Berkeley Technology Law Journal*, 27(2), 1091–1117.

Amara, N., Landry, R., & Traoré, N. (2008). Managing the Protection of Innovations in Knowledge-Intensive Business Services. *Research Policy*, 37, 1530–1547.

Anton, J. J., Greene, H., & Yao, D. A. (2006). Policy Implications of Weak Patent Rights. *Innovation Policy and the Economy, Volume 6, edited by Adam B. Jaffe, Josh Lerner, and Scott Stern*, (pp. 1–26).

Antràs, P., Fort, T. C., & Tintelnot, F. (2017). The Margins of Global Sourcing: Theory and Evidence from US Firms. *American Economic Review*, 107(9), 2514–2564.

Arora, A. (1995). Licensing Tacit Knowledge: Intellectual Property Rights and the Market for Know-How. *Economics of Innovation and New Technology*, 4(1), 41–59.

Arora, A. (1997). Patents, Licensing and Market Structure in the Chemical Industry. *Research Policy*, 26(4–5), 391–403.

Arora, A., Fosfuri, A., & Gambardella, A. (2001). *Markets for Technology. The Economics of Innovation and Corporate Strategy*. The MIT Press.

Arqué-Castells, P. & Spulber, D. F. (2022). Measuring the Private and Social Returns to R&D: Unintended Spillovers versus Technology Markets. *Journal of Political Economy*, Forthcoming.

Atkin, D., Chen, K., & Popov, A. (2022). The Returns to Face-to-Face Interactions: Knowledge Spillovers in Silicon Valley. *Working Paper*.

Autor, D., Dorn, D., Hanson, G. H., Pisano, G., & Shu, P. (2020). Foreign Competition and Domestic Innovation: Evidence from US Patents. *American Economic Review: Insights*, 2(3), 357–374.

Autor, D. H., Dorn, D., & Hanson, G. H. (2013). The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review*, 103(6), 2121–2168.

Bai, J., Barwick, P., Cao, S., & Li, S. (2022). Quid Pro Quo, Knowledge Spillover, and Industrial Quality Upgrades: Evidence from the Chinese Auto Industry. *Working Paper*.

Barrot, J.-N. & Sauvagnat, J. (2016). Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks. *Quarterly Journal of Economics*, 131(3), 1543–1592.

Bell, A., Chetty, R., Jaravel, X., Petkova, N., & van Reenen, J. (2019). Who Becomes an Inventor in America? The Importance of Exposure to Innovation. *Quarterly Journal of Economics*, 134(2), 647–713.

Bloom, N., Van Reenen, J., & Williams, H. (2019). A Toolkit of Policies to Promote Innovation. *Journal of Economic Perspectives*, 33(3), 163–184.

Buera, F. & Lucas, R. (2018). Idea Flows and Economic Growth. *Annual Review of Economics*, 10, 315–45.

Caballero, R. & Jaffe, A. B. (1993). How High are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth. *NBER Working Paper No. 4370*.

Candiff, V. A. (2009). Reasonable Measures to Protect Trade Secrets in a Digital Environment. *IDEA: Intellectual Property Law Review*, 49(3), 359–410.

Carlino, G. & Kerr, W. (2015). Aglomeration and Innovation. *Handbook of Regional and Urban Economics*, 5, 349–404.

Cassiman, B. & Veughelers, R. (2002). R&D Cooperation and Spillovers: Some Empirical Evidence from Belgium. *American Economic Reivew*, 92(4), 1169–1184.

Chattergoon, B. & Kerr, W. R. (2021). Winner Takes All? Tech Clusters, Population Centers, and the Spatial Transformation of U.S. Invention. *Research Policy*, Forthcoming.

Chodorow-Reich, G. (2014). The Employment Effects of Credit Market Disruptions: Firm-Level Evidence from the 2008-9 Financial Crisis. *Quarterly Journal of Economics*, 129(1), 1–59.

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

Dasaratha, K. (2023). Innovation and Strategic Network Formation. *The Review of Economic Studies*, 90(1), 229–260.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv*, Available at https://arxiv.org/abs/1810.04805.

Duguet, E. & MacGarvie, M. (2005). How Well Do Patent Citations Measure Flows of Technology? Evidence from French Innovation Surveys. *Economics of innovation and New Technology*, 14(5), 375–393.

Eeckhout, J. & Jovanovic, B. (2002). Knowledge Spillovers and Inequality. *American Economic Review*, 92(5), 1290–1307.

Ellison, G., Glaeser, E. L., & Kerr, W. (2010). What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review*, 100(3), 1195–1213.

Fadeev, E. (2022). Essays on Economics of Innovation and International Trade. *Doctoral Dissertation, Harvard University Graduate School of Arts and Sciences.*

Gomes-Casseres, B., Hagedoorn, J., & Jaffe, A. B. (2006). Do Alliances Promote Knowledge Flows? *Journal of Financial Economics*, 80(1), 5–33.

Graham, S. & Vishnubhakat, S. (2013). Of Smart Phone Wars and Software Patents. *Journal of Economic Perspectives*, 27(1), 67–86.

Graham, S. J. H., Marco, A. C., & Myers, A. F. (2018). Patent Transactions in the Marketplace: Lessons from the USPTO Patent Assignment Dataset. *Journal of Economics and Management Strategy*, 27(3), 343–371.

Grossman, G. M. & Helpman, E. (1991). Quality Ladders in the Theory of Growth. *Review of Economic Studies*, 58(1), 43–61.

Hall, B. H., Helmers, C., Rogers, M., & Sena, V. (2014). The Choice between Formal and Informal Intellectual Property: A Review. *Journal of Economic Literature*, 52(2), 375–423.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2005). Market Value and Patent Citations. *RAND Journal of Economics*, 36(1), 16–38.

Hoberg, G., Li, Y., & Phillips, G. M. (2021). U.S. - China Innovation Competition. *Working Paper.*

Jaffe, A. B., Trajtenberg, M., & Fogarty, M. S. (2000). Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors. *American Economic Review: Papers and Proceedings*, 90(2), 215–218.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Quarterly Journal of Economics*, 108(3), 577–598.

Jones, C. I. (2005). Growth and Ideas. *Handbook of Economic Growth, Volume 1B, edited by Philippe Aghion and Steven N. Durlauf.*

Jorda, K. F. (2008). Patent and Trade Secret Complementariness: An Unsuspected Synergy. *Washburn Law Journal*, 48(1), 1–32.

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

Kuhn, J., Younge, K., & Marco, A. (2020). Patent Citations Reexamined. *RAND Journal of Economics*, 51(1), 109–132.

Lemley, M. A. & Shapiro, C. (2005). Probabilistic Patents. *Journal of Economic Perspectives*, 19(2), 75–98.

Lerner, J. & Seru, A. (2022). The Use and Misuse of Patent Data: Issues for Corporate Finance and Beyond. *Review of Financial Studies*, 35(6), 2667–2704.

Liu, E. & Ma, S. (2023). Innovation Networks and R&D Allocation. *Working Paper*.

Marshall, A. (1920). *Principles of Economics*. London: MacMillan.

Mazzoleni, R. & Nelson, R. R. (1998). The Benefits and Costs of Strong Patent Protection: a Contribution to the Current Debate. *Research Policy*, 27(3), 273–284.

Melitz, M. & Polanec, S. (2015). Dynamic Olley-Pakes productivity decomposition with entry and exit. *The RAND Journal of Economics*, 46(2), 362–375.

Moretti, E. (2021). The Effect of High-Tech Clusters on the Productivity of Top Inventors. *American Economic Review*, 111(10), 3328–75.

Ouellette, L. L. (2012). Do Patents Disclose Useful Information. *Harvard Journal of Law and Technology*, 25(2), 545–607.

Png, I. (2017). Secrecy and Patents: Theory and Evidence from the Univform Trade Secrets Act. *Strategy Science*, 2(3), 176–193.

Price II, N., Rai, A. K., & Minssen, T. (2020). Knowledge transfer for large-scale vaccine manufacturing. *Science*, 369(6506), 912–914.

Roin, B. N. (2005). The Disclosure Function of the Patent System (Or Lack Thereof). *Harvard Law Review*, 118(6), 2007–2028.

Romer, P. (1990). Endogenous Technical Change. *Journal of Political Economy*, 98(5), S71–S102.

Rosenberg, N. (1963). Technological Change in the Machine Tool Industry, 1840–1940. *The Journal of Economic History*, 23(4), 414–443.

Scotchmer, S. (2004). *Innovation and Incentives*. The MIT Press.

Singh, J. & Marx, M. (2013). Geographical Constraints on Knowledge Spillovers: Political Borders vs. Spatial Proximity. *Management Science*, 59(9), 2056–2078.

Thompson, P. & Fox-Kean, M. (2005). Patent Citations and the Geography of Knowledge Spillovers: A Reassessment. *American Economic Review*, 95(1), 450–460.

Ziedonis, R. H. (2004). Don't Fence Me In: Fragmented Markets for Technology and the Patent Acquision Strategies of Firms. *Management Science*, 50(6), 804–820.

Zuniga, M. P. & Guellec, D. (2009). Who Licenses Out Patents and Why? Lessons from a Business Survey. *OECD Working Paper*.

