Creative Construction: Knowledge Sharing in Production Networks*

Evgenii Fadeev†

May 28, 2022

Click here for the latest version

Abstract

Knowledge spillovers between firms are often measured using patent citations. I show that even the most cited patents granted before 2000 received around 50% of citations from one firm only, and this concentration has increased up to 77% by 2014. I provide additional evidence consistent with the following hypothesis: instead of spillovers citations reflect intentional sharing of trade secrets between business partners. I develop and test a theory of knowledge sharing between firms in a production network that partially explains the observed citation patterns. The theory highlights that knowledge flows depend on the market structure and contractual arrangements between firms.

*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 Maxim Boycko, Gabriel Chodorow-Reich, Emmanuel Farhi, Edward Glaeser, Gordon Hanson, Elhanan Helpman, Sergei Izmalkov, Myrto Kaloutpsidi, William Kerr, Kristina Komissarova, Robin Lee, Marc Melitz, Ariel Pakes, Ludwig Straub, Hosny Zoabi, and participants in the Harvard International, Industrial Organization, Organizational and Macro lunches. I also thank John Squires and Christopher Tehrani for discussions about the legal issues surrounding patents and trade secrets.

†Harvard Business School, email: efadeev@hbs.edu. Job Market Paper.

1 Introduction

Knowledge flows between firms are key for economic growth (Romer (1990), Grossman & Helpman (1991), Aghion & Howitt (1992)). The economic and legal rationales for patents rely centrally on these transfers: a patent owner gets monopoly rights on a technology in exchange for its disclosure to the public, so others can create follow-on innovations based on it.[1] Therefore, patents should act as a source of knowledge similar to scientific publications. Patent citations are often 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. (2021), Akcigit et al. (2021), Moretti (2021)).

This paper shows 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, the IBM's patent in Table 1 is heavily cited, but almost all of its citations come from *one firm only*, Amkor Technology Inc. In general, for each year between 1976 and 2014 I consider the set of the most cited patents filed in the U.S. Patent and Trademark Office.[2] Figure 1 shows that a patent granted between 1980 and 2000, on average, received around 50% of citations from one firm only, and this increased to around 70% in 2010s. As a benchmark, the concentration of citations for the most cited scientific publications declined from 6% to 2% over the same time period.

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%

To clarify the source of the concentration, I document additional facts about patent citations. First, I show that the concentration is primarily driven by differences across firms in the probabilities of a citation rather than in the number of patents. For instance, Amkor is responsible for the majority of citations to the IBM's patent, not because Amkor has more

[1] See Mazzoleni & Nelson (1998) and *Pfaff v. Wells Elecs., Inc.*, 525 U.S. 55, 63 (1998) where the U.S. Supreme Court stated that "the 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."

[2] Specifically, in each grant year and technology class for the period 1976–2014 I track citations within a five-year window for the top 1% of the most cited patents. Taking a five-year window controls for the truncation bias that older patents have more time to accumulate citations. Comparing patents from the same technology class controls for differences across classes in citation patterns (Lerner & Seru (2022)).

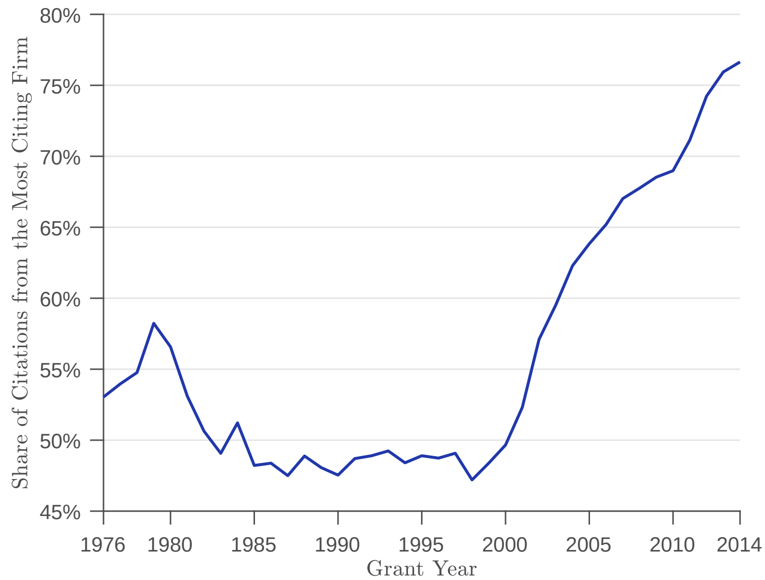


Figure 1: Concentration of Citations

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 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.

patents than other companies, but because a large share of its patents makes citations to IBM while other companies with similar patents make no citations. Second, these citation probabilities are specific to citing firms rather than to inventors working in these companies. Specifically, 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. Therefore, the explanation for the concentration of citations should be based on the relationship between firms citing each other rather than on the interactions between their inventors. Third, I show that firms responsible for the majority of citations to a highly-cited patent grow around 1% faster in sales and profits. Their citing patents generate around 15% more forward citations and around 3% greater abnormal stock market response on the grant day of a patent (Kogan et al. (2017)). Finally, I control for the legal specifics of the patent system. For example, the concentration is not driven by patent examiners or lawyers.

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. So, interpreting patent citations as being similar to scientific ones, in terms of measuring knowledge spillovers and the breadth of influence, might be incorrect. Instead, an alternative view of patent citations might be that they reflect cooperation or exchange of information between firms with close business ties. This is the broad idea that I pursue in this paper.

I argue that citations between firms are correlated with the sharing of trade secrets accompanying patents. As a result, they are concentrated because only a limited set of firms gets access to private knowledge of a patent owner. Despite the disclosure requirements of the patent system, firms frequently combine patenting with secrecy "because innovations are rarely composed of a monolithic piece of knowledge" (Anton et al. (2006)). 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)).[3] In general, firms prefer patenting for knowledge that is codified and can be reverse-engineered, and secrecy for knowledge that is tacit and easier to hide (Roin (2005), Hall et al. (2014)).

I develop and test a theory of knowledge sharing between firms that accounts for observed citation patterns. The idea behind the theory is as follows. A firm has incentives to share valuable secrets with a producer of complementary products such as an input supplier. For example, in Table 1 Amkor Technology is a supplier to IBM. In turn, the supplier might want to share these secrets with its other customers because by increasing the customers' profits it can charge them a higher input price. Depending on whether the supplier leaks private knowledge to others or not, knowledge flows are diffused or concentrated. This outcome depends on the market structure and types of contracts the supplier signs with the owner of secrets.

The key assumption in the theory is that contracts on knowledge sharing are incomplete (Arrow (1962)). Courts cannot verify the quality of secrets companies share with each other, so a firm cannot sell its secrets to a competitor when it is mutually beneficial for them. Once a buyer of secrets pays for the knowledge, a seller has little incentives to disclose them, and, conversely, once the seller reveals its secrets, the buyer might refuse to pay. In contrast to competitors, firms in a vertical relationship have incentives to share secrets because they produce complementary products and can use an input price to split the gains from knowledge sharing.

I assume that the firm with valuable trade secrets can sign two types of contracts with a supplier to prevent a leakage of its knowledge: a confidentiality agreement and exclusive dealing. Although courts cannot verify the quality of secrets firms share, they can verify the communication between parties, and a confidentiality agreement restricts the supplier's ability to discuss topics related to the firm's technology with other companies. Exclusive dealing eliminates the supplier's incentives to leak secrets by restricting its ability to sell an input to other companies.

In equilibrium, the firm and its supplier sign a contract that maximizes their joint surplus which consists of the firm's gross profits and the supplier's input sales to other firms. If

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

downstream competition is high, they sign exclusive dealing to give the downstream firm an advantage over its competitors. This mechanism is similar to the foreclosure literature (Rey & Tirole (2007)). If downstream competition is low, but the firm's competitors have a strong outside option to the supplier's input, the firm and the supplier sign a confidentiality agreement. They prefer to sell inputs to others due to low competition but to keep the firm's knowledge secret because the supplier cannot extract enough gains from sharing it with other firms. In both cases, the supplier does not leak the firm's secrets, so knowledge flows are concentrated. If competition is low and competitors have a weak outside option, a knowledge leakage to other companies generates greater gains in input sales to the supplier relative to losses in the firm's profits. Therefore, the supplier and the firm sign neither a confidentiality agreement nor exclusive dealing, and knowledge flows are more diffused.

The theory can be generalized to other types of business connections where firms engage in complementary activities. For example, I also consider the framework with a common customer buying from several suppliers. I focus on a vertical relationship between firms in order to test the theory using Compustat Segments data on supplier-customer relationships between publicly traded U.S. firms (Cohen & Frazzini (2008), Barrot & Sauvagnat (2016)).

The theory leads to several testable predictions. First, I show that a vertical supply chain relationship increases the probability of a citation between firms relative to similar firms with no production relationship. Second, firms sharing a common supplier or customer are also more likely to cite each other relative to firms with exclusive suppliers or customers. Third, the probability of a citation between firms sharing a common supplier is increasing in the bargaining position of this supplier which is proxied by its size relative to other firms in the supplier's industry. The relative size of the supplier approximates the weakness of customers' outside options of not working with it.[4]

The model also predicts that a higher level of competition leads to lower knowledge sharing. I show that the decline in knowledge sharing after 2000 (the rise in the concentration of citations in Figure 1) is greater in the technologies more exposed to import competition from China, where the U.S. imports are instrumented by Chinese exports to other high-income countries (Autor et al. (2020a)). Moreover, the competition affected citation patterns of firms but not of non-corporate organizations (e.g., universities and government agencies), so its effect on the decline of knowledge sharing is unlikely to be driven by the correlation between general technological changes and globalization.

[4]In the framework with a common customer, I assume that the suppliers' outside options of not working with it are zero. Therefore, this framework does not give empirical predictions about the suppliers' outside options, which can be proxied by the customer's relative size in its industry. Considering a more general downstream market structure would complicate the model, and I leave it for future research.

Literature

Patent citations are often used to measure knowledge spillovers between firms and inventors (e.g., Jaffe & Trajtenberg (2002), Jaffe & de Rassenfosse (2019)). The evidence in this paper indicates that patent citations are more likely to reflect intentional exchange of information between business partners rather than unintentional knowledge externalities.

On the theory side, this paper studies cooperation (creative construction) between firms who *intentionally* share private knowledge with each other to improve their joint production. The model is related to the antitrust literature on exclusionary vertical contracts (e.g., Whinston (2006)) but with a focus on the incentives to share knowledge. I complement this literature by discussing the role of confidentiality agreements. The theory in this paper has two main differences from the Schumpeterian literature on economic growth (e.g., Aghion et al. (2014)). First, the Schumpeterian literature is focused on competition between firms ("creative destruction") in the environment with knowledge spillovers while I highlight the importance of cooperation in production networks for the diffusion of knowledge across competitors. Second, the literature usually assumes that conditional on R&D knowledge spillovers are independent of the economic environment while I emphasize that intentional knowledge flows depend on the market structure.

Rosenberg (1963) was one of the first to emphasize the importance of input suppliers in the technological progress and described how in the nineteenth century specialized machine tool suppliers disseminated knowledge across industries about the metal working technology developed in the armament manufacture. The importance of vertical knowledge flows is also highlighted in the literature on R&D cooperation (e.g., Cassiman & Veughelers (2002)) and foreign direct investment (e.g., Aitken & Harrison (1999), Blalock & Gertler (2008), Alfaro-Ureña et al. (2021), Bai et al. (2021)).

Much of the literature on intellectual property (IP) protection treats patenting and secrecy as substitutes (e.g., Hall et al. (2014)). I argue that the combination of patents and trade secrets might be used to protect the same technology. This is in line with the 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 industry and drugs (e.g., Arora (1997), Price II et al. (2020)).

Arora et al. (2021) estimate how knowledge spillouts to competitors affect firms' incentives to invest in scientific research, measured by the number of publications. I complement their paper by studying how import competition from China affected the firms' choice of protection for existing trade secrets, measured by the concentration of patent citations. Both the increase in the concentration of patent citations documented in this paper and the decline in corporate science over time (Arora et al. (2019)) are consistent with the rising role of trade secrets.

The empirical tests of the theory analyze citation patterns in the production network and contribute to the literature on knowledge flows between firms (e.g., Gomes-Casseres et al. (2006)). Autor et al. (2020a) show that trade with China reduced corporate patenting in the U.S. I complement their evidence by showing that trade with China also reduced knowledge sharing between firms. The paper also contributes to the literature on patents' quality (e.g., Kogan et al. (2017), Higham et al. (2020)) by estimating it as a function of the knowledge inventors rely on.

Akcigit & Ates (2019) document the increasing share of firms' self-citations in backward citations. I complement their evidence by documenting the rising concentration of forward citations *between* firms after 2000. I analyze the relationship between these firms, how they grow depending on the knowledge they rely on, and distinguish whether the concentration is driven by firms or inventors. Akcigit & Ates (2019) also develop the model in which the *exogenous* decline in knowledge diffusion between competitors can explain multiple macro trends (e.g., the decline in business dynamism). My theoretical results with *endogenous* knowledge sharing and the empirical evidence with China shock point out that the decline in knowledge diffusion might be caused by rising competition. This result is consistent with Autor et al. (2020b) who argue that the competition level in the U.S. has increased.

The paper is organized as follows. Section 2 presents motivating evidence. Section 3 develops a theory of knowledge sharing between firms. Section 4 tests the theoretical predictions in the data. Appendix provides empirical details, robustness checks, theoretical extensions, and case studies.

2 Motivating Evidence: Concentration of Citations

This section provides several new facts about patent citations. Section 2.1 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. In Section 2.2, I show that the high concentration of citations is primarily driven by differences across firms in citation rates (probabilities) rather than by differences in the number of patents. Section 2.3 shows that inventors who move between companies significantly change their citation rates, meaning that citation probabilities are specific to firms rather than to inventors. In Section 2.4, I show that firms responsible for the majority of citations to a highly-cited patent grow faster in sales and profits and have patents of higher quality. Section 2.5 provides robustness checks, and Section 2.6 discusses the implications of the results.

Data and Background

Patents are supposed to facilitate knowledge sharing through the disclosure of information in the award. The U.S. Supreme Court 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 "the high price of . . . exclusive use."[5]

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, inventions should be patent-eligible, useful, novel, non-obvious, and the text of the application should satisfy the disclosure requirements.[6] Patent 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 the 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 a legal function of delimiting intellectual property rights on an invention.

I use the data on utility patents granted in the U.S. Patent and Trademark Office (USPTO) for the period 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. (2020a) provide a matching of patent assignees to names of publicly traded firms in Compustat data set. I extend their matching for additional years 2015–2019. Details are given in Appendix B.1.

2.1 Results: High Concentration of Patent Citations

For each year and technological class in the period 1976–2014, I track citations within a five-year window for the top 1% of the most cited granted patents. There are two reasons to focus on the set of the most cited patents. First, the value of patents is highly skewed with many 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. (2021)). This 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 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

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

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

across classes in citation patterns (Lerner & Seru (2022)). To classify technologies, I use the group level of Cooperative Patent Classification (CPC) system which has 672 groups.[7]

For each patent among the most cited ones, I compute the distribution of citations across different organizations. Denote by $n_{k,i}$ 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 from patents with missing assignee information. The concentration measure for patent k is the share of citations coming from the most citing organization:

$$\mathcal{C}_k = \max_i \left\{ \frac{n_{k,i}}{\sum_{j=1}^{M_k} n_{k,j}} \right\} \tag{2.1}$$

where M_k is the number of citing organizations. The most citing organizations are predominantly corporate firms, so I will use terms "firms" and "organizations" interchangeably. To construct an aggregate measure, for each year I take the average of patents' concentration measures within each technology class and then the average across technological classes weighted by the number of patents in a class. Appendix B.2.1 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.

Figure C1 in Appendix C shows that the increase in the concentration is primarily driven by changes within technological classes rather than the rise of technologies with high concentration 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: 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 this relationship is driven by patents granted after 2000.[8] Therefore, an increase in the concentration of citations is not driven by the decline in the number of citations. The next section discusses the sources of the concentration and its increase.

[7] The group level of CPC system is similar to the three-digit USPC system. There are several reasons to use modern CPC system instead of the more outdated USPC system. First, CPC system provides a consistent classification over time. Second, it is available for all years while USPTO stopped updating USPC from 2015. Third, CPC system is applied internationally, not only in the U.S. Finally, its classification provides more levels of disaggregation in contrast to USPC. I will use this feature in Section 4.1.

[8] The relationship holds for patents with more than 7 citations. For patents with less than 7 citations, the concentration is high due to a low number of citations. See section 2.2 for more details.

2.2 Sources of the Concentration of Citations

In this subsection, I decompose the concentration of citations into three sources: differences in citation rates (probabilities) across firms, differences in patenting across firms, and the total number of firms that could potentially make citations. The number of firms provides a natural lower bound on the concentration measure: with M firms the concentration cannot be lower than $^1/_M$. I show that the high concentration of citations is primarily explained by the differences in citation probabilities rather than by the variance in patenting across firms. For instance, Amkor Inc is responsible for 94% of citations to the IBM's patent in Table 1 because each of Amkor's patents makes a citation to the IBM's patent with a significantly higher probability relative to other firms, not because Amkor has more patents than other companies.

The details of the decomposition are the following. First, I divide all granted patents into disjoint groups based on common observational characteristics. Patents that share the same characteristics are more likely to cite similar patents. It is well-known that citation patterns depend on the geographical location of inventors (e.g., Jaffe et al. (1993)), proximity of technologies (e.g., Acemoglu et al. (2016)), and time differences between inventions (e.g., Caballero & Jaffe (1993)). So, for characteristics of patents I choose a geographical location of the majority of inventors, a detailed technology class (main subgroup level in CPC), and an application year.[9] Denote the set of all these groups by $\mathcal{H}$.

Second, for each cited patent I construct the set of patents that could potentially cite it and rely on the knowledge disclosed in it. This set includes patents with the same characteristics as the patents that actually make citations to the cited patent. Denote by $\mathcal{H}_k \subseteq \mathcal{H}$ the set of the characteristics of the patents citing patent k. For example, patents making citations to the IBM's patent in Table 1 ($k = 5877043$) are divided in $|\mathcal{H}_k| = 106$ disjoint groups based on their characteristics. Most citations (10 out of 218, all from Amkor) come from the following group ($h \in \mathcal{H}_k$): technological class $H01L23$, location of inventors Arizona, and application year 2003. Overall, there were 92 patents with these characteristics coming from firms such as Intel (62), Amkor (12), Freescale Semiconductor (5), and others (13).[10] If each of these patents made a citation to the IBM's patent with the same probability, more citations would be expected to come from Intel rather than from Amkor.