Appendix. For Online Publication

Table of Contents

A Case Studies: Combining Secrecy and Patenting

On May 5th 2021, the U.S. administration announced that it would support the temporary waiver of IP rights on messenger RNA technology for Covid-19 vaccines. The announcement generated a lot of debates whether this policy can increase the production of vaccines. Many IP lawyers and scholars argue that one of the problems is that patents do not disclose enough information for the replication of the mRNA technology.[40]

> " 'A waiver helps to keep generic manufacturers safe from patent litigation. But they won't even get to that stage without the cooperation of the inventors.' On top of that, not all that you need to produce these vaccine generics is patented and, thus, disclosed in the patent application. Much is protected against competitors not via patent law but by keeping it secret. 'You can't force the company that hold these secrets to pass it on to you.' "

The practice of combining trade secrets and patents in chemical innovations has a long history. Arora (1997) describes how dyestaff producers in the first half of the 20th century patented codified individual chemical compounds but kept tacit knowledge on how to combine these compounds secret. The same approach was used by ammonia producers:

> "The Haber-Bosch process for ammonia, a truly significant process innovation, was protected by more than 200 patents that covered the apparatus, temperatures, and pressures, but avoided particulars about the catalysts employed or their preparation. The catalyst was critical to the successful operation of the process, and keeping it secret significantly increased the expense and time for firms trying to circumvent the Haber-Bosch patent ..."

Combining secrecy with patents seems to be inconsistent with the disclosure requirements of patents. For example, inventors should disclose their preferred method for carrying out the invention ("best mode") in order to "restrain inventors from applying for patents while at the same time concealing from the public preffered embodiments of the inventions they have in fact conceived."[41] However, given high uncertainty about the limits of this requirement firms try minimize the amount of disclosed knowledge. For instance, in *Fonar Corp vs. Gen. Elec. Co.* case (software) inventors did not disclose their source code, and in *Amgen, Inc.*

[40]The quote is taken from the interview with Jayashree Watal, a professor at the Georgetown University School of Law in Washington D.C., who worked for more than three decades at the WTO secretariat ("Three Crises and One Waiver", Verfassungsblog, May 7th, 2021. For additional discussions, see also "The COVID-19 Vaccine Patent Waiver: The Wrong Tool for the Right Goal", Bill of Health, Petrie-Flom Center at Harvard Law School, May 5h, 2021.

[41]See *Teleflex, Inc. v. Ficosia N. Am. Corp., 299 F.3d 1313, 1330 (Fed. Cir. 2002).*

vs. Chungai Pharm. Co. case (biotech) inventors did not disclose the specific cell lines used in their products.[42] In both cases, courts supported the inventors. Jorda (2008) provides a general discussion on the limits of the "best mode" requirement from a legal point of view. For example, it applies only to the knowledge that inventors had at the time of patent filing. Given that patents are often filed at the early stage of research, preferred embodiments are often discovered later. In the case *C&F Packing Co. v. IBP, Inc.*, C&F had "developed a process for making and freezing a precooked sausage for pizza toppings" that was superior to existing technologies.[43] C&F got two patents: one on the equipment and another on the process itself. After that they continued to improve the technology but kept it secret. C&F shared these secrets under a confidentiality agreement with a supplier who leaked them to its customer, Pizza Hut. The court ordered Pizza Hut to pay 10.9$ million to C&F for trade secret misappropriation.

B Data Appendix

B.1 Data Details

Patents. The main source of patent data is PatentsView. Autor et al. (2020) provide a matching of patent assignees to Compustat firm names for publicly traded firms. I use their existing matching of assignee names to Computstat firms for the period 1976–2014 to extend it for years up to 2019. For the rest of the patents, I follow the procedure outline in Autor et al. (2020) for cleaning and standardizing firm names (e.g., replace "Incorporated" with "INC"). Finally, I matched around 100 thousand patents manually for the largest assignees.

FactSet Revere. I use two data sets from FactSet. FactSet Revere Company provides basic information on companies, including their names. FactSet Revere Supply Chain Relationships provides information on business relationships between firms. I discuss the types of relationships below. The following quote from FactSet's manual describes how they collect data on business relationships:

> "FactSet analysts systematically collect companies' relationship information exclusively from primary public sources such as SEC 10-K annual filings, investor presentations, and press releases, and classify them through normalized relationship types. Company information is fully reviewed annually, and changes based on corporate actions are monitored daily. The result is a comprehensive, detailed and up-to-date dataset of material intercompany relationships."

[42] *Fonar Corp. v. Gen. Elec. Co., 107 F.3d 1543, 1549 (Fed. Cir. 1997)* and *Amgen, Inc. v. Chungai Pharm. Co., 927 F.2d 1200, 1212 9Fed. Cir. 1991).*

[43] *C&F Packing Co. v. IBP, Inc., 224 F.3d 1296 (Fed. Cir. 2000).*

FactSet also provides information on the names of subsidiaries in some business relationships. I match all these names, including subsidiaries, with the names of assignees listed in patents, using the same matching procedure as I did for patents.

FactSet records licensing and supplier-customer relationships in a duplicate manner. Specifically, a firm receiving a licensing is also recorder as a customer, and a firm providing the intellectual property is also listed as a supplier. For all firm pairs involved in licensing agreements, I exclude the recordings on supplier-customer connections.

Compustat Segment. I take the supplier-customer data from Compustat Segments data set. For publicly traded firms, the data list names of the main customers, which are mostly other firms but can also be government agencies. Regulation SFAS No. 131 requires publicly traded firms to report the identity of any customer representing more than 10% of their total sales. Using Compustat Segments, Barrot & Sauvagnat (2016) constructed a data set of suppliers and customers for the U.S. publicly traded firms for the period 1976–2013. I extend their data up to 2022 using name matching and manual inspection.

USPTO Patent Re-Assignment. The data on patent re-assignment is described in Graham et al. (2018). I leave only re-assignment of patents between companies. Formally, I leave only transactions with 'convey_ty' equal to 'assignment'. I clean firm names in the same way as for patents. Then I match these data to patent assignees.

All three data sets — FactSet, Segments, and Re-Assignment — provide information on the date of a transaction or a relationship. These dates are self-reported, so they might not correspond to the dates of actual relationships. For each firm-pair, I find the minimum and the maximum of years in which the relationship between firms was active. I leave only relationships active between 2003 (the first year in FactSet) and 2022.

I group certain relationships into more aggregated groups. Table C1 in Appendix C provides a summary of all relationships.

Lex Machina. Lex Machina offers comprehensive data on federal litigation involving patents and trade secrets. Each case entry in the database includes details such as the names of plaintiffs, defendants, and any third parties involved. In cases related to patent litigation, the database also lists the patents at issue. Additionally, the data indicate whether a given case has overlapping claims with trade secret litigation.

In my analysis, I identified $1,092$ cases that featured a total of $2,541$ patents and were involved in both patent and trade secret litigation. These cases were filed from 2001 to 2021.

China Shock. Autor et al. (2020) provides a measure of the exposure to import competition from China at the main group level in the US Patent Classification (USPC). I use these data in Section 4.2.

B.2 Details to Section 2: Concentration of Patent Citations

The concentration measure from Figure 1 is constructed in the following way. First, I identify the top 1% of the most cited patents within each grant year and technological class. Second, for these patents I compute the share of citations coming from the most citing firm. Finally, I aggregate these measures within and across technological classes.

The first step is to identify the top 1% of the most cited patents. Denote $y_{km} = 1$ if patent $m \in \mathcal{P}$ makes a citation to patent $k \in \mathcal{P}$ and $y_{km} = 0$ otherwise, where $\mathcal{P}$ is the set of all granted patents. Each patent has an assignee (owner) or, in rare cases (around 3%), multiple assignees. For the majority of patents, the assignee is a corporate firm but it can also include universities, government agencies, and individual inventors. In the second step, I compute the distribution of citations across different organizations, so I exclude citations from individual inventors and patents with missing assignee information.[44] Each patent $k \in \mathcal{P}$ has a grant year t_k^g and a primary technological class c_k. I define the technological classes at the group level in the Cooperative Patent Classification (primary class is a class listed first in the patent file). Denote the set of all groups in Cooperative Patent Classification (CPC) system by $\mathcal{CPC}$. For each patent $k \in \mathcal{P}$, I compute the number of citations within a 5-year window from a grant day

$$n_k = \sum_{m \in \mathcal{T}_k} y_{km} \text{ where } \mathcal{T}_k = \{m \in \mathcal{P} : \text{Grant Date}_m - \text{Grant Date}_k \leqslant 5 \cdot 365 \text{ Days}\} \tag{B.1}$$

For each grant year t and technological class c, I define the set of all granted patents receiving at least one citation

$$\Omega_{t,c} = \{k \in \mathcal{P} : t_k^g = t \text{ and } c_k = c \text{ and } n_k > 0\}$$

and take the top 1% of patents in terms of the number of citations within this set. Denote it by $\Omega_{t,c}^{top}$, and define the *"Main"* sample as top patents for all years and technology classes:

$$Main = \left\{\Omega_{t,c}^{top}\right\}_{c \in \mathcal{CPC} \text{ and } t=1976...2014} \tag{B.2}$$