Consider citations to patent k from patents with characteristics $h \in \mathcal{H}_k$. The number of citations from firm i, $n_{k,i}(h)$, can be decomposed as follows

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

[9] The location is defined at the state level in the U.S. and at the country level for foreign inventors. For patents with inventors from multiple locations I take the location of the majority of inventors. See Appendix B.2.2.

[10] For illustration, I consider citations from all years to the IBM's patent in Table 1, not only within a 5-year window of the grant day. The aggregate results are based on citations within a 5-year window.

where $N_i(h)$ is the total number of patents with characteristics $h \in \mathcal{H}_k$ (citing and non-citing) from firm i, and $p_{k,i}(h) = \frac{n_{k,i}(h)}{N_i(h)}$ is the citation rate. I construct two counterfactual distributions of the concentration measure. First, I do Monte-Carlo simulations where $n_{k,i}(h)$ citations are allocated randomly across all patents with the same characteristics $h \in \mathcal{H}_k$. This exercise equates citation rates across firms that have patents with characteristics $h \in \mathcal{H}_k$ ($p_{k,i}(h) = p_{k,j}(h)$ for $i \neq j$) and shows the counterfactual concentration that is entirely driven by the differences in the number of patents across companies and the total number of firms. Second, I allocate $n_{k,i}(h)$ citations randomly across firms assuming that all firms with patents in $h \in \mathcal{H}_k$ have the same number of patents. This exercise equates both the citation rates and the number of patents across firms ($p_{k,i}(h) = p_{k,j}(h)$ and $N_i(h) = N_j(h)$ for $i \neq j$) and shows the counterfactual concentration that is entirely driven by the number of firms that could potentially cite patent k.[11]

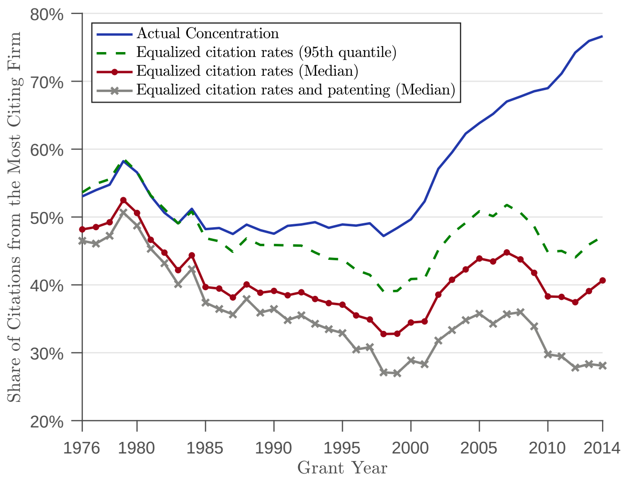
For each patent k, I do the Monte-Carlo simulations described above for all characteristics of citing patents, $\mathcal{H}_k$. Then I compute the counterfactual concentration measures in which citation rates are equal, and in which both citation rates and the number of patents are equal across firms. I aggregate these measures across cited patents in the same way as with the actual concentration in Section 2.1.

Figure 2(a) shows the decomposition of the concentration of citations. 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 measure where 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 ($\mathcal{PC}_t$). This measure provides a natural lower bound on the concentration of citations that is driven by the number of firms. The average decomposition over all years from 1976 to 2014

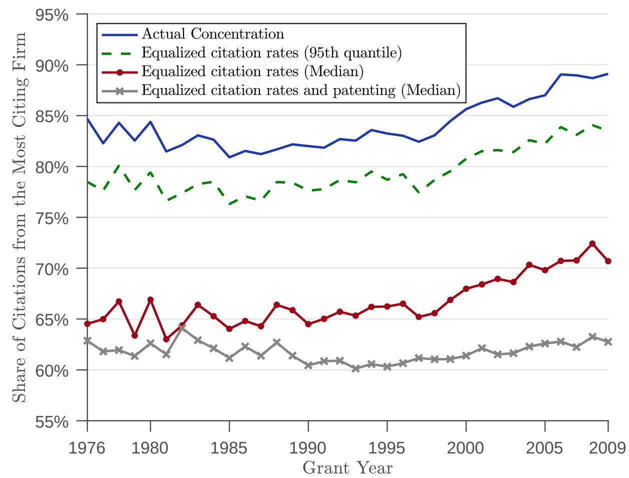
$$\underbrace{\overline{\mathcal{AC}} - \overline{\mathcal{PC}}}_{17.3\%} = \underbrace{\overline{\mathcal{AC}} - \overline{\mathcal{RC}}}_{14.1\%} + \underbrace{\overline{\mathcal{RC}} - \overline{\mathcal{PC}}}_{3.3\%}$$

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 actual and counterfactual concentrations also have different dynamics. Both $\mathcal{PC}_t$ and $\mathcal{RC}_t$ declined until 2000 and then started to increase. This means that the average number of firms that filed for

[11] Monte-Carlo exercises can be described using the urn model in probability theory (Feller (1968)). Denote the number of firms that have at least one patent with characteristics $h \in \mathcal{H}_k$ by $M(h)$. The exercise that equates citation rates only can be described as the following urn model of random drawings without replacement: $\sum_i n_{k,i}(h)$ balls are taken from the urn that contains $\sum_{i=1}^{M(h)} N_i(h)$ balls of $M(h)$ different colors with the distribution of colors $\{N_i(h)\}_{i=1}^{M(h)}$. The exercise that equates both the citation rates and the number of patents can be described as the following urn model of random drawings with replacement: $\sum_i n_{k,i}$ drawings are taken from the urn that contains $M(h)$ balls of different colors. After each drawing, the ball is returned back to the urn.



(a) Decomposition of the Concentration of Citations



(b) Decomposition of the Concentration Within Inventors Who Moved Across Firms

Figure 2: Sources of the Concentration of Citations

Figure 2(a) 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.2.2. Figure 2(b) shows the same decomposition for the concentration of citations across firms within inventors who patented for 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.3.

patents with similar characteristics increased until 2000 and then decreased.[12] The actual concentration $\mathcal{AC}_t$ stayed approximately constant until 2000 and then started to increase at a faster rate than $\mathcal{PC}_t$ and $\mathcal{RC}_t$. This implies that the role of the differences in citation probabilities was increasing during all years

$$\underbrace{\Delta\mathcal{AC} - \Delta\mathcal{PC}}_{30.2\%} = \underbrace{\Delta\mathcal{AC} - \Delta\mathcal{RC}}_{22.1\%} + \underbrace{\Delta\mathcal{RC} - \Delta\mathcal{PC}}_{8.1\%}$$

where $\Delta\mathcal{AC} = \mathcal{AC}_{2014} - \mathcal{AC}_{1976}$, $\Delta\mathcal{RC} = \mathcal{RC}_{2014} - \mathcal{RC}_{1976}$, and $\Delta\mathcal{PC} = \mathcal{PC}_{2014} - \mathcal{PC}_{1976}$.

An important limitation in the construction of the counterfactual concentrations is that

[12] The difference $\mathcal{RC}_t - \mathcal{PC}_t$ has slightly increased over time, meaning that the variance in patenting has slightly increased from 1976 to 2014.

citations are randomized across patents similar to the ones that actually make citations, excluding the ones that could have potentially made citations. For example, for the IBM's patent in Table 1 most citations come from Amkor Technology, which is a supplier to IBM. Therefore, the citations are allocated randomly across the patents that are similar to the Amkor's patents and are likely to represent non-competing technologies to IBM. So, the randomization exercise does not take into account a lot of patents from IBM's competitors that could have made citations to it. This underestimates the difference between the actual and counterfactual concentrations.

The counterfactual concentrations are constructed using the randomization of citations across patents that have the same detailed characteristics. Figure C3 (Appendix C) shows the counterfactual concentration in which citations are randomized across broader technological classes, and the concentration where inventors are not required to be from the same location.[13,14] These counterfactual concentrations are lower by 6 and 16 percentage points, respectively, relative to the baseline one. Figure C4 also shows that the results of the decomposition are robust when I group patents based on their textual similarity (Arts et al. (2018)) rather than based on the classification into technological classes.

As another benchmark, I also compute the concentration of citations for scientific publications using Microsoft Academic Graph data (Marx & Fuegi (2020)). The concentration measure is constructed in a way similar to the measure for patents. For each year and field of science in the period 1976–2014, I track citations within a five-year window for the top 1% of the most cited publications. I use Web of Science classification of scientific disciplines. Appendix B.2.5 provides details. Figure C3 shows that in contrast to patents the concentration of citations for scientific publications declined from 6% to 2% over the same time period.

2.3 Movement of Inventors

The concentration of citations might be driven by inventors rather than firms. For instance, suppose a patent received all of its citations from one inventor who always worked in one firm. Then, the concentration of citations across companies for this patent would be 100%, but it could reflect the inventor's citation practice. To separate whether citations are firm- or inventor-specific, I find inventors who filed similar patents in multiple companies. Then, I compute the concentration of citations within these inventors and do the decomposition from Section 2.2 to

[13]The main subgroup in CPC is a much more detailed level of technology classification than the one used in the literature. It 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.

[14]A patent file is a publicly available source of information to all inventors in the world.

check whether inventors' citation rates (probabilities) differ across firms.

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

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

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

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 $^{12}/_{15} = 0.8$, but 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 2.2. The details are given in Appendix B.3. Figure 2(b) 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 ($\mathcal{PC}_t^w$).[15] The actual average concentration within an 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

$$\underbrace{\overline{\mathcal{AC}}^w - \overline{\mathcal{PC}}^w}_{22.2\%} = \underbrace{\overline{\mathcal{AC}}^w - \overline{\mathcal{RC}}^w}_{17.3\%} + \underbrace{\overline{\mathcal{RC}}^w - \overline{\mathcal{PC}}^w}_{4.9\%}$$

shows that the concentration is primarily explained by the differences across firms in citation

[15]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

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

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 the 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-movers are less restricted by contractual arrangements such as confidentiality agreements and non-compete contracts, and this leads to less concentrated 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)).

2.4 Value of Citations for the Most Citing Firm

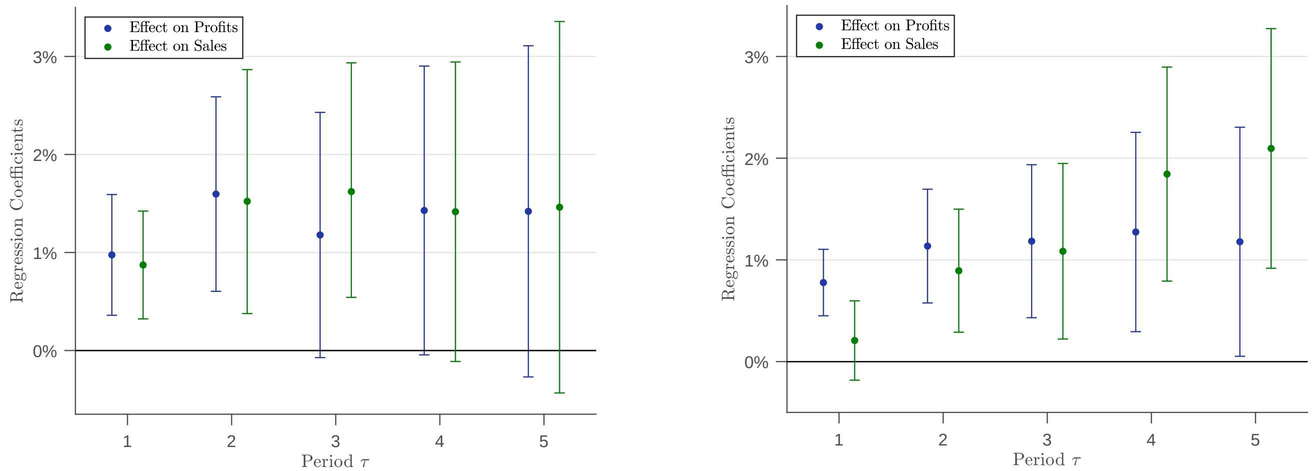
In this section, I compare the quality of (citing) patents and the growth of the most citing firms relative to other firms. I identify citing patents that satisfy two conditions. First, they are assigned to a firm that is responsible for the majority of citations to a patent from the top 1% of the most cited patents. Second, the most citing firm has a significantly higher citation rate relative to other firms (see Section 2.2). I call patents satisfying these two conditions "specific" patents. For example, patents from Amkor Technology making citations to IBM's patent in Table 1 are "specific" ones. Appendix B.2.4 provides more details.

I start with the following regression of firms' growth on different types of patents.

$$y_{f,t+\tau} - y_{f,t} = \beta_s \cdot \tilde{\ln}\left(N_{ft}^s\right) + \beta_g \cdot \tilde{\ln}\left(N_{ft}^g\right) + \beta_n \cdot \tilde{\ln}\left(N_{ft}\right) + \eta_f + \zeta_{i(f),t} + \gamma X_{ft} + \varepsilon_{ft} \qquad (2.2)$$

where $\tilde{\ln}(x) = \ln\left(x + (x^2+1)^{\frac{1}{2}}\right)$ is the inverse hyperbolic sine transformation to control for zeros in patenting. For firm f in year t, the variable $y_{f,t}$ denotes profits or sales, N_{ft}^s is the number of granted specific patents, N_{ft}^g is the number of granted patents that make at least one citation to the top 1% of the most cited patents. The inclusion of N_{ft}^g allows to distinguish whether specific patents are associated with higher firm's growth because they are responsible for the majority of concentrated citations or because they simply rely on a highly-cited patent. N_{ft} is the total number of granted patents, η_f is a firm fixed effect, $\zeta_{i(f),t}$ is an industry-year fixed effect, and X_{ft} is the set of controls that includes firm employment, firm capital stock, and a lag of firm profits or sales. The variable ε_{ft} is an error term. Appendix B.4 provides details.

The estimated coefficients β_s and β_n for $\tau = 1, \ldots, 5$ are given in Figure 3. The positive



(a) Coefficients for specific patents, β_s.

(b) Coefficients for all patents, β_n.

Figure 3: Value of Citations for the Most Citing Firm

This figure shows the estimated coefficients β_s and β_n with corresponding 95% confidence intervals for $\tau = 1, \ldots 5$ for the specification in (2.2). The positive coefficient $\beta_n > 0$ means that higher number of patents is associated with higher firm's growth in sales and profits. The positive coefficient $\beta_s > 0$ indicates that given the same number of patents the growth rate is higher for firms with greater share of specific patents. Blue lines show the results for the specification with growth in profits, and green lines for specification with growth in sales. The left panel shows estimates for β_s, and the right panel shows the estimates for β_n. Standard errors are clustered at the firm and year levels. Details are given in appendix B.4.

coefficient $\beta_n > 0$ means that higher number of patents is associated with higher firm's growth in sales and profits, which is a standard result in the literature. However, in addition to this result the positive coefficient $\beta_s > 0$ indicates that given the same number of patents the growth rate is higher for firms with a greater share of specific patents. Therefore, specific patents are more valuable to firms relative to non-specific ones.

One of the differences between specific and non-specific patents might be related to their quality. I explore whether specific patents are of higher quality relative to non-specific ones using two popular measures of patent quality used in the literature (Higham et al. (2020)). The first measure is the number of forward citations relative to the average number of citations for patents granted in the same year and technology class (group level CPC). Citations are taken within a 5 year from the grant day. I call it a scientific value of a patent. The second measure is based on the abnormal stock market return on the grant day of a patent for the firm that owns it (Kogan et al. (2017)). I extend Kogan et al. (2017) estimates for years up to 2019 using their parameter estimates for the baseline measure (see formula (4) in Kogan et al. (2017)). I call it an economic value of a patent.

I run the following regression at the patent level

$$\ln\left(Q_k\right) = \beta_s \cdot \mathbb{I}_k^s + \beta_g \cdot \mathbb{I}_k^g + \gamma X_k + \varepsilon_k \tag{2.3}$$

Table 2: Value of Patents as a Function of Backward Citations

	Scientific Value			Economic Value		
	(1)	(2)	(3)	(4)	(5)	(6)
Specific Patent, β_s	0.27	0.22	0.18	0.22	0.20	0.03
	(0.03)	(0.03)	(0.02)	(0.05)	(0.04)	(0.014)
Citation to a Top Patent, β_g	0.33	0.31	0.30	0.15	0.10	0.01
	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)	(0.005)
Firm Controls	Yes	Yes		Yes	Yes	Yes
Industry-Year FE	Yes	Yes			Yes	Yes
Firm FE		Yes				Yes
Firm × Year × Tech Class FE			Yes			

Notes: This table shows the estimation results from the regression (2.3) of the value of a patent on a dummy whether this patent is specific or not. Columns (1)–(3) consider the specification with the scientific value of a patent, which is defined as the number of forward citations relative to the average number of citations for patents granted in the same year and technological class. Columns (4)–(6) consider the specification with the economic value of a patent, which is defined based on the abnormal stock market return on the grant day of a patent for the firm that owns it (Kogan et al. (2017)). Firm controls include lag log of profits, log of employment and log of capital stock. Standard errors are clustered by firm and year.

where Q_k is the value of patent k (scientific or economic), $\mathbb{I}^s_k = 1$ if patent k is specific and zero otherwise, $\mathbb{I}^g_k = 1$ whether patent k makes a citation to a patent from the top 1% of the most cited patents. The control $\mathbb{I}^g_k$ allows to distinguish whether specific patents are of higher quality because of the concentration of citations or because they simply rely on a highly-cited patent. The variable X_k includes the set of controls for the patent and the firm that owns it. The variable ε_k is an error term.

The results are given in the Table 2. All specifications show that specific patents are of higher quality relative to non-specific ones. For the scientific value of patents, the most restrictive specification is given in the column (3) and includes firm-year-technological class fixed effects, meaning that specific patents are of higher scientific quality relative to non-specific ones even within the same firm, grant year, and technological class. The specific patents have on average 18% more forward citations relative to non-specific ones that cite a top patent, and 48% more citations relative to patents that do not cite the top 1% of the most cited patents.

For the economic value of patents, such a restrictive specification is not feasible due to data limitations.[16] The most restrictive specification is given in column (6) and includes firm controls,

[16]The main problem is that patents are granted only on Tuesdays of each week (except federal holidays), and it is impossible to distinguish the value of patents granted on the same day. In my specifications, I define patents to be specific if all patents granted on a given day were specific. Distinguishing them by technological classes would significantly reduce the sample size.

firm and industry-year fixed effects. The specific patents have on average 3% higher economic value relative to non-specific ones that cite a top patent, and 4% higher economic value relative to patents that do not cite the top 1% of the most cited patents.