In the second step, for each patent in the *Main* sample I compute the share of citations coming from the most citing organization. To account for patents with multiple assignees, I define a weighted citation as $y_{km}^w = y_{km}/F_m$ where F_m is the number of assignees for patent m. Define the number of citations to patent $k \in \mathcal{P}$ from organization i as

$$n_{k,i} = \sum_{m \in i, m \in \mathcal{T}_k} y_{km}^w$$

[44] Formally, I set $y_{km} = 0$ where patent m belongs to an individual inventor and does not have have an assignee information.

where $m \in i$ means that organization i is an assignee for patent $m \in \mathcal{P}$. Then, the concentration measure for patent $k \in \mathcal{P}$ is

$$\mathcal{C}_k = \max_i \left\{ \frac{n_{k,i}}{n_k} \right\}$$

In the third step, I aggregate these measures. Specifically, within each grant year (t) and technological class (c) I compute a simple average across patents[45]

$$\mathcal{C}(t,c) = \frac{1}{|\Omega_{t,c}^{top}|} \sum_{k \in \Omega_{t,c}^{top}} \mathcal{C}_k \tag{B.3}$$

where $|\Omega_{t,c}^{top}|$ is the number of patents in $\Omega_{t,c}^{top}$. Then I aggregate across technological classes using the weighted average of $\mathcal{C}(t,c)$ where weights are defined by the number of patents in each $\Omega_{t,c}^{top} \neq \varnothing$

$$\mathcal{C}(t) = \sum_{c \in \mathcal{CPC}} \frac{|\Omega_{t,c}^{top}|}{\sum_{c \in \mathcal{CPC}} |\Omega_{t,c}^{top}|} \mathcal{C}(t,c) \tag{B.4}$$

The variable $\mathcal{C}(t)$ for $t = 1976 \ldots 2014$ is shown in Figure 1.

B.3 Robustness to Section 2: Concentration of Citations

Panel (a) in Figure C3 shows that the concentration of citations is similar if we restrict the sample to corporate patents only. I also consider different thresholds for the most cited patents: top 5% and 10%. Finally, I exclude the sample of citing patents that are assigned to superstar firms. Specifically, in each year and group level in Cooperative Patent Classification I find top 1% of firms in terms of the number of patents, and exclude their patents from the sample of citing patents. Panel (b) shows that the results are robust if one uses the citation-weighted average or the median instead of the average to aggregate concentration measures within technological classes. Panel (c) shows that the results are the same when I exclude self-citations of firms to itself, so the concentration is driven by citations between firms rather than self-citations. Kuhn et al. (2020) argue that the quality of citations as a measure of knowledge flows has declined over time due to a small number of patents responsible for a large share of backward citations. Panel (c) shows that the results on the concentration are robust when I exclude top 1% of patents in terms of the number of backward citations. Finally, I exclude citations between patents sharing a common law firm to ensure that the concentration is not driven by lawyers citing themselves. I also group citations from patents from the same within-country family (continuations, continuations-in-part, divisionals) as a single citation. This ensures that

[45] The results are robust if instead of a simple average I use a citation-weighted average, or median instead of an average.

the rise in concentration is not driven by increasing patent families. Finally, for the period after 2001 I separate citations made by patent examiners and non-examiners. Figure C4 shows the concentration based on citations from patent examiners is around two times lower than the concentration based on citations from non-examiners.[46]

I also check whether citations are not driven by lawyers. Specifically, I track citations of lawyers who worked in multiple firms similar to the movement of inventors in Section B.5. The only difference is that there is no data for the location of lawyers, so I consider patents which are filed by the same lawyer in at least two companies, have a similar application period, and are classified to the same main subgroup in Cooperative Patent Classification system. Figure C5 shows that the actual concentration of citations across firms is around 95% within lawyers who represented similar companies. It is significantly higher relative to the 95th quantile of the concentration measure where citation rates are equated across companies within a lawyer.

B.4 Details on Monte-Carlo Simulations

The details on the Monte-Carlo simulations are the following. First, I divide all granted patents into disjoint groups based on common observational characteristics. Then, for each cited patent I find all patents sharing the same observational characteristics as the citing patents. Second, for each patent I randomize citations across similar patents to equalize citation rates across firms. Third, I randomize citations to equalize both citation rates and patenting across firms. For each patent, I compute the concentration measure on the simulated sample. I repeat this procedure 300 times to construct the distribution of concentration measures. Finally, I aggregate different moments of this distribution in a way similar to the actual concentration measure.

B.4.1 Step 1: Find Patents with Similar Characteristics

In Section 3.3, I use application years and technological classes for patent characteristics. For technological classes, I use the main subgroup level in CPC. Denote by t_k^a and $\tilde{c}_k$ an application year and a technological class of patent k. For each patent, I divide its citing patents based on their characteristics $g = (t^a, \tilde{c})$. Then, I find all patents, citing and non-citing, with the same characteristics.

As a robustness, in Section 3.3 I also group patents based on their textual similarity of abstracts. Specifically, as in the main analysis, for each cited patent I find all patents with the same characteristics $g = (t^a, \tilde{c})$ as the citing ones. Then, for all these patents, citing and non-citing, I compute the vector embedding of their abstracts using BERT model developed by Google. I use the version of BERT model called "all-MiniLM-L6-v2". The embedding is

[46] The USPTO started to separate examiner and non-examiner citations only around 2001.

a vector that provides a mapping from text to a numerical representation. Then, I compute pairwise cosine similarities between patents that actually make citations. I find the minimum of this similarity. Next, for each non-citing patent I compute all pairwise cosine similarities with all citing patents. I leave a non-citing patent in the sample only if its similarity with at least one citing patent is greater or equal to the minimum similarity among citing patents. In this exercise, patent characteristics are specific to a cited patent. Denote

In Sections 3.4 – 3.5, I also use a geographical location of the majority of inventors for patent characteristics. Denote the location for patent k by ℓ_k. I define the geographical location at the state level if an inventor is located in the U.S., and at the country level if an inventor is located outside the U.S. For example, the location for an inventor living in Cambridge, MA, USA is (USA, MA), and for an inventor living in Berlin, Germany is $Germany$. If a patent has several inventors in different locations, I define the location for a patent based on the location of the majority of inventors. In the case of a tie, I take the location based on the alphabetical order. In this case, the set of patent characteristics is $g = (t^a, \tilde{c}, \ell)$.

B.4.2 Step 2: Equalize Citation Rates

Denote the number of citations to patent k from patents with characteristics g by $n_k(g)$. For each g, I equate citation rates across firms. Formally, for each patent characteristic g I randomize $n_k(g)$ citations across all patents that have characteristics g and satisfy sample selection constraints from Section B.2.[47] As a result of this randomization, every patent can make a citation with the same probability. Denote the total number of such patents, citing and non-citing, by $N_k(g)$. Then every patent makes a citation with probability

$$p_k(g) = \frac{n_k(g)}{N_k(g)}$$

B.4.3 Step 3: Equalize Citation Rates and Patenting

This Monte-Carlo exercise is similar to the previous one except the details on the randomization of citations. To equate citation rates and patenting, I assume that $n_k(h)$ citations are allocated randomly to firms with the same probability. In other words, I assume that all firms have the same number of patents. Formally, denote by $\mathcal{F}_k(g)$ the set of firms that have at least one patent with characteristic g. Each citation out of $n_k(g)$ is randomly allocated to firm $j \in \mathcal{F}_k(g)$ with probability

$$\frac{1}{|\mathcal{F}_k(g)|}$$

[47] Citations should be within a 5-year window from a grant day of a cited patent. I also exclude citations from patents assigned to individual inventors or with missing assignee information

where $|\mathcal{F}_k(g)|$ is the number of firms in $\mathcal{F}_k(g)$.

B.4.4 Step 4: Aggregation

I repeat the randomization procedures 300 times, and each time I compute the counterfactual concentration of citations for patent k. For an exercise with equal citation rates, denote the concentration measure for patent k in round s by $\mathcal{RC}_{k,s}$. For an exercise with both equal citation rates and the number of patents, denote the concentration measure for patent k in round s by $\mathcal{PC}_{k,s}$. I compute the median and the 95th quantile based on the distribution $\{\mathcal{RC}_{k,s}\}_{s=1}^{300}$ and $\{\mathcal{PC}_{k,s}\}_{s=1}^{300}$. These moments are denoted by $\mathcal{RC}_k(q)$ and $\mathcal{PC}_k(q)$, where q denotes quantile. Then I aggregate these measures across patents in the same way as with the actual concentration, see equations (B.3) and (B.4).

B.5 Movement of Inventors and Citation Patterns

To distinguish whether the concentration of citations is driven by firms or inventors, I track citations of inventors who worked for multiple companies. I compute the concentration measure similar to the one in Section 2 but within inventors-movers, and then I do the decomposition of the concentration measure similar to the one in Section 3.3. This exercise follows the same procedure as the Monte-Carlo exercise in Section B.4 except that the sample is restricted to inventors who worked in multiple companies, and citations are randomized within an inventor.

To increase the sample size, I consider all citations rather than the ones within a five year window. As a result, I consider the trend in citation patterns for cited granted patents until 2009, so that they have 10 years to accumulate citations. I also focus on the sample of patents granted to publicly listed firms in Compustat.[48] Moreover, I exclude patents assigned to multiple companies because it is impossible to distinguish which company an inventor represents.