2.5 Robustness and Additional Results

Appendix B.5 provides multiple robustness checks. In particular, the results are robust when I consider different thresholds for the most cited patents (top 5% and 10%). The concentration is not driven by superstar firms in patenting. The concentration is robust to the exclusion of firms' self-citations to themselves, so it is driven by citations between firms rather than self-citations. The concentration is also robust when I group citations of patents from the same within-country family (continuations, continuations-in-part, divisionals) as a single citation, so 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. Using citations of lawyers who represented several companies and the methodology similar to the movement of inventors, I show that citations are not lawyer-specific, and the concentration is driven by firms. Appendix B.5 provides more robustness checks, including controls for outliers, different samples, and weighting schemes.

2.6 Discussion and Implications for the Theory

Knowledge spillovers might come from different sources. The U.S. courts argue that patents should provide a source of knowledge to other inventors (Roin (2005)), and the same argument is often used in the growth literature (e.g., Romer (1990), p.84). The literature on urban economics and technology diffusion argues that spillovers come from interactions between inventors (Duranton & Puga (2004), Perla & Tonetti (2014), Buera & Lucas (2018)). The evidence above raises concerns whether patent citations can be used to capture these types of knowledge spillovers. First, the comparison of the actual concentration with the random one indicates that that there are significant differences in citation rates across firms with very similar patents. Second, the evidence on the movement of inventors indicates that citation rates are different across firms even within the same inventor.

I argue that patent citations might reflect cooperation between firms and exchange of tacit trade secrets that are complementary to codified knowledge disclosed in patents. As a result, citations are concentrated because only a few companies get access to private knowledge of a patent owner. The evidence in Section 2.4 suggests that this access is valuable to the recipients of trade secrets. In theory, due to the disclosure requirements one invention cannot be protected by both patents and trade secrets, but in practice "[b]ecause innovations are rarely composed of a monolithic piece of knowledge, a combination of patenting and secrecy is common" (Anton et al. (2006)). Appendix A describes several examples on how firms combine secrecy and patenting. In particular, I discuss the examples of the mRNA technology for Covid-19 vaccines and the Haber-Bosch process for ammonia. I also provide several legal cases on why and when the combination of patenting and secrecy does not violate the disclosure requirements of patents.

The evidence from this section motivates the structure of the theoretical framework I develop below. First, the theory is aimed to explain the differences in citation rates across firms rather than the differences in the number of patents. Second, the explanation for the differences in citation rates is based on connections between firms and their incentives to share private knowledge rather than on connections between inventors who work in these firms, and how they interact with each other.

3 Theory

The grant of monopoly through patents is justified by the claim that it promotes knowledge diffusion through the disclosure of inventions. However, the evidence in Section 2 indicates that even for the most cited patents only a limited set of firms builds on the patents' knowledge. I argue that citations are associated with the sharing of trade secrets that accompany patents and develop a theory of knowledge sharing between firms that gives conditions under which knowledge flows are diffused or concentrated.

The idea behind the theory can be illustrated with a case study on the diffusion of knowledge about the copper interconnect technology for semiconductor chips (Lim (2009)). IBM was one of the leaders in the development of this technology and relied heavily on secrecy to protect it. IBM shared its secrets with an input supplier Novellus who developed a copper deposition tool. To protect its secrets, IBM and Novellus signed a contract that restricted Novellus's ability to discuss its partnership with IBM and to sell a deposition tool to others:

> "Under this agreement, Novellus was not permitted to reveal that it was working with IBM. . ., and IBM maintained a list of companies to which Novellus could not initially sell the tool."

In the model, IBM acts as an incumbent firm ($\mathcal{I}$) who shares its private knowledge with an input supplier ($\mathcal{S}$), Novellus. To protect its secrets from leakages to others, $\mathcal{I}$ and $\mathcal{S}$ sign a confidentiality agreement and exclusive dealing. As a result, knowledge flows are concentrated. Around 1997, IBM became more open about the development of the copper interconnect technology and filed for patents. Table C1 (Appendix C) shows the distribution of citations for one of the most cited patents in this technology assigned to IBM and Novellus.[17] Citations for this patent are highly concentrated and mostly come from IBM's suppliers: 64% from Novellus and 30% from other suppliers.

Besides IBM, Motorola also worked on this technology. In contrast to IBM, Motorola was relatively open with its suppliers, and they diffused knowledge to others:

> "... Motorola worked jointly with suppliers from an early stage to develop tools for copper technology. A member of Motorola's copper team remarked that 'unlike IBM, we are quite open with vendors,...' Motorola's relative openness made it an important source of information for the rest of the industry..."

In the model, Motorola also acts as an incumbent firm who shares its private knowledge with an input supplier, for example, Applied Materials. However, the difference from IBM is that Motorola signs neither a confidentiality agreement nor exclusive dealing, so the supplier shares Motorola's knowledge with other firms, for instance, Taiwan Semiconductor Manufacturing Company (TSMC). TSMC entered into this technology much later relative to IBM and Motorola, and in the model it acts as an entrant ($\mathcal{E}$). According to interviews, "[TSMC] depended primarily on technical knowledge from equipment suppliers. TSMC worked closely with Applied Materials..." Table C2 shows the distribution of citations for one of the most cited patents in this technology assigned to Motorola. In contrast to IBM's patent, citations to Motorola's patent are less concentrated and come from a diverse set of firms: 23% come from several suppliers (e.g, 7% from Applied Materials), and 39% come from 17 firms sharing a common supplier with Motorola (e.g., 5% from TSMC).

The goal of the theory is to derive conditions on the market structure and bargaining between firms under which business partners sign or do not sign different types of contracts that prevent knowledge flows in the economy. Section 3.1 presents a simple version of the model. Section 3.2 provides extensions and additional results to the baseline model. Section 3.3 lists the model's testable predictions.

3.1 Simple Model

Demand. There are two competing firms, an incumbent ($\mathcal{I}$) and an entrant ($\mathcal{E}$), who sell their products to a continuum of consumers of measure 1. Each consumer is one of three types. A share $1 - \alpha$ of consumers buys a product from one firm only (loyal customers). Specifically,

[17]The fact that IBM and Novellus filed for a joint patent provides another dimension of a close relationship between these firms. Only around 3% of patents granted in USPTO have more than one assignee. In Figure C5, I document the rise in joint innovations (measured by joint patents) between input suppliers and customers. Both the share of suppliers and customers who applied for joint patents and the share of suppliers and customers among the patents with multiple assignees have increased from 1980 to 2010.

a share $\frac{1-\alpha}{2}$ buys from $\mathcal{I}$ only, and a share $\frac{1-\alpha}{2}$ buys from $\mathcal{E}$ only. Their willingness to pay is determined by the quality of a product. So, $\mathcal{I}$-consumers are ready to pay up to $q_{\mathcal{I}}$, and $\mathcal{E}$-consumers are ready to pay up to $q_{\mathcal{E}}$, where $q_{\mathcal{I}}$ and $q_{\mathcal{E}}$ are product qualities of $\mathcal{I}$ and $\mathcal{E}$, respectively. A share α of consumers buys the product from a firm with the highest quality-price difference (switchers). Therefore, the parameter α measures the degree of competition between $\mathcal{I}$ and $\mathcal{E}$.

Note that this demand can be interpreted through the lenses of international trade where there is a continuum of measure 1 products, firms $\mathcal{I}$ and $\mathcal{E}$ are from different countries, and products with loyal customers represent non-traded goods. An increase in α can be interpreted as an increase in the set of goods that can be traded (e.g., due to the decline in trade costs).

Pricing to Consumers. Firms $\mathcal{I}$ and $\mathcal{E}$ can perfectly discriminate between different types of consumers.[18] So, they can charge loyal customers by their willingness to pay, $p_{\mathcal{I}} = q_{\mathcal{I}}$ and $p_{\mathcal{E}} = q_{\mathcal{E}}$. For the rest (switchers), $\mathcal{I}$ and $\mathcal{E}$ compete in prices. Therefore, these consumers will be served by a firm with a higher product quality at a price $p = \max\{q_{\mathcal{I}}, q_{\mathcal{E}}\} - \min\{q_{\mathcal{I}}, q_{\mathcal{E}}\}$ where I assume that in the case of indifference consumers buy from a firm with the highest quality. I also assume that consumers buy from $\mathcal{I}$ if $p_{\mathcal{I}} = p_{\mathcal{E}}$ and $q_{\mathcal{I}} = q_{\mathcal{E}}$.

Technology. Firm $\mathcal{I}$ has a valuable technology that consists of two types of knowledge: tacit and codified. Tacit knowledge is unpatentable and is easy to hide, so $\mathcal{I}$ keeps it secret. The incumbent needs an input in its production, and it has its own low-quality input or it can buy a high-quality one from an input supplier ($\mathcal{S}$). However, for $\mathcal{S}$ to produce a superior input, firm $\mathcal{I}$ has to share the *tacit* knowledge with it. Without this knowledge, $\mathcal{S}$ can only offer the same low-quality input as $\mathcal{I}$ has. Firm $\mathcal{I}$ can produce a final product of quality $q_{\mathcal{I}} = \lambda$ with a high-quality input from $\mathcal{S}$, and of quality $q_{\mathcal{I}} = (1-\gamma)\lambda$ with a low-quality input where $\gamma \leqslant 1$.

Firm $\mathcal{E}$ also needs an input, and it has its own low-quality input or it can buy an input from $\mathcal{S}$. The quality of $\mathcal{E}$'s product is $q_{\mathcal{E}} = \lambda_0$ with a superior input from $\mathcal{S}$, and $q_{\mathcal{E}} = (1-\gamma)\lambda_0$ with an inferior input where $\lambda_0 < \lambda$. However, using $\mathcal{I}$'s *tacit* knowledge it can innovate and improve the quality of its product to $q_{\mathcal{E}} = \lambda$ with a superior input from $\mathcal{S}$, and $q_{\mathcal{E}} = (1-\gamma)\lambda$ with an inferior input. The summary is given in Table 3.

Firm $\mathcal{E}$ also values $\mathcal{I}$'s codified knowledge, which it can obtain through reverse-engineering of $\mathcal{I}$'s product. However, the codified knowledge without the tacit secrets allows $\mathcal{E}$ only to imitate $\mathcal{I}$'s quality for common customers (switchers) but not to improve the quality of its product for its loyal customers. Codified knowledge is patentable, so $\mathcal{I}$ protects it from imitation with a patent. From the firms' point of view, this imitation is a pure business stealing effect since it does not allow firms to capture more surplus from consumers. Therefore, the incumbent has no

[18]With the interpretation based on international trade, this assumption means that firms can charge monopoly prices for non-traded products in their countries.

Table 3: Quality of Final Products

Access to a high-quality input?	$\mathcal{I}$'s quality	$\mathcal{E}$'s quality Access to $\mathcal{I}$'s tacit knowledge? Yes	 No
Yes	$q_{\mathcal{I}} = \lambda$	$q_{\mathcal{E}} = \lambda$	$q_{\mathcal{E}} = \lambda_0$
No	$q_{\mathcal{I}} = (1-\gamma)\lambda$	$q_{\mathcal{E}} = (1-\gamma)\lambda$	$q_{\mathcal{E}} = (1-\gamma)\lambda_0$

Notes: This table shows the quality of $\mathcal{I}$'s and $\mathcal{E}$'s products as a function of whether they have access to a high-quality input from $\mathcal{S}$ and to $\mathcal{I}$'s tacit secrets. For $\mathcal{S}$ to produce an input of high quality it needs to know $\mathcal{I}$'s tacit secrets.

incentives to license it to $\mathcal{E}$. The idea that firms use patents to disclose knowledge that would be available to the public through other channels anyway is consistent with the arguments of legal scholars that the patent system fails its disclosure function (e.g., Roin (2005)).

To sum up, $\mathcal{I}$ patents its codified knowledge but keeps the tacit knowledge secret. The codified knowledge allows imitation only, and the tacit secrets are key for improving the quality of $\mathcal{S}$'s input and $\mathcal{E}$'s product.

Citations. Both $\mathcal{S}$ and $\mathcal{E}$ apply for patents on their products that do not infringe $\mathcal{I}$'s patent. Without $\mathcal{I}$'s private knowledge they cite $\mathcal{I}$'s patent with probabilities $p_{\mathcal{S}}(0) \geqslant 0$ and $p_{\mathcal{E}}(0) \geqslant 0$, respectively. With $\mathcal{I}$'s private knowledge they cite $\mathcal{I}$'s patent with probabilities $p_{\mathcal{S}}(1) > p_{\mathcal{S}}(0)$ and $p_{\mathcal{E}}(1) > p_{\mathcal{E}}(0)$, respectively.

Contracting. The main contracting limitation is that courts cannot verify the quality of tacit knowledge firms share with each other and quality of their products, so $\mathcal{I}$ cannot license its tacit knowledge to $\mathcal{E}$ when it is mutually beneficial for them. Specifically, once $\mathcal{E}$ pays for knowledge, $\mathcal{I}$ has no incentives to disclose it and courts cannot check whether $\mathcal{I}$ shared relevant secrets with $\mathcal{E}$. Conversely, once $\mathcal{I}$ shares its secrets, $\mathcal{E}$ might refuse to pay claiming that the secrets are not valuable. This is a classical problem in the markets for ideas (Arrow (1962)).[19]

I assume that firms $\mathcal{I}$ and $\mathcal{E}$ can write contracts on inputs with $\mathcal{S}$. Moreover, $\mathcal{I}$ and $\mathcal{S}$ can also sign a confidentiality agreement and exclusive dealing. Although courts cannot verify the quality of secrets firms share with each other, they can verify the communication between parties. A

[19] Arora et al. (2001) discuss how to license tacit knowledge to a non-competing firm by bundling it with a complementary patent. In my model, once $\mathcal{E}$ receives $\mathcal{I}$'s secrets, it can improve the quality of its product without infringing $\mathcal{I}$'s patent, so the strategy for $\mathcal{I}$ to sell its secrets by bundling them with its patent does not work. I abstract from licensing agreements for three reasons. First, since a citing patent is granted, it is unlikely to infringe the cited ones. There are exceptions to this rule but they are rare (see Scotchmer (2004)). My personal conversations with patent attorneys also suggest that patent citations are rarely associated with licensing payments. Second, the growth literature that measures spillovers using patent citations (e.g., Caballero & Jaffe (1993), Akcigit & Kerr (2018)) usually assumes that citing patents do not infringe cited ones. Finally, Arqué-Castells & Spulber (2022) show that the inter-firm network of licensing agreements has very little overlap with the network of supplier-customer relationships.

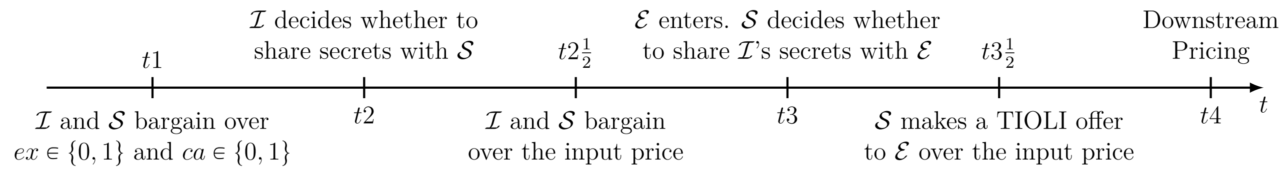


Figure 4: Timing

confidentiality agreement restricts $\mathcal{S}$'s ability to discuss topics related to $\mathcal{I}$'s technology with $\mathcal{E}$. Exclusive dealing restricts $\mathcal{S}$'s ability to sell an input to $\mathcal{E}$.

Timing. The game has the following timing (see also Figure 4):

t1. $\mathcal{I}$ and $\mathcal{S}$ have Nash bargaining with 50/50 split of the surplus about the following contract:

- *Exclusivity* (ex). The contract states whether $\mathcal{S}$ can sell an input to $\mathcal{E}$ ($ex = 0$) or not ($ex = 1$).
- *Confidentiality Agreement* (ca). The contract defines whether $\mathcal{S}$ can share $\mathcal{I}$'s knowledge with $\mathcal{E}$ ($ca = 0$) or not ($ca = 1$).

t2. Firm $\mathcal{I}$ decides whether to share its knowledge with $\mathcal{S}$. Then they have Nash bargaining over the input price ($t_{\mathcal{I}}$) with 50/50 split of the surplus.

t3. $\mathcal{E}$ enters. Suppose $\mathcal{I}$ shared its knowledge with $\mathcal{S}$. Without a confidentiality agreement between $\mathcal{I}$ and $\mathcal{S}$ ($ca = 0$), firm $\mathcal{S}$ decides whether to share $\mathcal{I}$'s secrets with $\mathcal{E}$. Then, under non-exclusive dealing ($ex = 0$) it makes a take-it-or-leave-it offer about the input price ($t_{\mathcal{E}}$) to $\mathcal{E}$.

t4. Firms $\mathcal{I}$ and $\mathcal{E}$ set downstream prices, and profits are realized.

Discussion. An important assumption is that the entrant does not participate in the contracting between $\mathcal{I}$ and $\mathcal{S}$ in the first stage. For example, its identity is not known at this point. This will cause an inefficiency from the firms' point of view because $\mathcal{I}$ and $\mathcal{S}$ will maximize their joint surplus rather than the profits of all three firms. For more on bilateral contracting see Segal (1999).

Section 3.2 discusses various extensions of the simple model. First, I show that specific assumptions about production, demand, and bargaining in the simple model are not crucial for the results. Second, I consider a richer contracting space by allowing the possibility to write contracts with stipulated damages (Aghion & Bolton (1987)). Third, I derive the equilibrium under the restrictions on the use of confidentiality and/or exclusive contracts. Finally, I provide a framework with two suppliers and a common customer.