For each patent, I compute the distribution of citations across inventors. I leave only patents that received at least 20 citations from one inventor. The results are robust to other thresholds. This is done in order to ensure greater variability in the concentration measure. For example, if an inventor cited a patent only one time, then this patent would always receive a citation from one firm only, and the within-inventor concentration measure would always be 100%. For each citing patent, I find all patents that were filed by the same inventor in the same U.S state or foreign country and the same main subgroup category in Cooperative Patent Classification

[48]Matching of patents to Compustat firms is cleaner in a sense that I use the data from Autor et al. (2020) to control for potential subsidiary-parent relationship. If a patent is granted to a subsidiary of a certain firm, I match it to the parent company. Therefore, when the same inventor has patents in two firms in Compustat, these firms are more likely to represent different organizations relative to cases where the same inventor has patents in two private firms or foreign firms not listed in the U.S.

system.[49] Patents should also be applied in the same time period: I find all patents applied in the period $[t_j^a, t_j^a + 2]$ where t_j^a is the application year of the citing patent. Then, I equate citation rates across firms by randomizing citations within each inventor across all these patents with similar characteristics: citing ones and control patents that are observationally similar to the citing ones. I remove citing patents where no inventor worked in at least two companies and filed for similar patents. To equate citation rates and patenting, I randomize citations across firms that had observationally similar patents filed by the same inventor. The procedures are the same as in Section B.4.

The final data set is the following. Each cited patent has at least one citing inventor who filed similar patents in multiple firms. I compute the actual concentration of citations within each of these citing inventors (if there are many). Next, I compute the same concentration in Monte-Carlo simulations where citations are allocated randomly. For each cited patent, I take the average of the concentration measures across all citing inventors-movers. This gives within-inventor actual and counterfactual average concentrations of citations for each cited patent. Then, I aggregate within and across technological classes in a way similar to Section 2. Figure 6 shows the results. The average within-inventor concentration measure is significantly higher relative to the 95th quantile of the same measure in Monte-Carlo simulations. This means that citations are driven by firms rather than inventors: inventors tend to cite different patents in various companies despite doing similar technologies.

B.6 Details to Section 3.5: Trade Secret Litigation and Concentration of Citations

Using Lex Machina data, I identify 2541 patents that were involved in both patent and trade secret litigation (for cases filed between 2001 and 2021). I exclude patents granted after 2014 to leave 5 years for accumulation of citations.

For each patent involved in trade secret litigation ("treated" patents), I find control patents that have only been involved in patent litigation, without any trade secret claims. I use four criteria, each with progressively stricter conditions, to select control and "treated" patents. First, control patents should be assigned to the same firm and have the same grant year as the "treated" patent. Second, in addition to the first condition, control patents should have approximately the same number of citations as the "treated" patent: between 0.8 and 1.2 of the number of citations that the "treated" patent receives. Third, control patents should come the same CPC group as the "treated" patent. Finally, I focus on "treated" patents assigned to plaintiffs only.

[49] If there are several inventors in the citing patent, I do this procedure for each of them.

To increase the sample size, I trace citations from all years, rather than limiting citations to a 5-year window. The reason for this approach is to ensure that patents accumulate enough citations for the computation of the concentration measure. Notice that patents involved in in trade secret litigation are not necessarily the most cited ones. Since both "treated" and control patents have the same grant years, the results are not biased due to truncation of the citation data. I compute the concentration of citations using two measures: Herfindahl-Hirschman Index (HHI) and the share of citations coming from the most citing firm ("Top Share"). For each "treated" patent, I compute the difference between its concentration and the average concentration of citations for its control patents. Then I take the average of these differences across all "treated" patents.

Next, I compute the counterfactual distribution of the difference in the concentration if citations were random. Specifically, for both "treated" and control patents I find all patents sharing the same application year, technological class (main subgroup CPC), and location of inventors. Then, I randomize citations to equalize citation rates across patents. The details are given in B.4.

To explain the importance of such randomization, consider the following example. Suppose a patent involved in trade secret litigation (the "treated" patent) receives all of its citations from one firm. Therefore, the concentration is equal to 1. All patents of this firm come from the same application year, technological class, and location of inventors. This firm is the only one who has patents with such characteristics. Therefore, even under randomization of citations the concentration for the "treated" patent would be equal to 1.

Suppose the "treated" patent has one control patent. The control patent receives an equal number of citations from two firms. Therefore, the concentration is equal to 0.5. Suppose that, as with the "treated" patent, these two citing firms specialize in their respective technologies: one firm is a patent monopolist in technology A and another firm is a patent monopolist in technology B. There are no other firms who have patents in these technologies. Therefore, even under randomization of citations the concentration for the control patent would be equal to 0.5.

The actual difference in the concentration is $1 - 0.5 = 0.5$. However, this difference is driven by specialization of citing firms and would be observed even under random citations. I show that the actual difference in the concentration of citations between patents with and without trade secret claims is significantly higher relative to the difference explained by observable patent characteristics.

B.7 Details to Section 4.2: Import Competition from China

Section 4.2 estimates how import competition from China affected citations patterns. For this exercise, I follow the methodology in Autor et al. (2020) (Appendix B.3) for the analysis at the technological class level. Specifically, I do the following steps.

First, I take the set of the top 1% of the most cited patents ($Main$ sample defined in appendix B.2). For the specification with corporate patents, I leave only patents assigned to corporate firms (both public and private). Denote by f_k the assignee of patent k and by t_k^g the grant year of patent k. I group patents into five 7-year periods based on the grant year. Formally, I define the following sets

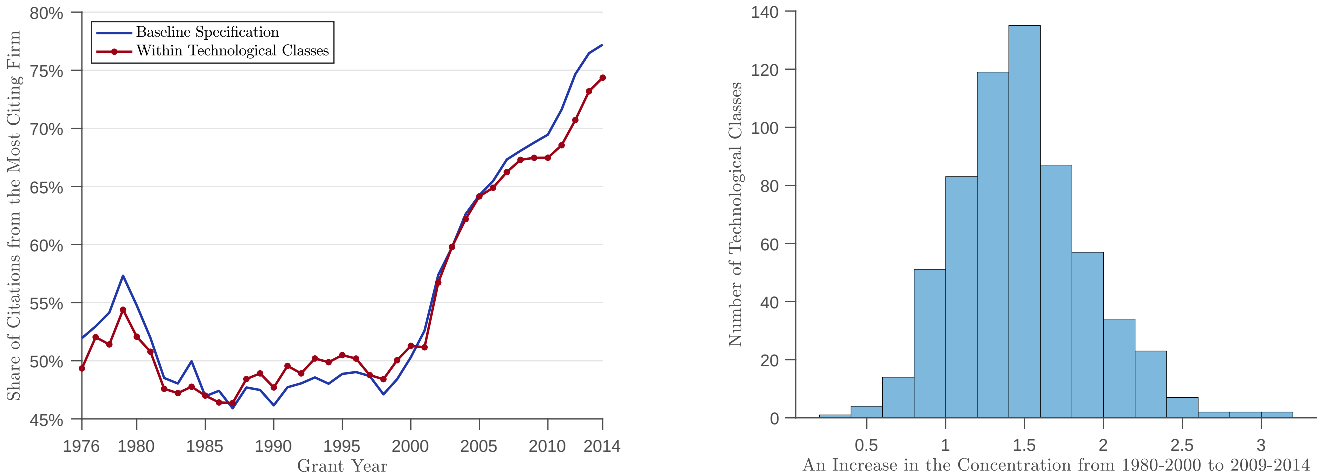
$$
\begin{aligned}
S_{1977} &= \{k \in Main : f_k \in \text{Corporate}, t_k^g \in [1977, 1983]\} \\
S_{1984} &= \{k \in Main : f_k \in \text{Corporate}, t_k^g \in [1984, 1990]\} \\
S_{1991} &= \{k \in Main : f_k \in \text{Corporate}, t_k^g \in [1991, 1997]\} \\
S_{1998} &= \{k \in Main : f_k \in \text{Corporate}, t_k^g \in [1998, 2004]\} \\
S_{2005} &= \{k \in Main : f_k \in \text{Corporate}, t_k^g \in [2005, 2011]\}
\end{aligned}
$$

where the sample $Main$ is defined in Section B.2.

Second, for each set S_t and each technology class I compute the aggregated concentration measure. I take the simple average across concentration measures. Autor et al. (2020) provides a mapping between USPC technological classes and SIC industries. Moreover, there exists a matching from USPC to NBER technology categories that will be used as controls. Therefore, for technology classes I use the USPC system. Denote the aggregate concentration measure for technology class j and set S_t by $\mathcal{C}_{j,t}$.

Third, given the constructed $\mathcal{C}_{j,t}$ the analysis proceeds as described in Section 4.2. Data construction with non-corporate patents is the same except that in the first step I leave only non-corporate patents from the $Main$ sample.

C Figures and Tables

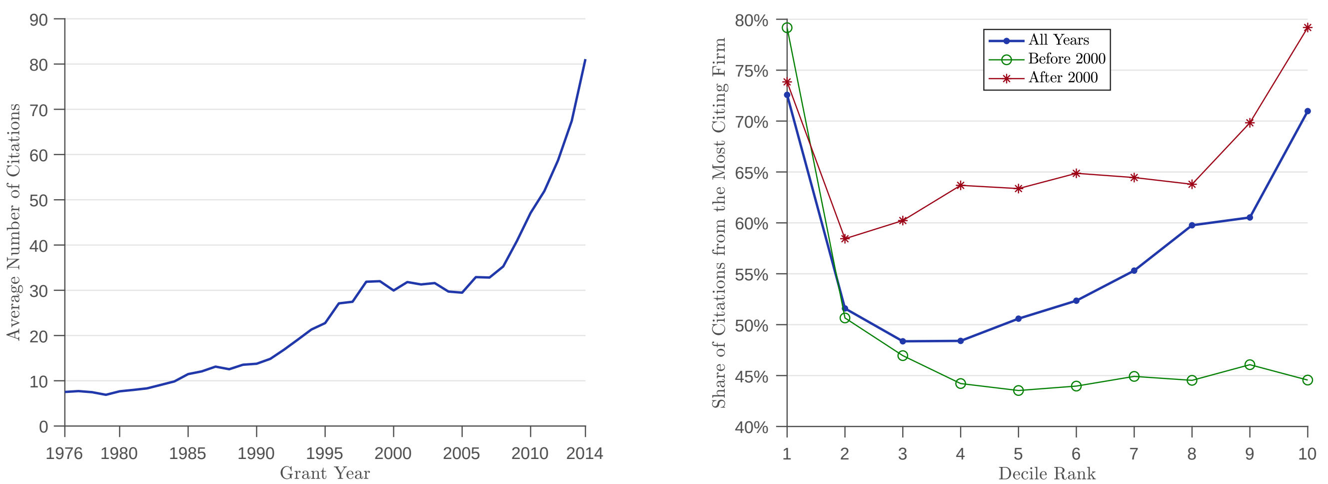


(a) Baseline vs. Within Technological Classes

(b) Distribution of the Increase Across Classes

Figure C1: Concentration of Citations Within Technological Classes

Panel (a) shows the aggregate concentration of citations that is driven by changes within technological classes. In the baseline specification, I aggregate concentration measures across classes by taking an average weighted by the number of patents in a class. The dotted red line shows the concentration in which the average across classes is unweighted. In Panel (b), for each technological class (a group category in CPC) I compute the ratio of the average concentration between 2009 and 2014 to the average concentration between 1976 and 2000. Panel (b) shows the distribution of the increase in the concentration measure across classes.

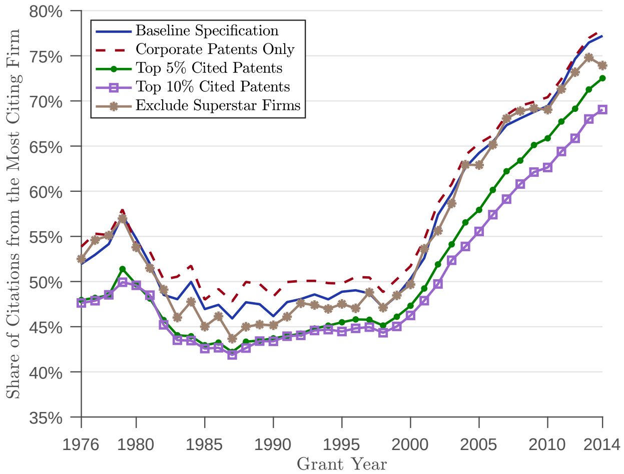


(a) Number of Citations by Years

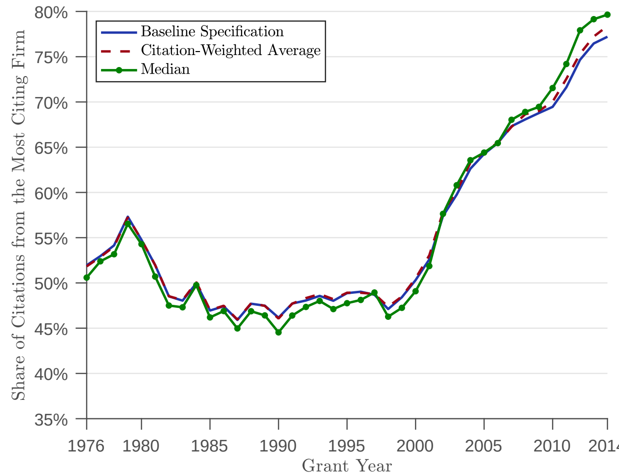
(b) Number of Citations and Concentration

Figure C2: Number of Citations and Concentration

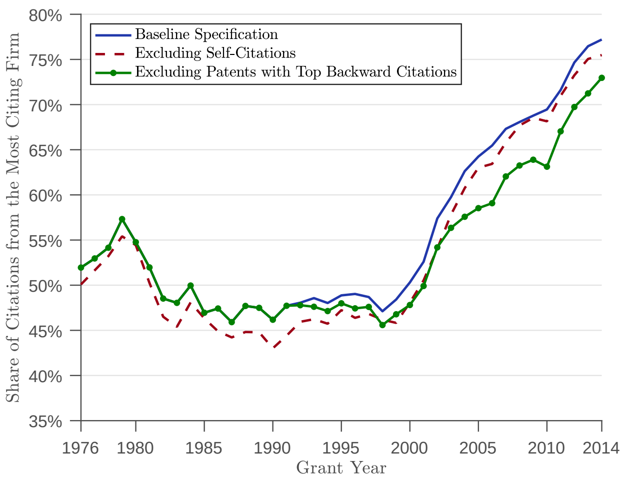
Panel (a) shows the average number of citations by years. Panel (b) shows the relationship between the average number of citations and the concentration. The dotted line shows the relationship based on patents granted in all years. The lines with circles and asterisks show the results for patents granted before and after 2000, respectively.



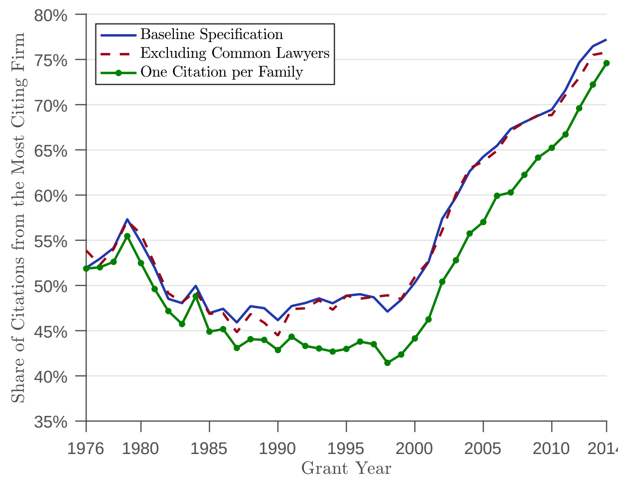
(a) Corporate patents, top 1%, top 10%, exclude superstar firms in patenting



(b) Alternative aggregation: citation-weighted average and median within classes.



(c) Exclude Self-Citations and top patents in terms of backward citations.



(d) Exclude Citations from common lawyers and group patents from one family.

Figure C3: Robustness for the Concentration of Patent Citations

These figures show robustness exercises for the concentration measure in Figure 1. Panel (a) shows that the concentration of citations in the sample of corporate patents only. I also consider different thresholds for the most cited patents: top 5% and 10%. Finally, I exclude citations from superstar firms in terms of the number of patents. Specifically, in each year and group level in Cooperative Patent Classification I find top 1% of firms in terms of the number of patents, and exclude their patents from the sample of citing patents. Panel (b) shows the results if one uses the citation-weighted average or the median instead of the average to aggregate concentration measures within technological classes. Panel (c) shows the concentration in the sample without self-citations of firms to themselves. It also shows the concentration in the sample without top 1% of patents in terms of the number of backward citations. Figure (d) shows the results in the sample without citations between patents sharing a common law firm. I also group citations from patents from the same within-country family (continuations, continuations-in-part, divisionals) as a single citation.

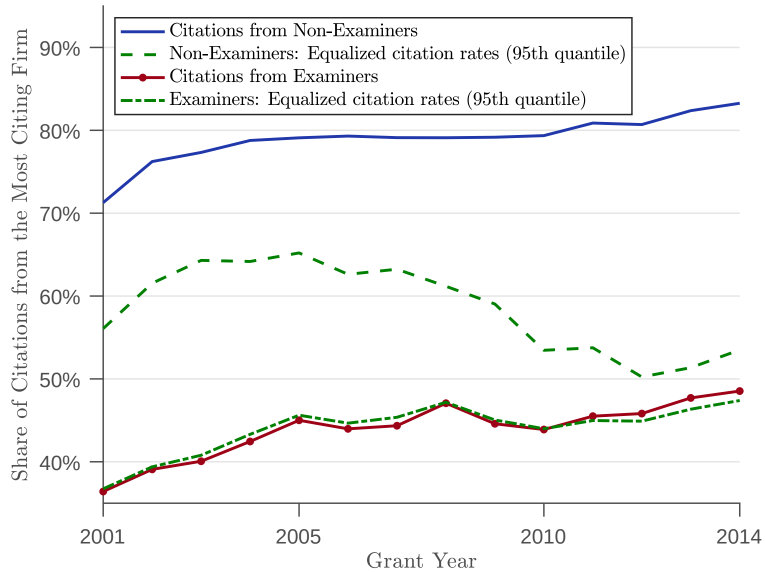


Figure C4: Concentration of Citations: Examiners vs. Non-examiners

This figure shows the concentration of citations across firms in which I separate citations from examiners and non-examiners. The USPTO started to distinguish citations from examiners in 2001. The dashed lines show 95th quantiles of the same measures in Monte-Carlo simulations in which citation probabilities are equalized across firms within the same (application year, technological class, location of inventors), see details in Appendix B.4.