Subgame Perfect Equilibrium. If $\mathcal{S}$ and $\mathcal{I}$ do not sign an exclusive contract in the first stage, then in the third stage $\mathcal{S}$ will make a take-it-or-leave-it offer to $\mathcal{E}$ about the input price. The offer depends on whether firm $\mathcal{E}$ has access to $\mathcal{I}$'s tacit knowledge or not:

$$\underbrace{t_{\mathcal{E}} = \frac{1-\alpha}{2} \cdot \lambda - \frac{1-\alpha}{2}(1-\gamma) \cdot \lambda = \frac{1-\alpha}{2}\gamma\lambda}_{\mathcal{E} \text{ has access to } \mathcal{I}\text{'s tacit knowledge (secrets)}} \text{ or } \underbrace{t_{\mathcal{E}} = \frac{1-\alpha}{2} \cdot \lambda_0 - \frac{1-\alpha}{2}(1-\gamma) \cdot \lambda_0 = \frac{1-\alpha}{2}\gamma\lambda_0}_{\mathcal{E} \text{ does not have access to } \mathcal{I}\text{'s tacit knowledge (secrets)}}$$

$\mathcal{S}$ always has incentives to share $\mathcal{I}$'s knowledge with $\mathcal{E}$ in the third stage because $\lambda > \lambda_0$. These incentives come from the complementarity between $\mathcal{S}$'s input and $\mathcal{I}$'s technology in the quality of $\mathcal{E}$'s product. A confidentiality agreement between $\mathcal{I}$ and $\mathcal{S}$ can prevent the knowledge leakage to $\mathcal{E}$ because it contractually restricts $\mathcal{S}$'s ability to communicate with $\mathcal{E}$. Exclusive dealing eliminates $\mathcal{S}$'s incentives to leak $\mathcal{I}$'s knowledge by restricting its ability to sell an input to $\mathcal{E}$.

In stage $t2$, $\mathcal{I}$ decides whether to share secrets with $\mathcal{S}$, and then they bargain over the input price. The upside of sharing secrets with $\mathcal{S}$ is that it improves the quality of $\mathcal{S}$'s input. The downsides are that $\mathcal{E}$ might also buy a high-quality input from $\mathcal{S}$ (under non-exclusive dealing, $ex = 0$), and that $\mathcal{S}$ might leak $\mathcal{I}$'s secrets to $\mathcal{E}$ (under $ex = 0$ and $ca = 0$). For a contractual arrangement $(ex, ca) \in \{0,1\}^2$, $\mathcal{I}$ shares its tacit knowledge with $\mathcal{S}$ if

$$\underbrace{\pi^o_{\mathcal{I}}(ex,ca) + \frac{1}{2}\Big(TS(ex,ca) - \pi^o_{\mathcal{I}}(ex,ca) - \pi^o_{\mathcal{S}}(ex,ca)\Big)}_{\mathcal{I}\text{'s profit under knowledge sharing with } \mathcal{S} \text{ and contract } (ex,ca) \in \{0,1\}^2} \geqslant \underbrace{\frac{1-\alpha}{2}(1-\gamma)\lambda + \alpha \cdot (1-\gamma)(\lambda - \lambda_0)}_{\mathcal{I}\text{'s profit without knowledge sharing with } \mathcal{S}} \tag{3.1}$$

where the right-hand side shows $\mathcal{I}$'s profit when it does not share knowledge with $\mathcal{S}$. In this case, it works with an inferior input but so as $\mathcal{E}$, and $\mathcal{E}$ does not get $\mathcal{I}$'s secrets. The left-hand side shows $\mathcal{I}$'s profit when it shares secrets with $\mathcal{S}$. $\mathcal{I}$ and $\mathcal{S}$ bargain over the input price and split the total surplus $(TS(ex, ca))$ 50/50 over the disagreement point $(\pi^o_{\mathcal{I}}(ex, ca), \pi^o_{\mathcal{S}}(ex, ca))$.

In equilibrium, there are three possible contractual arrangements: no contract between $\mathcal{I}$ and $\mathcal{S}$ in the first stage $(ex = 0, ca = 0)$, a confidentiality agreement only $(ex = 0, ca = 1)$, and exclusive dealing $(ex = 1, ca \in \{0, 1\})$. Under exclusive dealing, $\mathcal{S}$ is indifferent whether to share $\mathcal{I}$'s secrets with $\mathcal{E}$ in stage $t3$ or not, and I assume that it does not share them. Therefore, in this case a confidentiality agreement is redundant, and both arrangements $(ex = 1, ca = 0)$ and $(ex = 1, ca = 1)$ lead to the same outcome.[20]

In stage $t1$, $\mathcal{I}$ and $\mathcal{S}$ sign a contract (ex, ca) that maximizes their total surplus $TS(ex, ca)$ subject to constraint (3.1) that $\mathcal{I}$ has incentives to share its secrets with $\mathcal{S}$. Proposition 1

[20] If $\mathcal{I}$ expected a leakage of its knowledge when $\mathcal{S}$ is indifferent, they would sign a confidentiality agreement, but this would not change the structure of the equilibrium. Another way to break the indifference is to assume that there are small costs of knowledge sharing.

summarizes the Subgame Perfect Equilibrium.

Proposition 1. *The subgame perfect equilibrium contract between $\mathcal{I}$ and $\mathcal{S}$ is the following:*

$$\begin{cases} \text{Exclusive Dealing } (ex = 1, ca \in \{0,1\}) & \text{if } \alpha > \frac{1}{3} \\ \text{CA only } (ex = 0, ca = 1) & \text{if } \min\left\{\frac{\gamma}{\gamma+2}, \frac{\gamma}{4(1-\frac{\lambda_0}{\lambda})+\gamma(4\frac{\lambda_0}{\lambda}-1)}\right\} < \alpha \leqslant \frac{1}{3} \\ \text{Null } (ex = 0, ca = 0) & \text{if } \alpha \leqslant \min\left\{\frac{\gamma}{\gamma+2}, \frac{\gamma}{4(1-\frac{\lambda_0}{\lambda})+\gamma(4\frac{\lambda_0}{\lambda}-1)}\right\} \end{cases} \tag{3.2}$$

Firm $\mathcal{I}$ always shares knowledge with $\mathcal{S}$, and $\mathcal{S}$ shares $\mathcal{I}$'s knowledge with $\mathcal{E}$ only under a non-exclusive dealing without a confidentiality agreement ($ex = 0, ca = 0$).

Proof. For each contract, the total surplus of $\mathcal{I}$ and $\mathcal{S}$ is the following

$$\text{Exclusive: } TS(ex = 1, ca \in \{0,1\}) = \frac{1-\alpha}{2} \cdot \lambda + \alpha \cdot (\lambda - (1-\gamma)\lambda_0) + 0 \tag{3.3}$$

$$\text{CA only: } TS(ex = 0, ca = 1) = \frac{1-\alpha}{2} \cdot \lambda + \alpha \cdot (\lambda - \lambda_0) + \frac{1-\alpha}{2} \cdot \gamma\lambda_0 \tag{3.4}$$

$$\text{Null: } TS(ex = 0, ca = 0) = \frac{1-\alpha}{2} \cdot \lambda + \alpha \cdot (\lambda - \lambda) + \frac{1-\alpha}{2} \cdot \gamma\lambda \tag{3.5}$$

where the first two terms represent $\mathcal{I}$'s gross profits, and the last term represent $\mathcal{S}$'s profits from selling an input to $\mathcal{E}$. If $\alpha > \frac{1}{3}$, then $(ex = 1, ca \in \{0,1\})$ maximizes total surplus $TS(ex, ca)$. If $\frac{1}{3} < \alpha \leqslant \frac{\gamma}{\gamma+2}$, then $(ex = 0, ca = 1)$ maximizes total surplus. In both cases, condition (3.1) is satisfied.[21] If $\alpha \leqslant \frac{\gamma}{\gamma+2}$, then $(ex = 0, ca = 0)$ maximizes total surplus, and condition (3.1) is satisfied only if

$$\underbrace{\frac{1-\alpha}{2}(1-\gamma)\lambda + \frac{1}{2} \cdot \frac{1-3\alpha}{2}\gamma\lambda}_{\mathcal{I}\text{'s profit under knowledge sharing}} \geqslant \underbrace{\frac{1-\alpha}{2}(1-\gamma)\lambda + \alpha \cdot (1-\gamma)(\lambda - \lambda_0)}_{\mathcal{I}\text{'s profit without knowledge sharing}}$$

which can be rewritten as

$$\alpha \leqslant \frac{\gamma}{4(1-\frac{\lambda_0}{\lambda}) + \gamma(4\frac{\lambda_0}{\lambda} - 1)} \tag{3.6}$$

□

The equilibrium is illustrated in Figure 5, and the intuition behind Proposition 1 is the following. For product qualitites $(q_{\mathcal{I}}, q_{\mathcal{E}})$, the total sum of profits of all three firms is

$$\Pi_{all} = \frac{1-\alpha}{2}(q_{\mathcal{I}} + q_{\mathcal{E}}) + \alpha \cdot (q_{\mathcal{I}} - q_{\mathcal{E}}) = \frac{1+\alpha}{2}q_{\mathcal{I}} + \frac{1-3\alpha}{2}q_{\mathcal{E}}$$

[21] Condition (3.1) is always satisfied for $(ex = 1, ca \in \{0,1\})$ and is satisfied for $(ex = 0, ca = 1)$ due to simplifying functional form assumptions (see Appendix D.1). I consider a more general case in Appendix D.2.

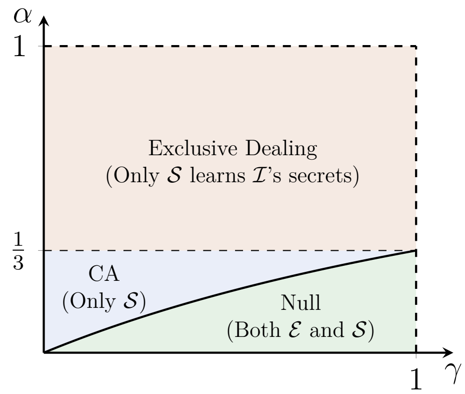


Figure 5: Equilibrium as a function of (α, γ)

This figure illustrates the equilibrium contract between $\mathcal{I}$ and $\mathcal{S}$ described in Proposition 1. Parameter α measures the degree of competition between $\mathcal{I}$ and $\mathcal{E}$, and parameter γ reflects the value of $\mathcal{S}$'s input relative to $\mathcal{I}$'s and $\mathcal{E}$'s outside options.

For $\alpha > \frac{1}{3}$, it is decreasing in $q_{\mathcal{E}}$, so $\mathcal{I}$ and $\mathcal{S}$ sign exclusive dealing to minimize $q_{\mathcal{E}}$. This mechanism is similar to the foreclosure literature (Rey & Tirole (2007)). For $\alpha < \frac{1}{3}$, the sum of profits is increasing in both $q_{\mathcal{I}}$ and $q_{\mathcal{E}}$, and is maximized under a transfer of $\mathcal{I}$'s tacit knowledge (secrets) to $\mathcal{E}$, and when both firms use a high-quality input from $\mathcal{S}$. Since $\mathcal{S}$ makes a take-it-or-leave-it offer to $\mathcal{E}$ about the input price, it can extract the whole increase in surplus coming from its input. Therefore, $\mathcal{I}$ and $\mathcal{S}$ prefer not to sign an exclusive contract: an input price to $\mathcal{E}$ is greater than losses of $\mathcal{I}$ in profits.[22]

However, a knowledge transfer to $\mathcal{E}$ might fail for two reasons. First, since sharing of secrets in non-contractible, $\mathcal{I}$ cannot be directly compensated for it. So, if (3.6) is not satisfied (competition is too high), costs of a knowledge leakage will outweigh benefits of higher input quality, and $\mathcal{I}$ will not share knowledge with $\mathcal{S}$. In this case, $\mathcal{I}$ and $\mathcal{S}$ sign a confidentiality agreement. Second, if (3.6) is satisfied, $\mathcal{I}$ and $\mathcal{S}$ might still prefer a confidentiality agreement because $\mathcal{S}$ does not capture the whole increase in surplus associated with a knowledge transfer to $\mathcal{E}$. Specifically, $\mathcal{I}$'s secrets increase $\mathcal{E}$'s profit by $\frac{1-\alpha}{2}(\lambda - \lambda_0)$, but $\mathcal{S}$ can capture only $\gamma \cdot \frac{1-\alpha}{2}(\lambda - \lambda_0)$ through its input price. Therefore, if γ is low, then sharing of $\mathcal{I}$'s secrets with $\mathcal{E}$ generates more losses in $\mathcal{I}$'s profits relative to gains in input price $t_{\mathcal{E}}$. So, $\mathcal{I}$ and $\mathcal{S}$ prefer a confidentiality agreement to create a commitment for $\mathcal{S}$ not to share $\mathcal{I}$'s secrets with $\mathcal{E}$.

[22] If $\mathcal{E}$ made an offer to $\mathcal{S}$, then $\mathcal{I}$ and $\mathcal{S}$ would always prefer exclusive dealing. In general, suppose that $\mathcal{S}$ and $\mathcal{E}$ have a generalized Nash Bargaining where $\psi_{\mathcal{S}} \in [0, 1]$ is $\mathcal{S}$'s bargaining power. Then, $\mathcal{I}$ and $\mathcal{S}$ sign exclusive dealing with a confidentiality agreement whenever $\alpha > \frac{\psi_{\mathcal{S}}}{\psi_{\mathcal{S}}+2}$. Appendix D.2 provides more details.

3.2 Extensions and Additional Results

This section discusses extensions and additional results to the baseline model. Section D.2 shows that the results from the simple model are robust to generalizations in the competition, production, and bargaining structures.

Section D.3 expands the contractual space by allowing contracts with stipulated damages (Aghion & Bolton (1987)). The ability to use stipulated damages might allow $\mathcal{I}$ and $\mathcal{E}$ to extract more surplus from $\mathcal{E}$ and increase social welfare because it ensures a knowledge transfer from $\mathcal{I}$ to $\mathcal{E}$ in cases where previously $\mathcal{I}$ with $\mathcal{S}$ signed a confidentiality agreement (see Figure 5). This happens because $\mathcal{I}$ is compensated for the knowledge transfer to $\mathcal{E}$ through stipulated damages while before contractual limitations prevented this compensation.

Section D.4 discusses welfare implications of restrictions on the contractual space. The upside of this policy is that it might promote knowledge diffusion in the economy. The downside of restrictions on the use of confidentiality agreements is that firms might start using exclusive dealing to prevent knowledge leakages. The downside of restrictions on both exclusive contracts and confidentiality agreements is that firms in a vertical relationship might stop sharing knowledge with each other. I provide conditions on the market structure under which restrictions on confidentiality or exclusive contracts generate welfare gains. It is never optimal to ban these contracts simultaneously.

Finally, Section D.5 considers a setup with two suppliers and a common customer. Most predictions are similar to the model with a common supplier, and I discuss them in the next section.

3.3 Testable Predictions

The theory predicts patterns of knowledge flows in a production network that I map to the data using patent citations. In Section 4, I test four predictions from the model:

1. Suppliers and customers are more likely to cite each other relative to similar firm-pairs with no vertical relationship.

2. Firms sharing a common supplier or customer are more likely to cite each other relative to firms with exclusive suppliers or customers.

3. A stronger bargaining position (γ) of the common supplier (a worse outside option of $\mathcal{E}$) is associated with a higher likelihood of a citation between its customers ($\mathcal{I}$ and $\mathcal{E}$).

4. A higher level of competition (α) leads to less knowledge sharing between $\mathcal{I}$ and $\mathcal{E}$.

Remark 1. The model with two customers and a common supplier predicts that $\mathcal{E}$ is more likely to cite $\mathcal{I}$ if it has a bad outside option to $\mathcal{S}$'s input (high γ). In the framework with two suppliers and a common customer in Appendix D.5, I assume that suppliers' outside options of not trading with a customer are zero. Therefore, it does not give a prediction about citations between suppliers as a function of the bargaining position of a common customer. I leave the framework with a more general production network for future research.

Remark 2. In the model, each firm has one patent, and sharing of secrets affects citation probabilities between them. The framework can be extended to the case in which firms have multiple patents. Suppose that firm $\mathcal{S}$ has $N_{\mathcal{S}}$ patents, and each independently makes a citation to $\mathcal{I}$'s patent with probability $p_{\mathcal{S}}(1)$ if $\mathcal{S}$ receives $\mathcal{I}$'s secrets and $p_{\mathcal{S}}(0)$ otherwise. The expected number of citations to $\mathcal{I}$'s patent from firm $\mathcal{S}$ is $p_{\mathcal{S}}(k) \cdot N_{\mathcal{S}}$ for $k \in \{0, 1\}$. The empirical strategy in Section 4.1 estimates $p_{\mathcal{S}}(1) - p_{\mathcal{S}}(0)$ taking $N_{\mathcal{S}}$ as given. The focus on citation probabilities is motivated by the evidence from Section 2 that the concentration of citations is primarily explained by differences across firms in citation rates rather than patenting. To explain the full distribution of citations across firms one should explain not only citation probabilities ($p_{\mathcal{S}}(k)$) but also the number of patents ($N_{\mathcal{S}}$). I leave the full analysis of $(p_{\mathcal{S}}(k), N_{\mathcal{S}}(k))$ for future research.

4 Empirical Tests

This section tests four theoretical predictions from Section 3.3. Section 4.1 tests predictions 1–3 using cross-sectional and time variation in supplier-customer relationships between companies. Specifically, the unit of analysis is a pair of patents, and I estimate the probability of a citation between a randomly selected pair of patents as a function of the relationship between their assignees.

Section 4.2 tests the fourth prediction about the effect of competition on knowledge sharing. There are at least two ways to test this prediction. The first one is a test of the equilibrium structure: compare cross-sectional variation across industries in the competition level with citation patterns between firms. There is a potential endogenity problem with this approach. For example, firms in industries with less competition might also rely less on knowledge from each other for technological reasons. The second way is a test of the equilibrium response: find within-industry shocks to the competition levels and compare *changes* in citation patterns in the model and data. I use the rise of imports from China and the instrumental variable approach from Autor et al. (2013) to isolate plausibly exogenous industry shocks to the competition levels. Specifically, the unit of analysis is a technological class, and I show that technologies more exposed to the rising import competition from China experienced a faster growth in the concentration of citations.

Data

I use the data on utility patents granted in the U.S. Patent and Trademark Office (USPTO), Compustat North America data on publicly traded firms, and Compustat Segments data on the supplier-customer relationship between firms. The data are taken for the period 1976–2019. Compustat data give information on sales, profits, employment and industry affiliation of publicly traded firms. Compustat Segments contains data on the supplier-customer links reported by publicly traded U.S. firms. Regulation SFAS No. 131 requires publicly traded firms to report the identity of any customer representing more than 10% of their total sales. Therefore, the data are constructed using the information provided by suppliers about their customers. All details are given in appendix B.1.