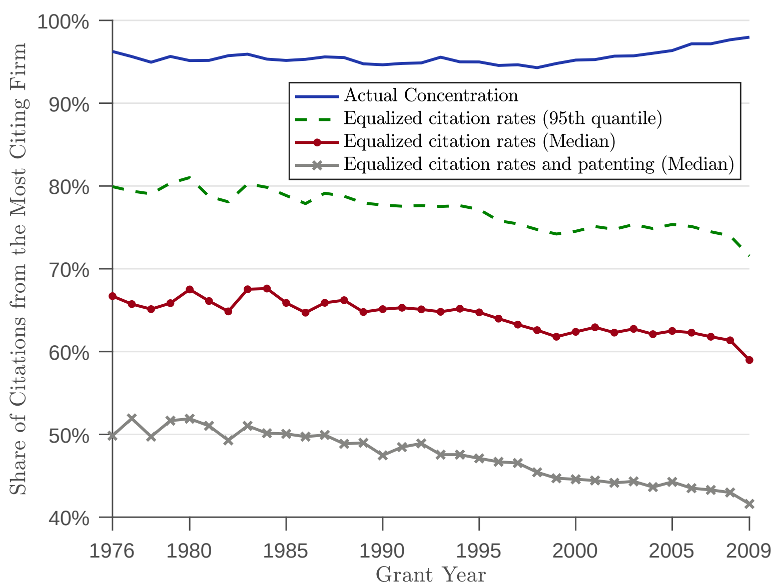
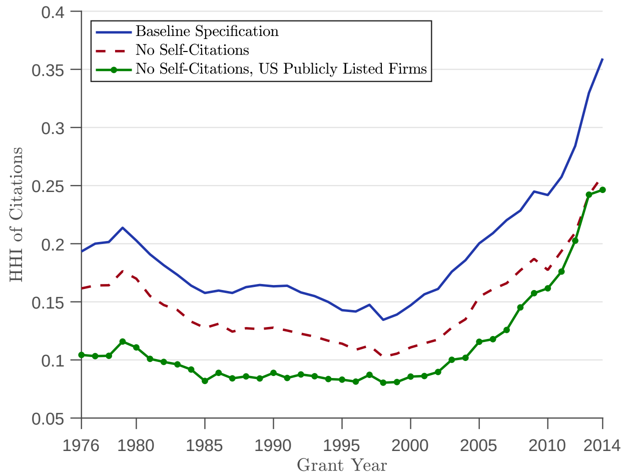
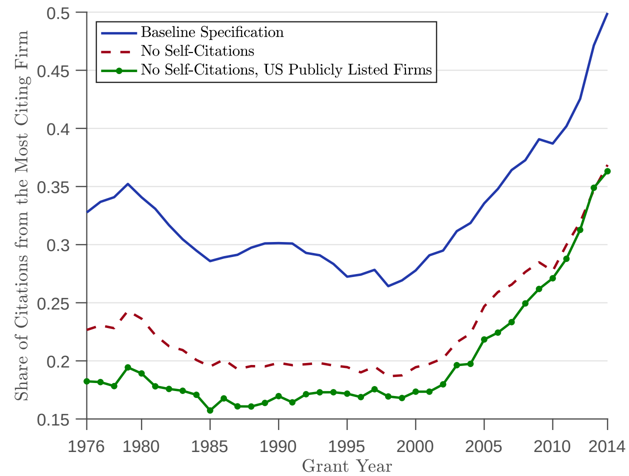


Figure C5: Concentration of Citations for Lawyers Who Represented Several Firms

This figure shows the concentration of citations across firms within lawyers who represented multiple companies. I compare patents with similar characteristics: the same main subgroup level in Cooperative Patent Classification and application time (within 2 years from the citing application year). The solid line shows the actual aggregate within-lawyer concentration of citations across firms measured by the share of citations coming from one firm only. The dashed line shows the 95th quantile of the same measure in Monte-Carlo simulations where citations are allocated randomly within a law firm.



(a) Herfindahl-Hirschman Index



(b) Share from the Most Citing Firm

Figure C6: Concentration of Citations at the Firm Level

Panel (a) shows the results for the concentration defined by the Herfindahl-Hirschman Index of citations across firms, see equation (2.2) on page 9. The solid line shows the baseline concentration using all citations. The dashed line shows the concentration without self-citations. The dotted line shows the concentration based on citations between US publicly listed firms. Panel (b) provides a robustness check for the firm-level concentration defined as the share of citations from the most citing firm. For each firm in a year, I define the concentration based on citations within five years to the firm's patents granted in this year. The aggregate measure is defined as the average concentration across firms weighted by the number of citations.

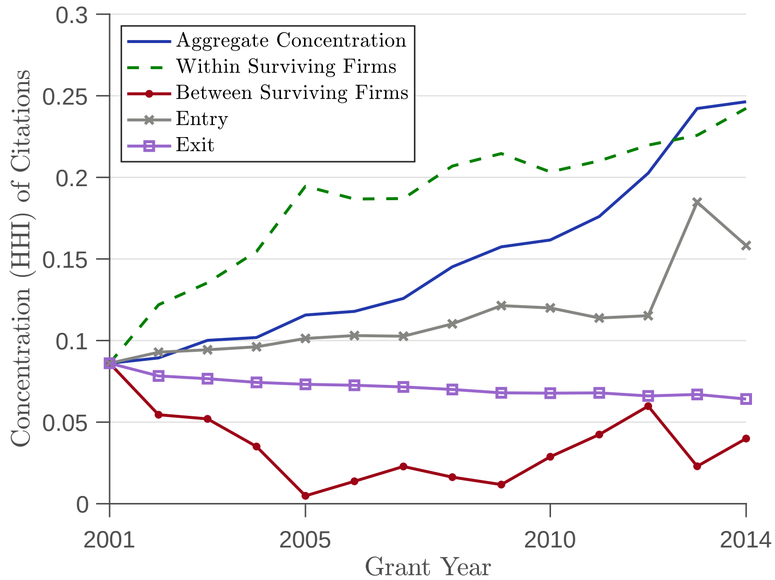
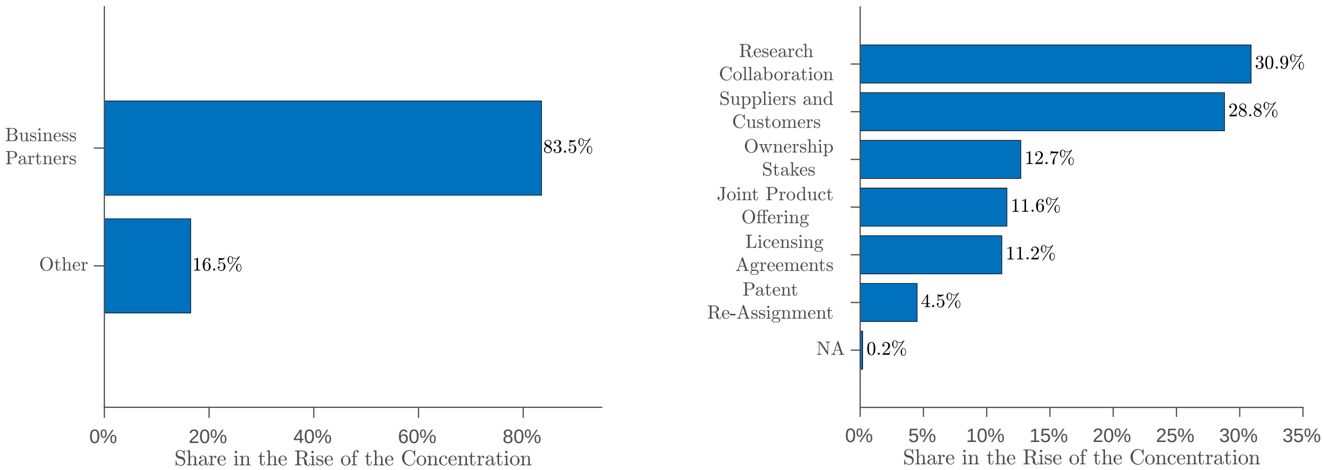


Figure C7: Concentration Over Time: Within-Firm, Between-Firm, Entry, and Exit

This figure decomposes the rise in the concentration of citations, defined in equation (2.2) on page 9, from 2001 to 2014 into within-firm, between-firm, and entry/exit components (Melitz & Polanec 2015). For each component, I plot how the average concentration would evolve if it were driven by this component only. For years t and $t + s$, the decomposition considers changes within and between firms that received citations both in t and $t + s$ ("surviving firm"). Additionally, it accounts for the impact of firms that stopped receiving citations by $t + s$ ("exiting firm") and those that began receiving them in $t + s$ ("entering firms"). Around 97.5% of the rise in the concentration from 2001 to 2014 is explained by changes in the concentration within surviving firms. If the concentration within these firms stayed constant, the average concentration among surviving firms would decline by 28.8% due to re-allocation of citations from firms with high concentration to firms with low concentration of citations. On average, entrants have higher concentration relative to surviving firms, and they explain 45% of the rise in the aggregate concentration. Exiting firms also have higher concentration relative to surviving ones, and their exit contributes to a decline in the average concentration (13.7%).



(a) Citations by Relationship Between Firms

(b) Citations Within Partners

Figure C8: Role of Partners in the Rise of the Concentration

This figure shows the share of the rise in the concentration of citations that can be explained by business partners. Panel (a) shows the aggregate role of partners. Formally, $\frac{\mathcal{H}^{\mathcal{P}}_{2014}-\mathcal{H}^{\mathcal{P}}_{2001}}{\mathcal{H}_{2014}-\mathcal{H}_{2001}} \approx 0.835$, where $\mathcal{H}_t$ is the aggregate concentration from equation (2.2), and $\mathcal{H}^{\mathcal{P}}_t$ is the aggregate concentration from partners based on equation (2.3). Panel (b) shows a similar decomposition within partners across different types of relationships.

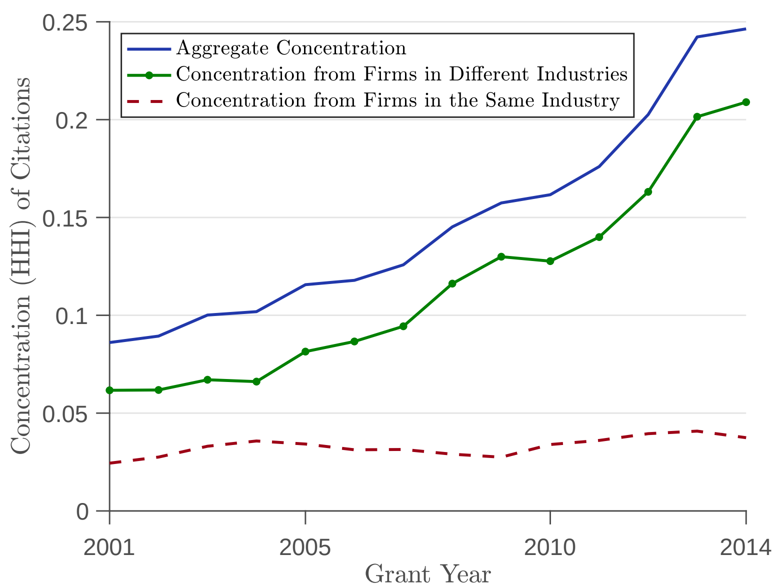
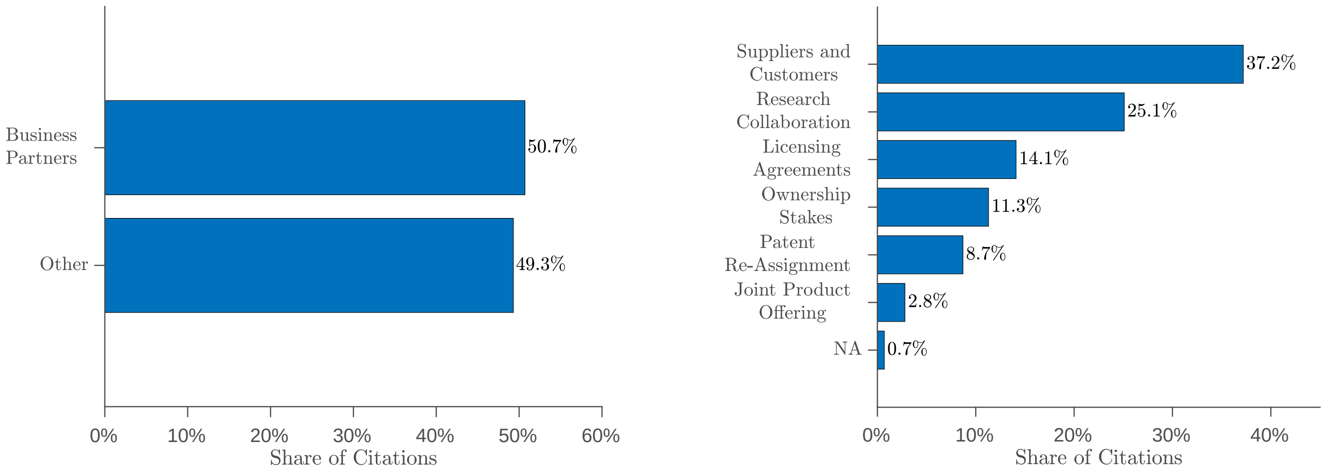


Figure C9: Decomposition of the Concentration Based on Industries

This figure decomposes the concentration into the roles of firms from the same industry as a cited firm and firms from other industries. The decomposition is similar to the one in equation (2.3): instead of partners and other firms I define cited-citing firm pairs as coming from the same 4-digit Standard Industry Classification Industry or from different industries.



(a) Citations by a Relationship Between Firms

(b) Citations Within Partners

Figure C10: Distribution of Citations, Robustness to Figure 2 With a Broader Sample

This figure shows the distribution of patent citations across different types of relationships between cited and citing firms. In Figure 2, I consider citations between publicly traded US companies. In the sample of firms for this figure, only one firm is required to be a publicly traded US company. Panel (a) shows the share of citations coming from business partners. Panel (b) shows the distribution of citations coming from business partners across different types of partners.

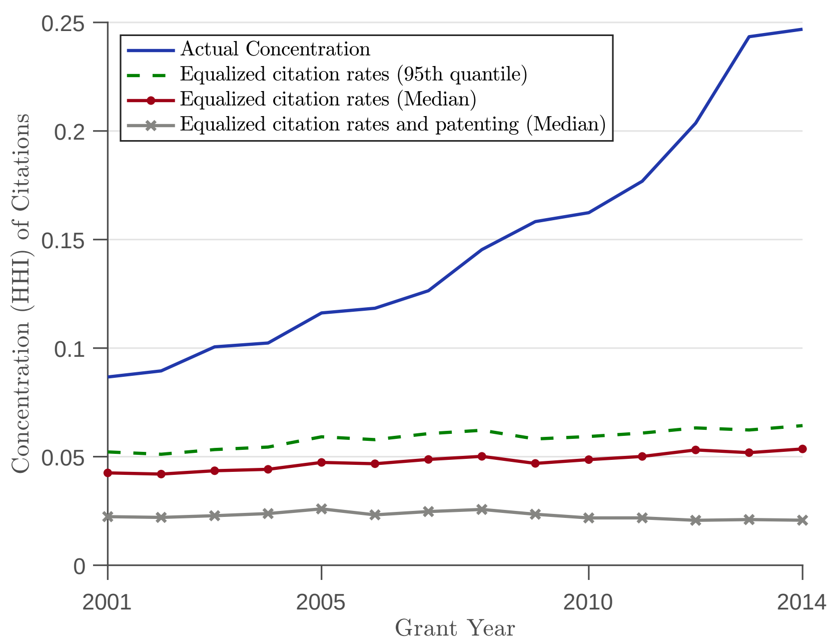


Figure C11: Decomposition of the Firm-Level Concentration of Citations

This figure does a decomposition similar to Figure 5 for the firm-level concentration of citations. There are two main differences from the computation in Figure 5. First, I allocate citations randomly within the set of publicly traded companies, rather than across all firms. The random allocation of citations follows the procedure described in Appendix B.4. Second, I aggregate citations at the firm level, as defined in Section 2.3.

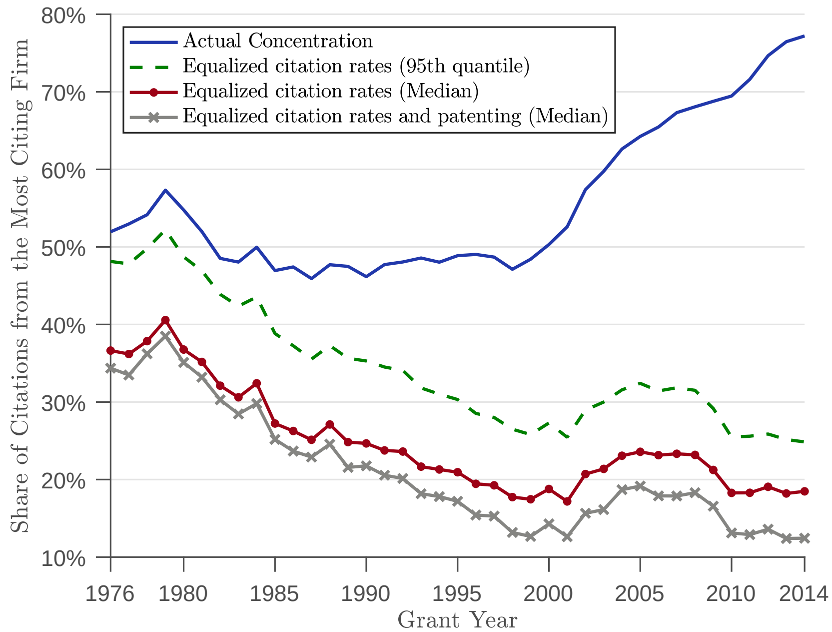


Figure C12: Decomposition of the Concentration of Citations Based on the Textual Similarity Between Patents

This figure does a decomposition similar to Figure 5 using textual similarity of patents. Specifically, in addition to application years and technological classes, I also control textual similarity of abstracts between non-citing patents and citing patents. I use BERT model to measure textual similarity. The details are given in Appendix B.4.