These data have several limitations. First, the supplier-customer data cover only publicly traded firms listed in the U.S. Private and foreign firms might differ from the U.S. publicly traded firms in terms of patenting and citation practices, and this might bias the results relative to the estimation based on the full network of firms' relationships. The differences in rates of patenting should not confound the results because the empirical strategy in Section 4.1 is designed to estimate the probability of a citation between two patents from different firms conditional on the existence of these patents. However, the differences in citation practices might introduce a bias if these differences are not accounted by geographic and firm controls. For example, despite filing patents in the U.S. office foreign firms might still follow citation standards of their countries. The second limitation is a selection problem because firms are required to list only the largest customers. As a result, the data might miss small customers and large suppliers selling to multiple firms who do not represent more than 10% of the supplier's revenue. I use controls for firms' sizes to partially alleviate this problem. Third, suppliers are not required to report sales to their customers.[23] Data on bilateral sales would be useful to construct bargaining positions between firms in a vertical relationship. Instead, I use total sales of a supplier relative to other firms in the same industry to proxy for its bargaining position vis-a-vis its customers.

Despite all these limitations, Compustat Segments data set is one of the best sources for supplier-customer relationships in the U.S. over a long period of time. These data were used in studies on propagation of shocks in production networks (Barrot & Sauvagnat (2016)), network formation (Lim (2017)), and tests of investors' inattention (Cohen & Frazzini (2008)).

[23] Computstat Segments data include a variable for the annual sales of the reporting supplier to the reporting customer, but this variable is often missing or imputed.

4.1 Citations in a Production Network

In this section, I test predictions 1–3 from Section 3.3. The first prediction is that firms are more likely to cite each other if they are connected by a vertical production relationship. Second, firms are more likely to cite others' patents if they share a common supplier or customer. Finally, the probability of a citation between firms sharing a common supplier is increasing in the bargaining position of this supplier. The strength of the supplier's bargaining position is defined by the the customer's outside options of not buying an input from the supplier. I proxy the bargaining position of common suppliers by their sales share in the industry they are affiliated with. The intuition is that the supplier's relative size approximates the difficulty of finding an alternative supplier for the customers.

Based on the econometric techniques from the spatial literature (Jaffe et al. (1993), Thompson & Fox-Kean (2005), Singh & Marx (2013)), I estimate the following specification

$$\begin{aligned} \mathbb{P}\left(y_{ij}=1|\mathbb{I}_{ij},X_{ij}\right) = F\Big(&\beta_{sc}\cdot\mathbb{I}^{sc}_{ij}+\beta_{cs}\cdot\mathbb{I}^{cs}_{ij}+\beta_{com}(s)\cdot\mathbb{I}^{com}_{ij}(s)\cdot(1+\beta_s\cdot S_{ij})+ \\ &+\beta_{com}(c)\cdot\mathbb{I}^{com}_{ij}(c)+X_{ij}\cdot\gamma\Big) \end{aligned} \tag{4.1}$$

where $y_{ij}=1$ if patent j cites patent i and zero otherwise. For a pair of patents (i,j) denote the application year of patent j by t_j. The variable $\mathbb{I}^{sc}_{ij}$ ($\mathbb{I}^{cs}_{ij}$) a dummy equal to 1 if j's assignee was a supplier (customer) to i's firm in the period $[t_j-2,t_j+2]$. The variable $\mathbb{I}^{com}_{ij}(s)$ ($\mathbb{I}^{com}_{ij}(c)$) is a dummy equal to 1 if j's and i's firms shared at least one common supplier (customer) in the period $[t_j-2,t_j+2]$. Finally, the variable S_{ij} reflects the size of common suppliers. Specifically, for each common supplier I compute the average sales share in its industry over the period $[t_j-2,t_j]$. If firms shared several suppliers, I take an average across their sales shares. The variables in X_{ij} include firm, spatial, time and technological controls.[24] I use logit, probit and linear probability models. In the linear model, I include fixed effects for the interaction between (i,j) technological classes and i's grant years ($Tech_i\times Tech_j\times Year_i$) where the technology class is defined at the group level in Cooperative Patent Classification (CPC) system. This allows to control for differences in citation patterns across technologies and in time. I also include fixed effects for cited and citing firms which allow to control for time-invariant differences in citation practices across firms.

The estimation of the specification (4.1) is not computationally feasible for the universe of all patent-pairs. Since the data on the supplier-customer links cover only publicly traded

[24] Specifically, firm controls include log sales of i's assginee and log number of patents granted to i's assignee in the grant year of patent i, log sales of j's assignee and log number of patents applied by j's assignee in application year t_j. Spatial controls include log distance between inventors and a dummy whether they are in the same country. Time controls include i's grant year fixed effects, and dummies for a lag between j's application year and i's grant year. Technology controls include a dummy whether i and j patents are in the same main subgroup of CPC, and whether their assignees are in the same four-digit Standard Industry Classification (SIC) industry.

firms, I restrict the sample to citations between these firms only (excluding self-citations), but the set of all patent-pairs is still too large. I use the sampling procedure first proposed in Jaffe et al. (1993). Specifically, for each cited-citing pair of patents I randomly select a control patent from the set of patents that do not make a citation to the cited patent but have the same application year, the same detailed technological class (main subgroup CPC) as the citing patent, and belong to a firm different from the citing one.[25] I use the sampling weighting (Manski & Lerman (1977)) to ensure that estimates reflect probabilities in the sample of cited-citing and all cited-control patent pairs, not only selected ones. Appendix B.6 provides details on the empirical specification and data construction.

The estimation results are given in Table 4. All specifications show that the probability of a citation between two patents is higher if patents' assignees have a supplier-customer relationship and/or share a common supplier or customer. Moreover, the probability of a citation between firms sharing common suppliers is increasing in the relative size of these suppliers. Suppliers, on average, are more likely to cite customers than customers suppliers, and sharing a customer is associated with a greater increase in the probability of a citation relative to sharing a supplier. Columns (1)–(4) show the results for the linear probability model. I multiply the depended variable by 100, so the estimates represent an increase in the probability of a citation in percentage points. In general, the probability of a citation between two randomly selected patents is very small, and this translates into small estimates of the linear probability model. However, they are economically large. A randomly selected pair of cited-citing patents has, on average, around 483 cited-control pairs without a citation meaning that the probability of a citation is around 0.21%. So, being a supplier increases the probability of a citation by 0.08%, which represents a $8/21 = 38\%$ proportional increase. The Logit model provides another way to evaluate the magnitudes of the estimates. Since the probability of a citation is low, the Logit coefficients approximate the proportional increase in the probability of a citation. Specifically, the marginal effect of k's variable (dummy) is

$$\Lambda(X\gamma + \beta_k) - \Lambda(X\gamma) \approx \beta_k \cdot \Lambda'(X\gamma) = \beta_k \cdot \Lambda\left(X\gamma\right) \cdot \left(1 - \Lambda\left(X\gamma\right)\right) \approx \beta_k \cdot \Lambda\left(X\gamma\right)$$

where $\Lambda(X\gamma) = \frac{1}{1+\exp(-X\gamma)}$ and $\Lambda\left(X\gamma\right) \approx 0$. So, being supplier to a cited firm is associated with a 34% proportional increase in a probability of a citation between two patents, which is

[25] Jaffe et al. (1993) consider control patents from the same 3-digit USPC technological class. However, Thompson & Fox-Kean (2005) argue that the 3-digit classes are too broad to capture technological similarity between patents and propose to use the 9-digit classification. The problem with the 9-digit classification is that it is too narrow and potentially introduces a lot of noise, leaving the majority of patents without a control one (Henderson et al. (2005)). Unfortunately, the USPC system does not provide intermediate levels of aggregation between 3-digit and 9-digit classes. To address both critiques, I use instead the cooperative patent classification (CPC), which has more flexible hierarchical classification system. Specifically, I find control patents from the same main subgroup level, which has more than 7000 categories.

Table 4: Supply Chain Relationship and the Probability of a Citation

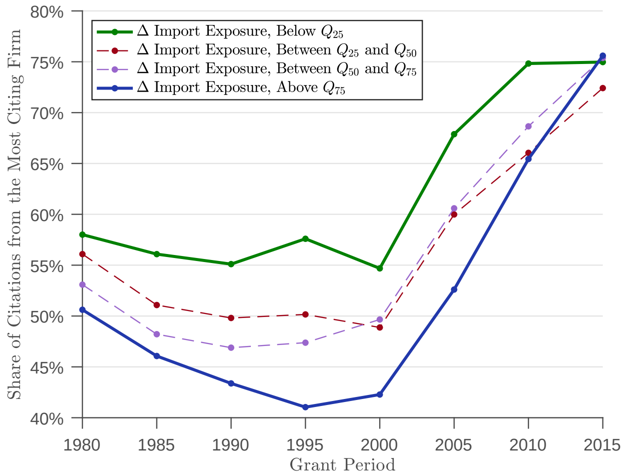
Empirical Specification	Linear				Logit	Probit
	(1)	(2)	(3)	(4)	(5)	(6)
Supplier Cites a Customer, β_{sc}	0.09	0.09	0.08	0.08	0.34	0.11
	(0.004)	(0.004)	(0.003)	(0.004)	(0.012)	(0.004)
Customer Cites a Supplier, β_{cs}	0.05	0.05	0.05	0.07	0.27	0.09
	(0.003)	(0.003)	(0.003)	(0.003)	(0.015)	(0.005)
Share Common Customers, $\beta_{com}(c)$	0.10	0.10	0.09	0.07	0.39	0.12
	(0.003)	(0.003)	(0.002)	(0.003)	(0.009)	(0.003)
Share Common Suppliers, $\beta_{com}(s)$	0.02	0.01	0.02	0.03	0.04	0.01
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.001)
Relative Size of Common Suppliers, β_s	0.54	0.59	0.51	0.47	2.28	0.76
	(0.024)	(0.024)	(0.024)	(0.025)	(0.061)	(0.022)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Patent Controls		Yes	Yes	Yes	Yes	Yes
$\text{Tech}_i \times \text{Tech}_j \times \text{Year}_i$ FE			Yes	Yes		
Firm_i, Firm_j FE				Yes		

Notes: This table shows the estimation results for the specification in (4.1). It estimates the probability of a citation between two patents as a function of the production relationship between their assignees. The sample is restricted to patents granted between 1976 and 2019 to publicly traded firms only and excludes self-citations. I use the sampling procedure outlined in Jaffe et al. (1993) and Thompson & Fox-Kean (2005) to construct the sample of cited-citing and cited-control pairs of patents. For the linear probability model, the outcome variable is multiplied by 100. In the regressions, firm controls include log sales of i's assignee and log number of patents granted to i's assignee in the grant year of patent i, log sales of j's assignee and log number of patents applied by j's assignee in application year t_j, and a dummy whether i and j assignees are in the same four-digit Standard Industry Classification (SIC) industry. Patent controls include log distance between inventors and a dummy whether they are in the same country, i's grant year fixed effects, and dummies for a lag between j's application year and i's grant year, a dummy whether i and j patents are in the same main subgroup of CPC. Standard errors are clustered at the cited patent level. All specifications are sample-weighted (Manski & Lerman (1977)). More details are given in Appendix B.6.

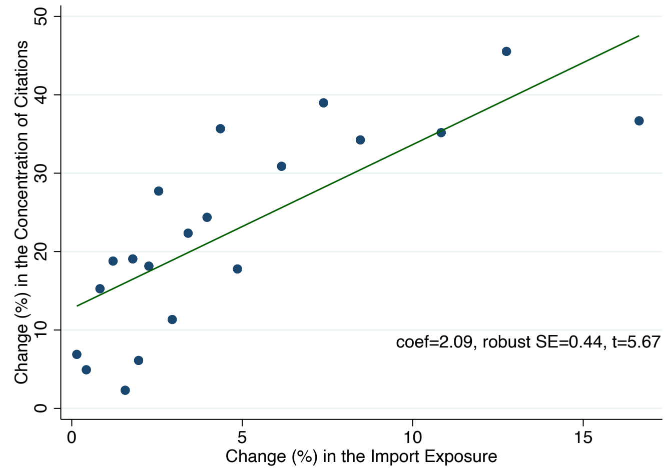
comparable to 38% in the linear probability model.

4.2 Trade with China and Concentration of Knowledge Flows

The model in Section 3 predicts that a higher level of competition leads to lower knowledge sharing. I provide the evidence consistent with this prediction using the variation across technologies in their exposure to the rising import competition from China. To isolate changes in the import competition level unrelated to US demand and technological shocks, I instrument US imports from China by Chinese exports to other high-income countries (Autor et al. (2013)). Autor et al. (2020a) 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 knowledge sharing between firms.



(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 6: 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. Figure 6(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 $[t-4, t]$. Figure 6(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.

Following Autor et al. (2020a), I define the measure of the 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 technology class using the mapping of four-digit SIC industries to technology classes implied from patents owned by publicly traded firms as in Autor et al. (2020a).

Figure 6(a) shows the concentration measure from Section 2 (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 is 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%. So, technological classes with the most exposure to China shock experienced faster growth in the concentration.

I follow the methodology from Section 2 to separate whether the rise in the concentration in different classes is driven by the increasing variance across firms in patenting (N) or in citation rates (p). In Figure C6 (Appendix C), I compare the dynamics after 2000 of the actual concentration measure relative to the counterfactual one from the Monte-Carlo simulations. The counterafactual concentration shows how the concentration would evolve due to the changes in patenting across firms (N) holding the firms' citation rates (p) fixed at the level of 2000. For technologies least exposed to the competition from China, the actual and the counterfactual concentration measures closely follow each other, and the changes in citation rates across firms (p) explain about 41% of the rise in the concentration. In contrast, for the most exposed technologies the changes in citation rates across firms (p) explain around 70% of the increase in the concentration. Therefore, trade with China affected citation patterns primarily through the changes in citation rates. An implication of this result is that firms' exit from patenting cannot fully explain the rise in the concentration of citations.

I study the change in the concentration of citations around 2000 in the regression specifications. Specifically, I define for each technology class the average concentration measure among patents granted in the 7-year period starting from 1977, 1984, 1991, 1998, 2005.[26] Appendix B.7 provides more details. I denote the concentration measure for technology class j and the period starting from t by $\mathcal{C}_{j,t}$, and define the following growth measures

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

Figure 6(b) 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

[26] Periods are 1977–1983, 1984–1990, 1991–1997, 1998–2004, 2005–2011. The sample construction for patents is different from Autor et al. (2020a) due to the nature of the outcome variable. Autor et al. (2020a) 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 7-year 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 5: Trade with China and Increase in the Concentration of Citations

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: OLS					
Δ Tech Class Exposure	2.09	1.46	1.44	1.76	1.43	1.43
to Chinese Imports	(0.44)	(0.41)	(0.42)	(0.34)	(0.39)	(0.38)
	Panel B: 2SLS					
Δ Tech Class Exposure	2.38	1.52	1.50	1.92	1.84	1.81
to Chinese Imports	(0.49)	(0.50)	(0.50)	(0.42)	(0.56)	(0.58)
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. (2020a)). 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. (2020a). 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.

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 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 whether it has more than 50% of software subclasses (Graham & Vishnubhakat (2013)).

Panel A in Table 5 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. In all specifications, there is a positive and significant relationship between the changes in the import competition from China and the growth in the concentration of citations. 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 non-corporate patents (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.

5 Conclusion

This paper shows that patents play a limited role in the diffusion of knowledge, and patent citations provide a poor measure of knowledge spillovers. I argue that patent citations reflect intentional knowledge sharing between business partners. I develop and test a theory of how common suppliers and customers might diffuse knowledge across competitors.

Much of the literature on economic growth (Jones (2005), Aghion et al. (2014)) is based on the idea of knowledge spillovers, and their systematic measurement relies on patent citations (Trajtenberg & Jaffe (2002)). The evidence in this paper raises concerns whether spillovers in the growth models can be disciplined by patent citations. The inability to measure knowledge externalities brings back the critique by Krugman (1991) that "... there is nothing [in the data] to prevent the theorist from assuming anything about them that she likes". Patent citations still provide valuable information about collaboration between firms, but the growth theories are often focused on creative destruction (competition) rather than creative construction (cooperation). More studies of strategic alliances between firms coupled with patent citations can bring new insights about innovations and technological progress.

References

Acemoglu, D., Akcigit, U., & Kerr, W. (2016). Innovation Network. *PNAS*, 113(41), 11483–11488.

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. (2021). The impact of regulation on innovation. *NBER Working Paper No. 28381*.

Aghion, P. & Bolton, P. (1987). Contracts as a Barrier to Entry. *American Economic Review*, 77(4), 388–401.

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

Aitken, B. J. & Harrison, A. E. (1999). Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela. *American Economic Reivew*, 89(3), 605–618.

Akcigit, U. & Ates, S. T. (2019). What Happened to U.S. Business Dynamism? *Working Paper*.

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.

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. (2021). The Effects of Joining Multinational Supply Chains: New Evidence from Firm-to-Firm Linkages. *Working Paper*.

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).

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

Arora, A., Belenzon, S., Patacconi, A., & Suh, J. (2019). The Changing Structure of American Innovation: Some Cautionary Remarks for Economic Growth. *Innovation Policy and the Economy, Volume 20, edited by Josh Lerner, and Scott Stern*, (pp. 39–93).

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

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.

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. *in R. Nelson, eds., The Rate and Direction of Inventive Activity. Princeton University Press, Princeton, NJ*, (pp. 609–625).

Arts, S., Cassiman, B., & Gomez, J. C. (2018). Text Matching to Measure Patent Similarity. *Strategic Management Journal*, 39, 62–84.

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

Autor, D., Dorn, D., Katz, L., Patterson, C., & Reenen, J. V. (2020b). The Fall of the Labor Share and the Rise of Superstar Firms. *Quarterly Journal of Economics*, 135(2), 645–709.

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. (2021). 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.

Blalock, G. & Gertler, P. (2008). Welfare Gains from Foreign Direct Investment Through Technology Transfer to Local Suppliers. *Journal of International Economics*, 74(2), 402–421.

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*.

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, L. & Frazzini, A. (2008). Economic Links and Predictable Returns. *The Journal of Finance*, 63(4), 1977–2011.

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*.

Duranton, G. & Puga, D. (2004). Micro-Foundations of Urban Agglomeration Economies. *Handbook of Regional and Urban Economics*, 4, 2063–2117.

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

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*. John Wiley & Sons, Inc., Third edition.

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.

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.

Henderson, R., Jaffe, A. B., & Trajtenberg, M. (2005). Patent Citations and the Geography of Knowledge Spillovers: A Reassessment: Comment. *American Economic Review*, 95(1), 461–464.