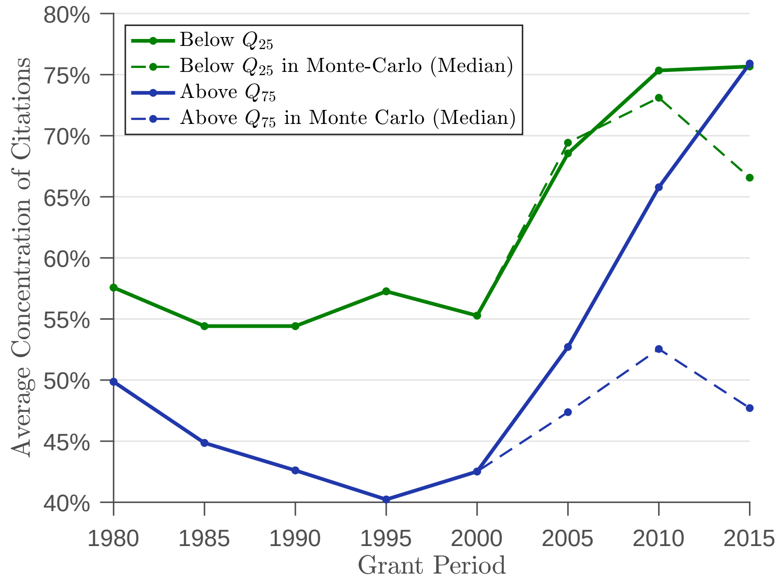


Figure C13: Trade with China. Actual Concentration vs. Counterfactual One with Equalized Citation Rates

This figure shows the concentration measure from Section 2 (Figure 1) for different technological classes divided into quartiles based on their exposure to import competition from China, ΔIP_{i2} in (4.2). The solid lines show the concentration measures over time for the technologies most (the blue line) and least (the green line) exposed to the competition from China. The dashed lines show the counterfactual concentration measures that are evolved due to changes in patenting across firms holding firms' citation rates fixed at the level of 2000. Formally, for patents granted after 2000 I compute the median concentration measures in the in the Monte-Carlo simulations where citations are allocated randomly across patents sharing the same application years, technological classes, and locations of inventors (see Appendix B.4). I add the difference between the actual and the counterfactual concentration measures in 2000.

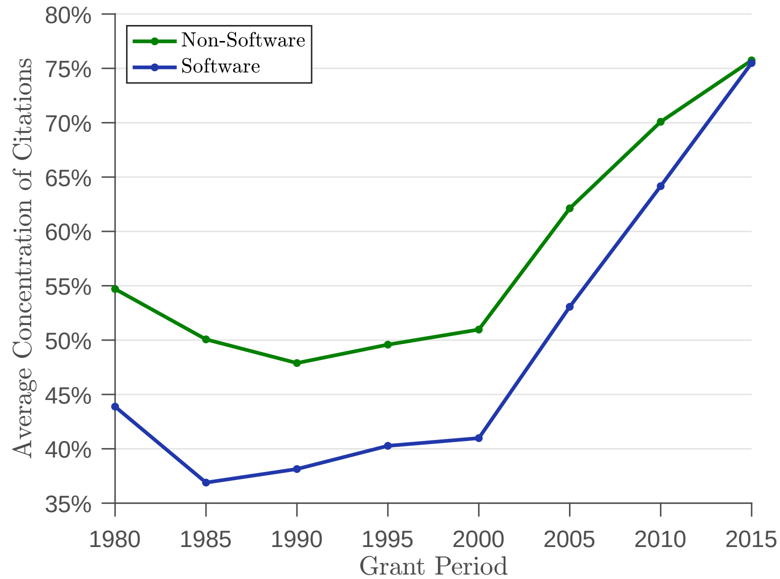


Figure C14: Concentration of Citations for Software and Non-Software Technologies

This figure shows the concentration measure from Figure 1 in which technological classes are separated into classes with software and non-software patents. Specifically, Graham & Vishnubhakat (2013) define subsclasses in the US Patent Classification (USPC) associated with software technologies. I define "software" technological classes as classes with more than 50% of software subclasses.

Table C1: Types of Business Partnerships Between Firms

Type	Data	Description from FactSet	Group
Customer	FactSet Segments	Entities to which the source company sells products/services.	Suppliers and Customers
Supplier	FactSet Segments	Entities from which the source company purchases goods or services.	Suppliers and Customers
Manufacturing	FactSet	Entities who provide paid manufacturing services to the source company.	Suppliers and Customers
Marketing	FactSet	Entities who provide paid marketing and/or branding/advertising services to the source company.	Suppliers and Customers
Distribution	FactSet	Entities whom the source company pays to distribute this company's products/services.	Suppliers and Customers

Continued on next page

Table C1 – continued from previous page

Type	Data	Description	Group
In-Licensing	FactSet	Entities from whom the source company license products, patents, intellectual property, or technology	Licensing
Out-Licensing	FactSet	Entities to whom the source company licenses products, patents, intellectual property, and technology; also entities where the source company is paid by the target entity, commonly upfront and in periodic future payments.	Licensing
Research Collaboration	FactSet	Entities collaborating with the source company for research and development, generally for new product development, common between science companies and between technology companies. This designation is applicable for products in development, not marketed.	Research Collaboration
Equity Investment	FactSet	Entities in which the source company owns an equity stake. This designation applies only when the source company owns equity in another company - i.e. working interests, royalties, property, or well claims do not qualify for the Equity Investment designation.	Ownership Stakes
Investor	FactSet	Entities which own equity in the source company.	Ownership Stakes
Joint Venture	FactSet	Entities where the source company jointly owns a separate company with one or more companies.	Ownership Stakes

Continued on next page

Table C1 – continued from previous page

Type	Data	Description	Group
Integrated Product Offering	FactSet	Entities with whom the source company agrees to bundle standalone products/services of each company and are then marketed together as one offering. No money is exchanged upfront, and costs, risks, and profits are shared.	Integrated Product Offering
NA	FactSet	Partners with an unknown relationship. They are responsible for only 0.7% of citations among partners.	NA
Patent Re-Assignor	USPTO	Entity who transfers its right, title, and interest in a patent or patent application to an assignee.	Patent Re-Assignment
Patent Re-Assignee	USPTO	Entity who receives the right, title, and interest in a patent or patent application from an assignor.	Patent Re-Assignment

Note: This table describes all business partnerships used in Section 2.3. The description for all relationships except NA and Patent Re-Assignment comes from FactSet's Data and Methodology Guide. In the analysis of citations between partners, I use the grouping of relationships from the last column.

Table C2: Trade Secret Litigation and Concentration of Citations

Concentration Measure	HHI				Top Share			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ Concentration	2.11	4.78	6.19	8.49	2.53	5.07	4.53	5.99
	(0.44)	(0.59)	(1.25)	(1.33)	(0.62)	(0.82)	(1.57)	(2.14)
Controls:								
Same Firm and Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Same N of Citations		Yes	Yes	Yes		Yes	Yes	Yes
Same Tech Class			Yes	Yes			Yes	Yes
Plaintiffs Only				Yes				Yes

Notes: This table shows the average difference in the concentration of citations between patents involved in litigation with and without trade secret claims. Columns 1–4 measure the concentration as the Herfindahl-Hirschman Index (HHI), and columns 5–8 — as the share of citations coming from the most citing firm. All measures are multiplied by 100. The numbers in brackets show standard errors of the difference in the concentration of citations when citations are allocated randomly within the same application years, technological classes, and locations of inventors (see Appendix B.4). I use four sets of controls. First, I require control patents (patents involved in patent infringement litigation but without trade secret claims) to be from the same firm and grant year. Second, I require control patents to have approximately the same number of citations as the treatment patents (patents involved in litigation with trade secret claims): between 0.8 and 1.2 of citations. Third, I require patents to be from the same CPC group. Finally, I focus on patents assigned to plaintiffs only.

Table C3: Placebo Tests: Trade with China and Increase in the Concentration of Citations

	Non-Corporate Patents					Lag Outcomes			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Δ Tech Class Exposure	−1.68	−1.58	0.90	4.33	5.04	−0.46	−0.47	−1.31	−1.42
to Chinese Imports	(1.85)	(1.81)	(1.21)	(5.33)	(5.67)	(0.64)	(0.65)	(1.10)	(1.11)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Δ Citations		Yes	Yes	Yes	Yes		Yes	Yes	Yes
2 Lags of outcomes			Yes	Yes	Yes				
11 sectors, 6 Tech				Yes	Yes			Yes	Yes
Software Patents					Yes				Yes

Notes: This table shows the results for the falsification tests in specification (4.3) and Table 4. Changes in US import exposure are instrumented with Chinese exports to non-U.S. high-income markets (Autor et al. (2020)). In columns (1)–(4), I regress the change in the concentration of citations for non-corporate patents on the changes in import competition from China. In columns (5)–(7), I regress the change in the concentration measure pre-period (pre 1991) on future changes in import exposure. Standard errors are clustered at the technology class level. All specifications are weighted by the number of Compustat-matched U.S.-inventor patents in a technology class.