Higham, K. W., de Rassenfosse, G., & Jaffe, A. B. (2020). Patent Quality: Towards a Systematic Framework for Analysis and Measurement. *NBER Working Paper No. 27598*.

Jaffe, A. B. & de Rassenfosse, G. (2019). Patent Citation Data in Social Science Research: Overview and Best Practices. *Research Handbook on the Economics of intellectual Property Law*.

Jaffe, A. B. & Trajtenberg, M. (2002). *Patents, citations, and innovations: A window on the knowledge economy*. MIT Press.

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.

Krugman, P. (1991). *Geography and Trade*. MIT Press.

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

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.

Lim, K. (2009). The Many Faces of Absorptive Capacity: Spillovers of Copper Interconnect Technology for Semiconductor Chips. *Industrial and Corporate Change*, 18(6), 1249–1284.

Lim, K. (2017). Endogenous Production Networks and the Business Cycle. *Working Paper*.

Manski, C. F. & Lerman, S. R. (1977). The Estimation of Choice Probabilities From Choice Based Samples. *Econometrica*, 48(8), 1977–1988.

Marx, M. & Fuegi, A. (2020). Reliance on Science: Worldwide Front-Page Patent Citations to Scientific Articles. *Strategic Management Journal*, 41(9), 1572–1594.

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.

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.

Perla, J. & Tonetti, C. (2014). Equilibrium Imitation and Growth. *Journal of Political Economy*, 122(1), 52–76.

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

Rey, P. & Tirole, J. (2007). A Primer on Foreclosure. *in M. Armstrong and R. Porter, eds., Handbook of Industrial Organization*, 3, 2145–2220.

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.

Segal, I. (1999). Contracting with Externalities. *The Quarterly Journal of Economics*, 114(2), 337–388.

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.

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

Whinston, M. D. (2006). *Lectures on Antitrust Economics.* The MIT Press.

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.[27]

> " '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."[28] 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.*

[27] 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.

[28] 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.[29] 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.[30] 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. (2020a) 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. (2020a) for cleaning and standardizing firm names (e.g., replace "Incorporated" with "INC"). Finally, I matched around 100 thousand patents manually for the largest assignees.

Suppliers and Customers. I take the supplier-customer data from Compustat Segments data set. For publicly traded firms, it lists 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 2019 by manual inspection of Compustat Segments data.

[29] *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).*

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

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

Section B.2.1 provides details on the construction of the concentration measure. Section B.2.2 gives details on the Monte-Carlo exercise that equates citation rates across firms. Section B.2.3 discusses the Monte-Carlo exercise that equates citation rates and the number of patents across firms across firms. Section B.2.4 gives details on how to identify "specific" patents that are used in Section 2.4. Section B.2.5 provides details on the concentration of citations for scientific publications.

B.2.1 Construction of the Concentration Measure

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 technology 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 technology 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 will compute the distribution of citations across different organizations, so I exclude citations from individual inventors and patents with missing assignee information.[31] Each patent $k \in \mathcal{P}$ has a grant year t_k^g and a primary technology class c_k. I define the technology class 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 technology 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

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

by $\Omega_{t,c}^{top}$, and define the *"Main"* sample as top patents for all years and technology classes:

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

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

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

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 technology class (c) I compute a simple average across patents[32]

$$\mathcal{C}(t,c) = \frac{1}{|\Omega_{t,c}^{top}|} \sum_{k \in \Omega_{t,c}^{top}} \mathcal{C}_k \tag{B.2}$$

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 $\Omega_{t,c}^{top} \neq \varnothing$

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

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

B.2.2 Monte-Carlo Simulations to Equalize Citation Rates Across Firms

The details of the Monte-Carlo simulations that equate citation rates across firms are the following. First, for each patent from the *Main* sample I take the citing patents and identify control patents that have the same application year, geographical location of inventors, and detailed technological class. Second, for each patent I randomize citation links across citing and control patents. Then for each patent I compute the concentration measure on the simulated sample. Third, I repeat this procedure 300 times to construct the distribution of

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

concentration measures. Finally, I aggregate different moments of this distribution in a way similar to the actual concentration measure.

In the first step, I divide all granted patents into disjoint groups based on common observational characteristics. Each patent k has an application year t_k^a, a geographical location of the majority of inventors ℓ_k, and the detailed technological class $\tilde{c}_k$. 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. For a detailed technology class, I take the main subgroup level in CPC, which has more than 7000 categories. It is nested within a group level, $\widetilde{\mathcal{CPC}} \subset \mathcal{CPC}$ where $\widetilde{\mathcal{CPC}}$ is the set of all main subgroups in CPC. I define the characteristics as a triple $h = (t^a, \ell, \tilde{c})$, and each patent is assigned to a unique group of characteristics. Define the set of all characteristics as

$$\mathcal{H} = \left\{ h = (t^a, \ell, \tilde{c}) : t^a \in \text{All Observed Application Years}, \ell \in \text{All Observed Locations}, \tilde{c} \in \widetilde{\mathcal{CPC}} \right\}$$

For each $h = (t^a, \ell, \tilde{c}) \in \mathcal{H}$, denote the group of patents with these characteristics by $\mathcal{P}(h)$ such that

$$\mathcal{P} = \bigcup_{h \in \mathcal{H}} \mathcal{P}(h) \text{ and } \mathcal{P}(h) \cap \mathcal{P}(h') = \varnothing \text{ for } h \neq h'$$

For each $k \in Main$, I take the set of citing patents

$$Citing_k = \{m \in \mathcal{P} : y_{km} = 1 \text{ and } m \in \mathcal{T}_k\} \tag{B.4}$$

Denote the number of citations to patent k from patents with characteristics h by $n_k(h)$. For each $h \in \mathcal{H}_k$, I equate citation rates across firms. Formally, for each $h \in \mathcal{H}_k$ I randomize $n_k(h)$ citations across all patents that have characteristics h and satisfy the 5-year window time constraint. Denote the set of these patents by $\mathcal{P}_k(h)$:

$$\mathcal{P}_k(h) = \{m \in \mathcal{P}(h) : m \in \mathcal{T}_k\}$$

where $\mathcal{T}_k$ is defined in (B.1). The Monte-Carlo exercise takes $n_k(h)$ random patents from the set $\mathcal{P}_k(h)$ and assigns citations to these patents. Under this randomization patents in $\mathcal{P}_k(h)$

from different firms make a citation to patent k with the same probability

$$p_k(h) = \frac{n_k(h)}{|\mathcal{P}_k(h)|}$$

I repeat the randomization procedure 300 times, and each time I compute the counterfactual concentration of citations for patent k. Denote the concentration measure for patent k in round s by $\mathcal{RC}_{k,s}$. I compute the median and the 95th quantile based on the distribution $\{\mathcal{RC}_{k,s}\}_{s=1}^{300}$, which are denote by $\mathcal{RC}_k(q95)$ and $\mathcal{RC}_k(q50)$. Then I aggregate these measures across patents in the same way as with the actual concentration, see equations (B.2) and (B.3).

B.2.3 Monte-Carlo Simulations to Equalize Citation Rates and Patenting Across Firms

The Monte-Carlo exercise that equates both citation probabilities and patenting is similar to the one from Section B.2.2 except the details on the randomization of citations. In the exercise that equates citation probabilities only, $n_k(h)$ citations are allocated randomly across patents in $\mathcal{P}_k(h)$. Therefore, firms with more patents in $\mathcal{P}_k(h)$ are more likely to cite patent k. To equate patenting, I assume that $n_k(h)$ citations are allocated randomly to firms with the same probability. Formally, denote by $\mathcal{F}_k(h)$ the set of firms that have at least one patent in $\mathcal{P}_k(h)$. Each citation out of $n_k(h)$ is randomly allocated to firm $j \in \mathcal{F}_k(h)$ with probability

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

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

B.2.4 Identifying "Specific" Patents

In Sections 2.4 and B.4, I identify patents where inventors were likely to have access to some private knowledge relative to others. Formally, the procedure is the following. Monte-Carlo simulations from Section B.2.2 construct the distribution of counterfactual concentration measures $\{\mathcal{C}_{k,s}\}_{s=1}^{300}$ for each patent $k \in Main$. I call citing patents $m \in Citing_k$ "specific" to patent $k \in Main$ if they are responsible for the majority of citations from a particular firm, and the concentration measure for patent $i \in Main$ exceeds the $95th$ quantile in the Monte-Carlo simulations. Formally,

$$m \in Citing_k \text{ is "specific" to } k \in Main \Leftrightarrow m \in i = \arg\max_j \left\{ \frac{n_{k,j}}{n_k} \right\} \text{ and } \mathcal{C}_k > \mathcal{RC}_k(q95)$$

where $m \in i$ means that patent m is assigned to firm i, and $\mathcal{RC}_k(q95)$ denotes the 95th quantile of the distribution $\{\mathcal{C}_{k,s}\}_{s=1}^{300}$.

B.2.5 Concentration for Scientific Publications

Figure C3 compares the concentration of citations for patents and scientific publications. To construct the concentration measure for publications, I use Microsoft Academic Graph data (Marx & Fuegi (2020)). The concentration measure is constructed in a way similar to the measure for patents. For each year between 1976 and 2014 and each scientific discipline in Web of Science classification, I track citations within a five-year window for the top 1% of the most cited publications. For each patent, the concentration is defined as a share of citations coming from the most citing organization. These organizations are predominantly universities. To reduce the computational burden I take a random 10% sample of publications from the set of the most cited ones. Then, I take the average of the concentration measures within a field of science, and the average across fields of science weighted by the number of publications in each field. The purple squared line in C3 shows the average concentration measure for the most cited publications.

B.3 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 2.1 but within inventors-movers, and then I do the decomposition of the concentration measure similar to the one in Section 2.2. This exercise follows the same procedure as the Monte-Carlo exercises in Sections B.2.2 and B.2.3 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.[33] 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

[33]Matching of patents to Compustat firms is cleaner in a sense that I use the data from Autor et al. (2020a) 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.

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 system.[34] 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. The procedure is the same as in Section B.2.2. 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 procedure is the same as in Section B.2.3.

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 2(b) 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.4 Details to Section 2.4: Value of Citations for the Most Citing Firm

Section 2.4 estimates a regression of the firm's growth in sales and profits on the composition of its patent portfolio. For firm f and year t, this portfolio includes the total number of granted patents, N_{ft}, and the number of patents making at least one citation to the top 1% of the most cited patents, N_{ft}^g. Specifically,

$$N_{ft}^g = |\{m \in \mathcal{P} : \text{ there is } k \in Main \text{ such that } m \in Citing_k\}|$$

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

where $\mathcal{P}$ is the set of all granted patents, the sample $Main$ is defined in Section B.2.1, and the sample $Citing_k$ is defined in (B.4). Finally, the firm's portfolio of patents includes the number of "specific" patents N^c_{ft} that are responsible for the majority of citations to a highly-concentrated patent in the $Main$ sample. The precise definition is given in section B.2.4.

The details of specification (2.2) are the following. The outcome variable $y_{f,t}$ represents sales (Compustat variable "sale") or profits (Compustat variables: "sale" − "cogs"). Both variables are deflated by CPI (= 100 for 1982 year). Variables in X_{ft} include firm employment (Compustat variable "emp"), firm capital stock (Compustat variable "ppegt" deflated by the NIPA price of equipment), and a lag of profit or sales. All variables are winsorized at the 1% level using annual breakpoints. The variable $\zeta_{i(f),t}$ is an industry-year fixed effect where industry is defined using historical 4-digit SIC classification. For years with missing industry information I take the industry for the nearest year. Standard errors are clustered by firm and year.

B.5 Robustness to Section 2: Concentration of Citations

Panel (a) in Figure C7 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 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 C8 shows the concentration based on citations from patent examiners is around two times lower than the concentration based on citations from non-examiners.[35]

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

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.3. 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 C9 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.6 Details to Section 4.1: Citations in a Production Network

Section 4.1 estimates the probability of a citation between two patents as a function of the network relationship between firms that own them. I impose the following restrictions on the set of all citations between patents granted between 1976 and 2019. First, since the supplier-customer data is available for the U.S. publicly traded firms only, I leave only cited and citing patents owned by a publicly traded U.S. firm. Second, to control for truncation that older patents have more time to accumulate citations I consider only citations within a 10 year window. Third, to focus on citations between firms I exclude self-citations. Finally, I remove cited patents granted after 2014 to allow patents to accumulate some citations.[36] Formally, denote by $y_{km} = 1$ if patent m makes a citation to patent k and $y_{km} = 0$ otherwise. Denote the grant year of patent k by t_k^g, the assignee of a patent k by f_k. I focus on the following set of cited-citing pairs

$$\text{Cited-Citing Set} = \Big\{(k,m) \in \mathcal{P}^2 : y_{km} = 1, \text{Grant Date}_m - \text{Grant Date}_k \leqslant 10 \cdot 365,\\ t_k^g \leqslant 2014, f_k \neq f_m\Big\}$$

Next, following Jaffe et al. (1993), Thompson & Fox-Kean (2005), Singh & Marx (2013) I construct the control set of patents that have similar characteristics to citing ones. For each citing patent, I define the control set as all patents with the same application year, from the same main subgroup CPC class, and coming from a different firm. Formally, denote by t_k^a the application year of patent k, and by $\tilde{c}_k$ the main subgroup CPC class of patent k. Then the

[36] In all regression specifications, I use year fixed effects, which also control for differences in citations for patents granted in different years. I take the patents until 2014 in order to keep the sample of patents used for the computation of the concentration measure. I consider citations within a 10 year window (rather than 5) in order to increase the sample size. The results are robust when I restrict citations to be within a 5-year window, or take granted patents until 2009.

cited-control set for a pair $(k,m) \in$ Cited-Citing Set is

$$\text{Control Set}_{k,m} = \{r \in \mathcal{P} : t^a_r = t^a_m, \tilde{c}_r = \tilde{c}_m, f_r \neq f_m, f_r \neq f_k\}$$

I randomly choose one patent from Control Set$_{k,m}$, denote it by $r(k,m)$. Then the cited-control set is

$$\text{Cited-Control Set} = \Big\{(k,r) \in \mathcal{P}^2 : \text{ there is } m \in \mathcal{P} \text{ such that } (k,m) \in \text{Cited-Citing Set}, \text{ and } r = r(k,m)\Big\}$$

The estimation of (4.1) is done on the following sample of patent pairs

$$\text{Sample} = \text{Cited-Citing Set} \cup \text{Cited-Control Set}$$

For estimation of (4.1) I use the data on supplier-customer relationships. The details of the construction of this data are given in appendix B.1. Denote the first year when firm A reports firm B as a customer by t_{first}, and the last year by t_{last}. In rare cases, there might be gaps in years when A reports B as a customer, but this is likely to be caused by the 10%-threshold for reporting explained in Appendix B.1. Therefore, I assume that firm A is supplier to firm B in all years between t_{first} and t_{last}.

Controls include the spatial proximity between inventors. I define the distance (in nautical miles) between two patents by taking the minimum out of all pairwise distances between patents' inventors. I include a dummy whether two patents come from the same country or not. The country for a patent is defined by the country of the majority of inventors as in section B.2.2. The controls also include a dummy whether cited and citing patents are in the same main subgroup of CPC, and dummies for a lag between a citing (control) application year and cited grant year.

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. (2020a) (Appendix B.3) for the analysis at the technology class level. Specifically, I do the following.

First, I take the set of the top 1% of the most cited patents ($Main$ sample defined in appendix B.2.1). For the specification with corporate patents, I leave only patents assigned to corporate firms (both public and private). As in the previous sections, denote by f_k the assignee of patent k and by t^a_k the application year of patent k. I group patents into five 7-year periods based on the application year. Formally, I define the following sets

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

where the sample $Main$ is defined in Section B.2.1.

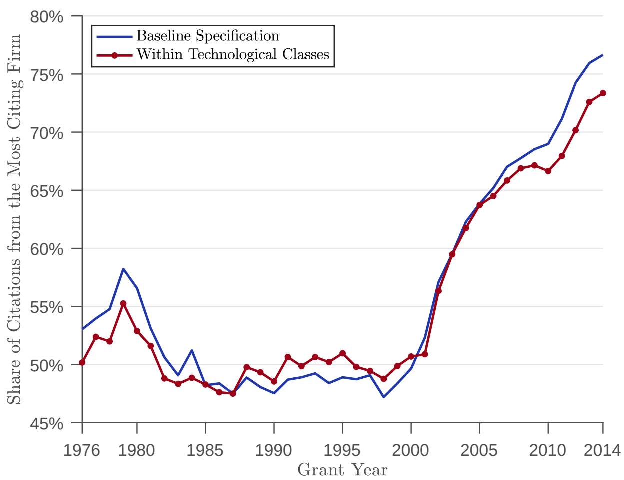
Second, for each set S_t and each technology class I compute the aggregated concentration measure. I take the simple average across concentration measures. Autor et al. (2020a) provides a mapping between USPC technological classes and SIC industries. Moreover, there exists a matching from USPC to NBER 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.

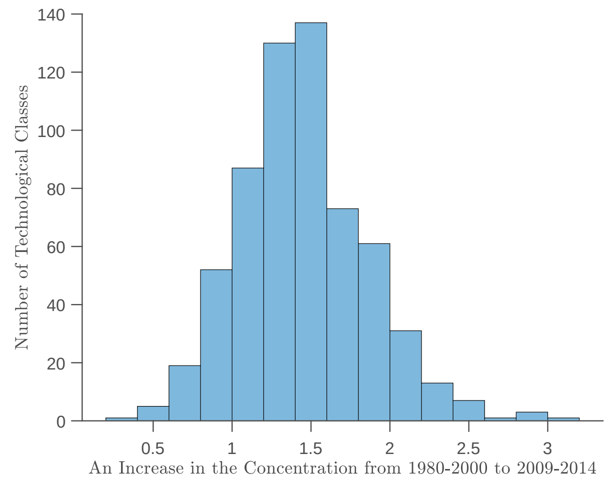
B.8 Joint Innovations Between Suppliers and Customers

In this section, I consider the dynamics of joint innovations between suppliers and customers. I focus on patents with multiple assignees where at least two of these assignees are from publicly traded firms in Compustat. Around 3% of patents granted in the USPTO since 1976 have at least two assignees. Multiple assignees on a patent mean that there are several organizations sharing property rights on a patent. To study the joint research between suppliers and customers, I compute two statistics. First, among patents with multiple assignees I compute the share of patents in which assignees have a vertical relationship. Panel (a) of Figure C5 shows that the share of joint patents between suppliers and customers among all joint patents has significantly increased over time. However, this rise might be caused by changes in the reporting requirements in 10-K files. In other words, the increase in the share of joint patents among suppliers and customers might be driven by greater disclosure of supplier-customer relationships over time rather than by greater activity in joint innovations. To address this concern, in Panel (b) I plot the share of supplier-customer pairs who had joint patents in a particular period relative to the total number of supplier-customer pairs in this period. This panel shows that the share of firms who reported a vertical relationship and had joint patents has increased over time.

C Figures and Tables



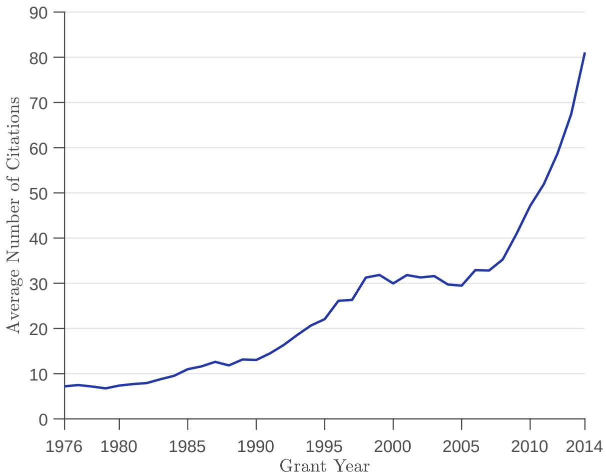
(a) Baseline vs. Within Technological Classes



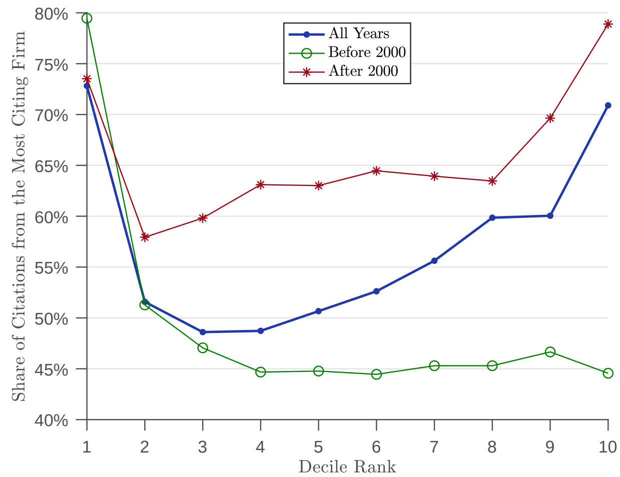
(b) Distribution of the Increase Across Classes

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 technology 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.

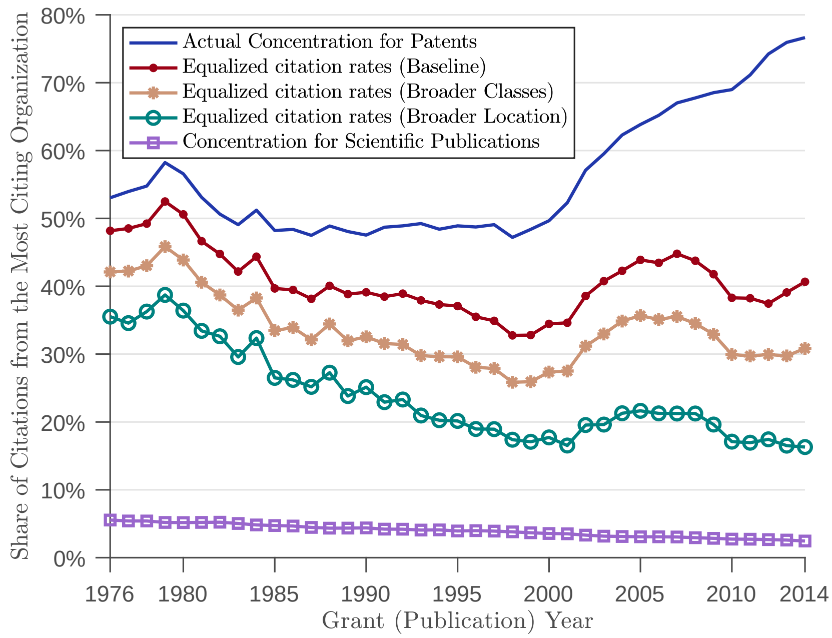


Figure C3: Concentration of Citations: Patents, Monte-Carlo, and Publications

This figure shows the average concentration of citations for the most cited patents and scientific publications between 1976 and 2014. For each cited patent and publication, the concentration is defined as the share of citations coming from the most citing organization. For patents the most citing organization is usually a corporate firm, and for scientific publications it is usually a university. The most citing patents are defined in the following way. For each patent, I track its citations within a 5 year from a grant date, and for each grant year and the group level in Cooperative Patent Classification, I take the top 1% of the most cited patents. The most citing scientific publications are defined in a similar way. For each year and the scientific field in Web of Science classification, I take the top 1% of the most cited publications where citations are taken within a 5 year from the publication year. Since the number of publications is much larger relative to the number of patents, to reduce the computational requirements I take a random sample of 10% publications from the most cited ones. The solid blue line shows that the actual concentration of patent citations. The dotted (red) line, the starred (brown) line, and the circled (green) line show the medians of the concentration measure where citation rates are equalized across firms. They are constructed using Monte-Carlo simulations where citations are allocated randomly across all patents observationally similar to the citing patents (see Section 2.2 and Appendix B.2.2). The dotted line shows the median counterfactual concentration from the baseline specification in Section 2.2, the starred line shows the median concentration when patents with similar characteristics are grouped based a location of the majority of inventors, an application year, and on a broader technological class level (group CPC). The circled line shows the median concentration in which patents with similar characteristics are grouped based only on an application year, and on a narrow technological class level (main subgroup CPC). The squared purple line shows the concentration of citations for scientific publications.

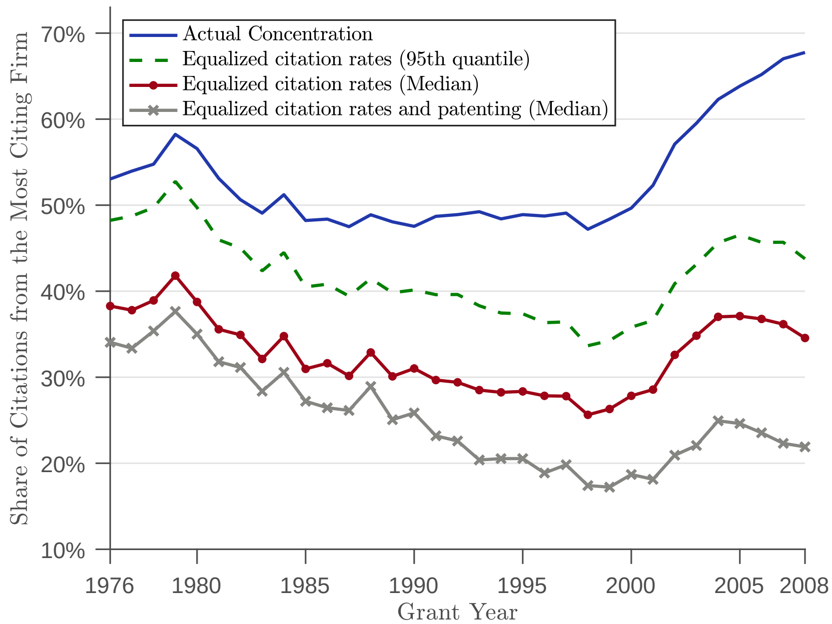
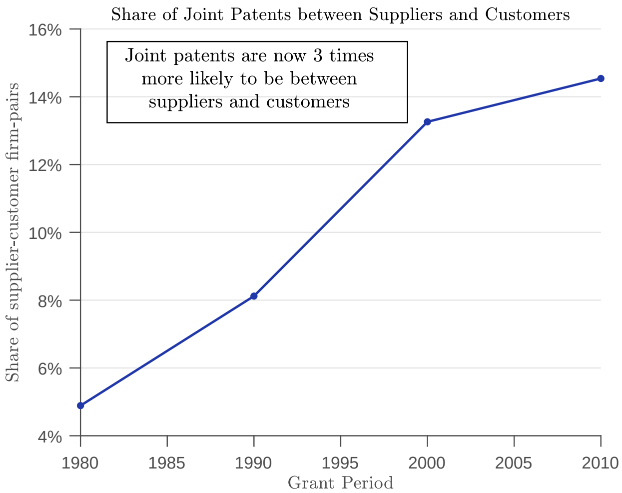
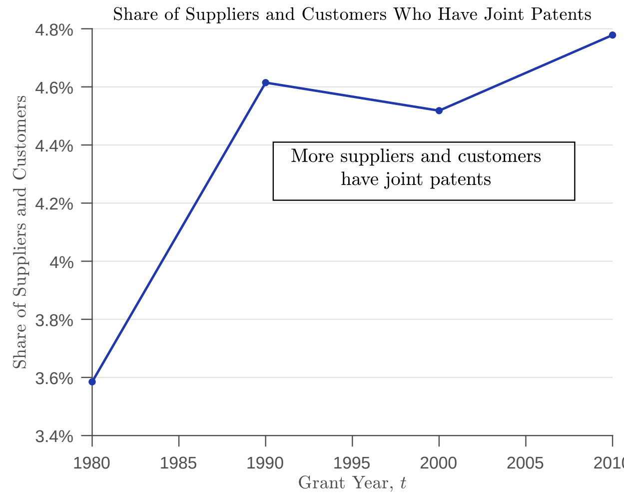


Figure C4: Decomposition of the Concentration of Citations Based on the Textual Similarity Between Patents

This figure shows the decomposition of the concentration of citations described in Section 2.2 and Figure 2(a) where instead of technological classes I group patents based on their textual similarity. Specifically, for each citing patent I find patents with the same application year, and then I choose at most 200 patents closest in terms of the textual similarity based on the data from Arts et al. (2018). Finally, among the selected patents I choose the ones where the majority of inventors are from the same location as for the citing patent. Then, I do Monte-Carlo exercises described in Section 2.2.



(a) Share of Joint Patents Among Suppliers and Customers.



(b) Share of Suppliers and Customers who have Joint Patents.

Figure C5: Joint Innovations Between Suppliers and Customers

Panel (a) shows the share of joint patents between suppliers and customers among all joint patents. Panel (b) shows the share of supplier-customer pairs who had joint patents in a particular period relative to the total number of supplier-customer pairs in this period.

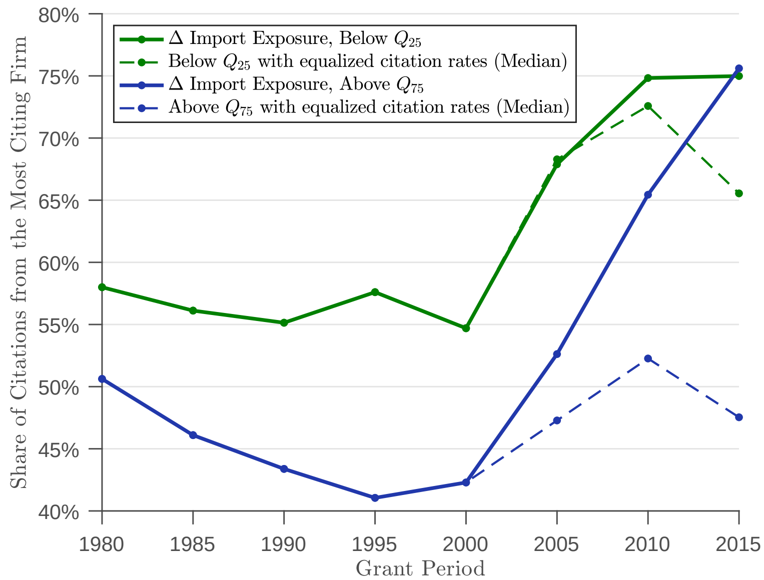
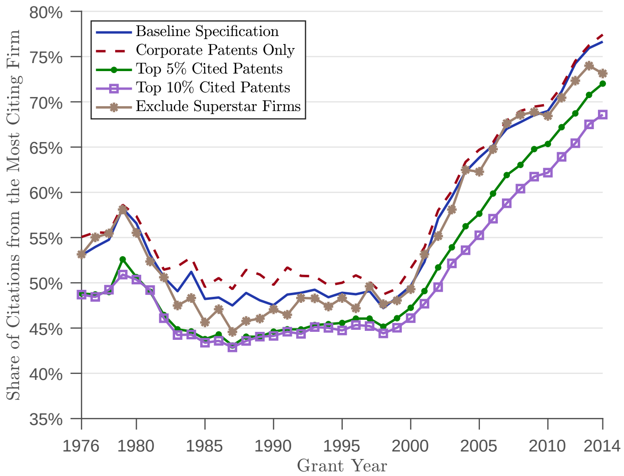
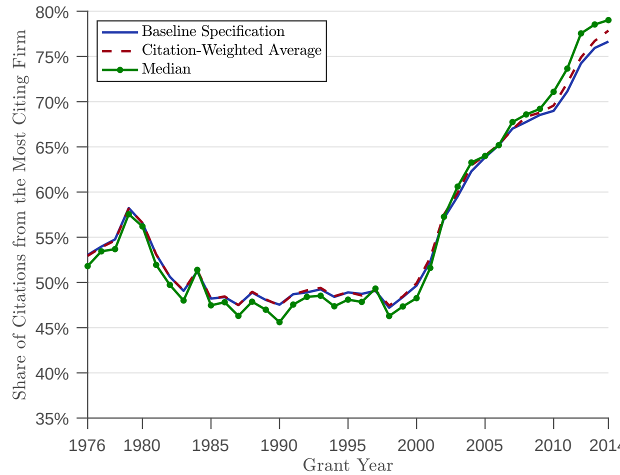


Figure C6: Trade with China. Actual Concentration vs. Counterfactual One with Equalized Citation Rates

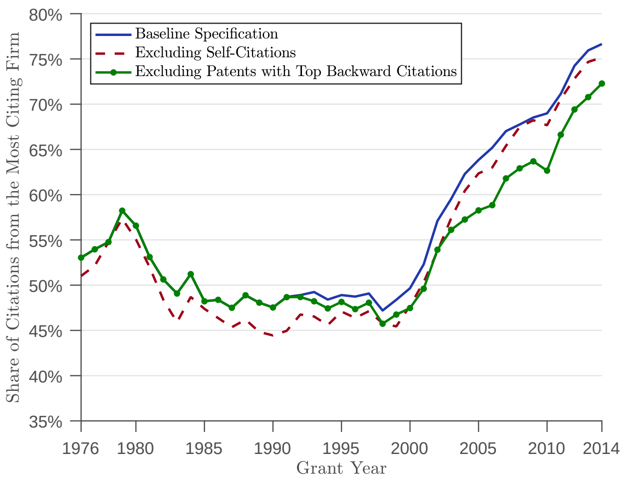
This figure shows the concentration measure from Section 2 (Figure 1) for different technological classes divided into quartiles based on their exposure to import competition from China, ΔIP_{i2} in (4.2). The solid lines show the concentration measures over time for the technologies most (the blue line) and least (the green line) exposed to 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 observationally similar to citing ones (see B.2.2). I add the difference between the actual and the counterfactual concentration measures in 2000.



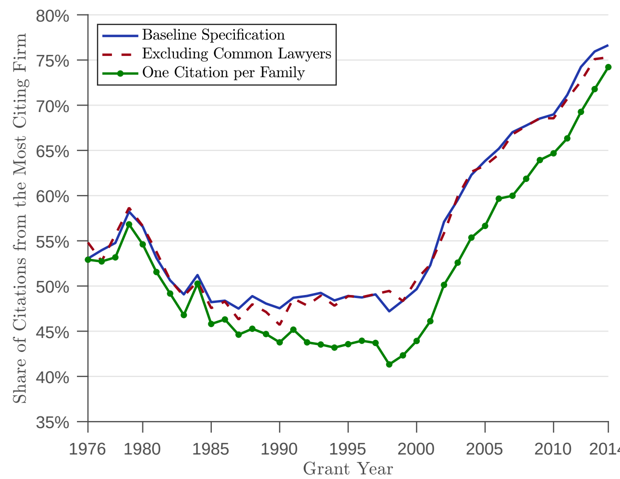
(a) Corporate patents, top 1%, top 10%, exclude superstar firms in patenting



(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 C7: 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 the sample of citing patents 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 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.

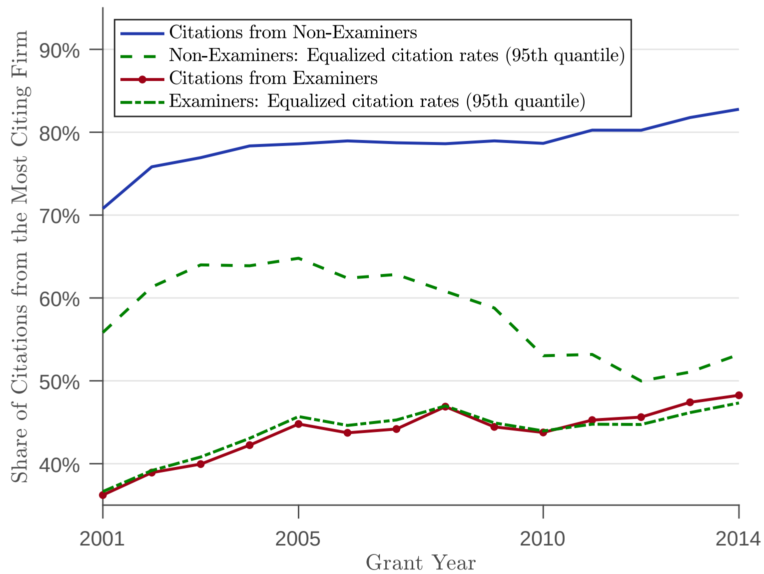


Figure C8: Concentration of Citations: Examiners vs. Non-examiners

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.

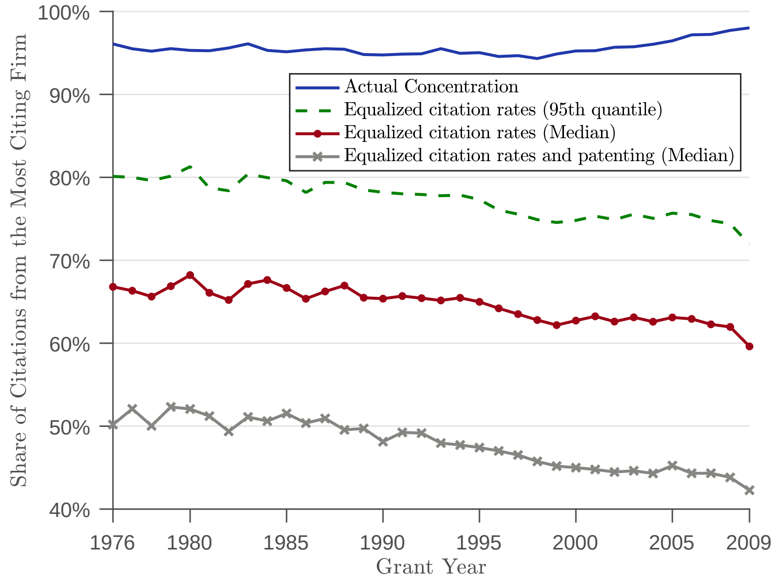


Figure C9: Concentration of Citations for Lawyers Who Represented Several Firms

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, the same geographical location (states in the U.S., and countries for foreign inventors), 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 an inventor.

Table C1: Distribution of Citations across firms for IBM/Novellus's patent 6126798.

Citing Firm	% of Citations	Relationship to IBM	Source
Novellus Systems	64%	Supplier	Compustat Segments, Lim (2009)
Lam Research	16%	Supplier	IBM's website
Applied Materials	11%	Supplier	Lim (2009)
Semitool Inc	3%	Supplier	Compustat Segments, Lim (2009)
Toshiba, Ebara	2%	Share a Supplier	Compustat Segments
Entegris Inc	0.05%	Joint Collaboration	Entegris's website
IBM	0.05%	Self-Citations	—
Other	3%	No Info	—

Table C2: Distribution of Citations across firms for Motorola's patent 5391517.

Citing Firm	% of Citations	Relationship to Motorola	Source
Semitool Inc	13%	Supplier	Compustat Segments
ASM International	11%	Semitool's Industry	Compustat
IBM	9%	Share a Supplier	Compustat Segments
Applied Materials	7%	Supplier	Compustat Segments
U.S. Navy	7%	No Info	—
Taiwan Semiconductor Manufacturing	5%	Share a Supplier	Compustat Segments
Advanced Micro Devices	4%	Share a Supplier	Compustat Segments
Intel Corp	4%	Share a Supplier	Compustat Segments
Motorola	4%	Self-Citations	—
Other Suppliers	3%	Suppliers	Compustat Segments
Other Firms Sharing a Supplier	17%	Share a Supplier	Compustat Segments
Other Firms	7%	No Info	—

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 to Chinese Imports	0.83 (0.71)	0.79 (0.70)	0.74 (0.64)	1.67 (1.28)	1.68 (1.30)	−0.10 (0.71)	−0.10 (0.72)	−0.71 (1.25)	−0.82 (1.25)
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 5. Changes in US import exposure are instrumented with Chinese exports to non-U.S. high-income markets (Autor et al. (2020a)). 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.

D Theory Appendix

D.1 Proposition 1: $\mathcal{I}$'s profits under different contracts

If $\mathcal{I}$ and $\mathcal{S}$ sign exclusive dealing with a confidentiality agreement, then under knowledge sharing $\mathcal{I}$'s profits in the final stage will be

$$\begin{aligned}\pi_{\mathcal{I}}(ex=1,ca=1) &= \frac{1-\alpha}{2}(1-\gamma)\lambda + \alpha(1-\gamma)\left(\lambda-\lambda_0\right) + \\ &+ \frac{1}{2}\Big(\frac{1-\alpha}{2}\lambda + \alpha\left(\lambda-(1-\gamma)\lambda_0\right) - \frac{1-\alpha}{2}(1-\gamma)\lambda - \alpha(1-\gamma)\left(\lambda-\lambda_0\right)\Big) \\ &= \frac{1-\alpha}{2}(1-\gamma)\lambda + \alpha(1-\gamma)\left(\lambda-\lambda_0\right) + \frac{1+\alpha}{4}\gamma\lambda\end{aligned} \tag{D.1}$$

If $\mathcal{I}$ and $\mathcal{S}$ sign a confidentiality agreement only, then under knowledge sharing $\mathcal{I}$'s profits in the final stage will be

$$\begin{aligned}\pi_{\mathcal{I}}(ex=0,ca=1) &= \frac{1-\alpha}{2}(1-\gamma)\lambda + \alpha\cdot\max\left\{0,(1-\gamma)\lambda-\lambda_0\right\} + \\ &+ \frac{1}{2}\cdot\max\left\{0,\Big(\frac{1-\alpha}{2}\gamma\lambda + \alpha(\lambda-\lambda_0) - \alpha\cdot\left|(1-\gamma)\lambda-\lambda_0\right|\Big)\right\} \\ &= \begin{cases}\frac{1-\alpha}{2}(1-\gamma)\lambda + \alpha(\lambda-\lambda_0) + \frac{1-3\alpha}{4}\gamma\lambda & \text{if } \alpha\cdot\big(3\gamma-4(1-\frac{\lambda_0}{\lambda})\big)\leqslant\gamma \\ \frac{1-\alpha}{2}(1-\gamma)\lambda & \text{if } \alpha\cdot\big(3\gamma-4(1-\frac{\lambda_0}{\lambda})\big)>\gamma\end{cases}\end{aligned} \tag{D.2}$$

If $\mathcal{I}$ and $\mathcal{S}$ do not sign any contract in the first stage, then under knowledge sharing $\mathcal{I}$'s profits in the final stage will be

$$\begin{aligned}\pi_{\mathcal{I}}(ex=0,ca=0) &= \frac{1-\alpha}{2}(1-\gamma)\lambda + \\ &+ \frac{1}{2}\cdot\max\left\{0,\Big(\frac{1-\alpha}{2}\lambda + \alpha(\lambda-\lambda) - \frac{1-\alpha}{2}(1-\gamma)\lambda - \alpha\left(\lambda-(1-\gamma)\lambda\right)\Big)\right\} \\ &= \frac{1-\alpha}{2}(1-\gamma)\lambda + \max\left\{0,\frac{1-3\alpha}{4}\gamma\lambda\right\}\end{aligned} \tag{D.3}$$

In all cases, the first line is $\mathcal{I}$'s outside option, the second line is a half of the gain in the total surplus of $\mathcal{I}$ and $\mathcal{S}$, and the third line is a simplified expression.

If $\mathcal{I}$ does not share its knowledge with $\mathcal{S}$, then its profits in the final stage will be

$$\pi_{\mathcal{I}}\left(\text{No Sharing}\right) = \frac{1-\alpha}{2}(1-\gamma)\lambda + \alpha(1-\gamma)\left(\lambda-\lambda_0\right) \tag{D.4}$$

For $\alpha > \frac{1}{3}$, $\mathcal{I}$ and $\mathcal{E}$ prefer to sign exclusive dealing, and $\mathcal{I}$'s profits under knowledge sharing with exclusive dealing are higher under knowledge sharing with $\mathcal{S}$ than without sharing. Under $\alpha \leqslant \frac{1}{3}$, $\mathcal{I}$'s profits under a confidentiality agreement exceed its profits without knowledge sharing. For the case without a contract (Null), $\mathcal{I}$ shares knowledge with $\mathcal{S}$ under condition (3.6).

D.2 Generalizations in the Competition, Production, and Bargaining Structures

In this section, I consider an extension of the simple model in the competition, production and bargaining structures. Denote by $h_{\mathcal{I}} = 1$ if $\mathcal{I}$ uses a high-quality input from $\mathcal{S}$ (and shares its secrets with $\mathcal{S}$) and $h_{\mathcal{I}} = 0$ otherwise, $h_{\mathcal{E}} = 1$ if $\mathcal{E}$ uses a high-quality input from $\mathcal{S}$ and $h_{\mathcal{E}} = 0$ otherwise, $\ell = 1$ if $\mathcal{S}$ leaks $\mathcal{I}$'s secrets to $\mathcal{E}$ and $\ell = 0$ otherwise. I extend the competition and production structures by considering a more general mapping from $e = (h_{\mathcal{I}}, h_{\mathcal{E}}, \ell)$ to downstream profits, $\pi_{\mathcal{I}}(e)$ and $\pi_{\mathcal{E}}(e)$.[37] I also consider a generalized Nash bargaining between parties: denote by $\psi_{\mathcal{I}}$ and $\psi_{\mathcal{E}}$ the bargaining powers of $\mathcal{S}$ in negotiations with $\mathcal{I}$ and $\mathcal{E}$, respectively.

In Section 3.1, the supplier always has incentives to share $\mathcal{I}$'s knowledge with $\mathcal{E}$ in the third stage. In the general framework, $\mathcal{S}$ has incentives to shares $\mathcal{I}$'s knowledge with $\mathcal{E}$ if its input is complementary to $\mathcal{I}$'s knowledge in $\mathcal{E}$'s profits

$$\Delta\pi_{\mathcal{E}}\left(\cdot, \ell = 1\right) > \Delta\pi_{\mathcal{E}}(\cdot, \ell = 0) \tag{D.5}$$

where $\Delta\pi_{\mathcal{E}}\left(\cdot, \ell\right) = \pi_{\mathcal{E}}\left(h_{\mathcal{I}} = 1, h_{\mathcal{E}} = 1, \ell\right) - \pi_{\mathcal{E}}\left(h_{\mathcal{I}} = 1, h_{\mathcal{E}} = 0, \ell\right)$. In this case, $\mathcal{S}$ can bargain a higher input price from $\mathcal{E}$ ($t_{\mathcal{E}}(\ell) = \psi_{\mathcal{E}} \cdot \Delta\pi_{\mathcal{E}}\left(\cdot, \ell\right)$) under knowledge sharing ($\ell = 1$). Assume that (D.5) is satisfied.

$\mathcal{I}$'s incentives to share its secrets with $\mathcal{S}$ in the second stage depend on its expectations on whether $\mathcal{E}$ will have access to a high-quality input ($h_{\mathcal{E}} \in \{0, 1\}$) and whether $\mathcal{S}$ will leak $\mathcal{I}$'s secrets to $\mathcal{E}$ ($\ell \in \{0, 1\}$). $\mathcal{I}$ shares its secrets with $\mathcal{S}$ if

$$\underbrace{\pi_{\mathcal{I}}\left(h_{\mathcal{I}} = 0, h_{\mathcal{E}}, \ell\right) + (1 - \psi_{\mathcal{I}}) \cdot \Big(TS(h_{\mathcal{E}}, \ell) - TS^{o}(h_{\mathcal{E}}, \ell)\Big)}_{\mathcal{I}\text{'s profits under knowledge sharing with } \mathcal{S}} > \underbrace{\pi_{\mathcal{I}}\left(h_{\mathcal{I}} = 0, h_{\mathcal{E}} = 0, \ell = 0\right)}_{\mathcal{I}\text{'s profits without knowledge sharing}} \tag{D.6}$$

[37] There are minimal assumptions on profits: $\pi_{\mathcal{I}}(e)$ ($\pi_{\mathcal{E}}(e)$) is increasing (decreasing) in $h_{\mathcal{I}}$ and decreasing (increasing) in $(h_{\mathcal{E}}, \ell)$. That is, firm $\mathcal{I}$ values a high-quality input from $\mathcal{S}$, and its profit is lower if firm $\mathcal{E}$ uses $\mathcal{I}$'s knowledge and a high-quality input from $\mathcal{S}$. Firm $\mathcal{E}$ values a high-quality input from $\mathcal{S}$ and the knowledge from $\mathcal{I}$, but its profit is lower if $\mathcal{I}$ uses a high-quality input. In Section 3.1, profits are the following

$$\pi_{\mathcal{I}}(h_{\mathcal{I}}, h_{\mathcal{E}}, \ell) = \frac{1-\alpha}{2}\Big(1 - \gamma \cdot (1 - h_{\mathcal{I}})\Big)\lambda + \alpha\Big[\lambda - (1 - (1 - h_{\mathcal{E}})\gamma) \cdot (\ell \cdot \lambda + (1 - \ell)\lambda_0)\Big]$$
$$\pi_{\mathcal{E}}(h_{\mathcal{I}}, h_{\mathcal{E}}, \ell) = \frac{1-\alpha}{2}\Big(1 - \gamma \cdot (1 - h_{\mathcal{E}})\Big) \cdot (\ell \cdot \lambda + (1 - \ell)\lambda_0)$$

where $TS(h_\mathcal{E}, \ell) = \pi_\mathcal{I}(h_\mathcal{I} = 1, h_\mathcal{E}, \ell) + \psi_\mathcal{E} \cdot \Delta\pi_\mathcal{E}(h_\mathcal{I} = 1, h_\mathcal{E}, \ell)$ is a joint surplus of $\mathcal{I}$ and $\mathcal{S}$ when $\mathcal{I}$ uses a high-quality input from $\mathcal{S}$, and $TS^o(h_\mathcal{E}, \ell) = \pi_\mathcal{I}(h_\mathcal{I} = 0, h_\mathcal{E}, \ell) + \psi_\mathcal{E} \cdot \Delta\pi^o_\mathcal{E}(h_\mathcal{I} = 0, h_\mathcal{E}, \ell)$ is a joint surplus of $\mathcal{I}$ and $\mathcal{S}$ when $\mathcal{I}$ does not use a high-quality input from $\mathcal{S}$.

A contract (ex, ca) between $\mathcal{I}$ and $\mathcal{S}$ in the first stage will define the equilibrium $(h^\star_\mathcal{E}, \ell^\star)$. Exclusive dealing $(ex = 1)$ restricts S's ability to sell its input to $\mathcal{E}$ $(h^\star_\mathcal{E} = 1 - ex)$, and a confidentiality agreement $(ca = 1)$ restricts S's ability to communicate with $\mathcal{E}$ $(\ell^\star = 1 - ca)$. $\mathcal{I}$ and $\mathcal{S}$ choose a contract to maximize their joint surplus subject to condition (D.6) that $\mathcal{I}$ has incentives to share its secrets with $\mathcal{S}$ in Stage 2 $(h_\mathcal{I} = 1)$:

$$(h^\star_\mathcal{E}, \ell^\star) = \arg\max_{(h_\mathcal{E}, \ell)} \left\{ \pi_\mathcal{I}(h_\mathcal{I} = 1, h_\mathcal{E}, \ell) + \psi_\mathcal{E} \cdot \Delta\pi_\mathcal{E}(h_\mathcal{I} = 1, h_\mathcal{E}, \ell) \right\} \text{ subject to (D.6)}$$

The predictions in the extended model are similar to the simple one. First, the stronger outside option (lower $\Delta\pi_\mathcal{E}(\cdot)$ holding $\pi_\mathcal{I}(\cdot)$ fixed) leads to the equilibrium without knowledge sharing. Second, higher competition between $\mathcal{I}$ and $\mathcal{E}$ leads to the equilibrium without knowledge sharing. The competition can be measured based on the responsiveness of $\pi_\mathcal{I}(h_\mathcal{I}, h_\mathcal{E}, \ell)$ to changes in $(h_\mathcal{E}, \ell)$.

D.3 Richer Contracting: Stipulated Damages

In Section 3.1, $\mathcal{I}$ and $\mathcal{S}$ can sign exclusive dealing and a confidentiality agreement. These contracts prevent $\mathcal{S}$ from selling an input to $\mathcal{E}$ and sharing $\mathcal{I}$'s secrets with $\mathcal{E}$. In this section, I consider the possibility of stipulated damages (Aghion & Bolton (1987)) where $\mathcal{I}$ and $\mathcal{S}$ specify a fee F for breaking a contract. This might allow $\mathcal{I}$ and $\mathcal{E}$ to extract more surplus from $\mathcal{E}$.

Suppose that in the first stage $\mathcal{I}$ and $\mathcal{S}$ sign exclusive dealing with a confidentiality agreement and specify a fee F that $\mathcal{S}$ has to pay to $\mathcal{I}$ for breaking this contract. Assume $\mathcal{I}$ shares its knowledge with $\mathcal{S}$, and that after $\mathcal{E}$ enters, $\mathcal{S}$ can propose $\mathcal{E}$ to pay F in order to break the contract between $\mathcal{I}$ and $\mathcal{S}$. If $\mathcal{E}$ agrees to pay F, then S will share $\mathcal{I}$'s secrets with $\mathcal{E}$ and will make a take-it-or-leave-it offer about the input price $t_\mathcal{E} = \frac{1-\alpha}{2}\lambda - \frac{1-\alpha}{2}(1-\gamma)\lambda = \frac{1-\alpha}{2}\gamma\lambda$. Therefore, $\mathcal{E}$ agrees to pay F if

$$F < F^\star = \frac{1-\alpha}{2}\lambda - t_\mathcal{E} - \frac{1-\alpha}{2}(1-\gamma)\lambda_0 = \frac{1-\alpha}{2}(1-\gamma)(\lambda - \lambda_0)$$

Assume that in the case of indifference $\mathcal{E}$ agrees to pay F. Then, the joint surplus of $\mathcal{I}$ and $\mathcal{S}$ under the contract with stipulated damages is

$$\frac{1-\alpha}{2}\lambda + t_\mathcal{E} + F^\star = \frac{1-\alpha}{2}(\lambda + \gamma\lambda_0) + \frac{1-\alpha}{2}(\lambda - \lambda_0)$$

which maximizes the joint surplus of $\mathcal{I}$ and $\mathcal{S}$ for $\alpha < \frac{1}{3}$ (compare to (3.3) – (3.5)). To sum up, for $\alpha > \frac{1}{3}$ firms $\mathcal{I}$ and $\mathcal{S}$ sign exclusive dealing *without* stipulated damages (or set them $F > F^\star$). In this case, only $\mathcal{S}$ learns $\mathcal{I}$'s secrets. For $\alpha < \frac{1}{3}$, firms $\mathcal{I}$ and $\mathcal{S}$ sign exclusive dealing and a confidentiality agreement *with* stipulated damages $F = F^\star$. In this case, both $\mathcal{S}$ and $\mathcal{E}$ learn $\mathcal{I}$'s secrets.

The ability to use stipulated damages increases social welfare because it ensures a knowledge transfer from $\mathcal{I}$ to $\mathcal{E}$ in cases where previously $\mathcal{I}$ with $\mathcal{S}$ signed a confidentiality agreement (see Figure 5). This happens because $\mathcal{I}$ is compensated for the knowledge transfer to $\mathcal{E}$ through stipulated damages while before contractual limitations prevented this compensation.

D.4 Welfare Implications of Restrictions on Vertical Contracts

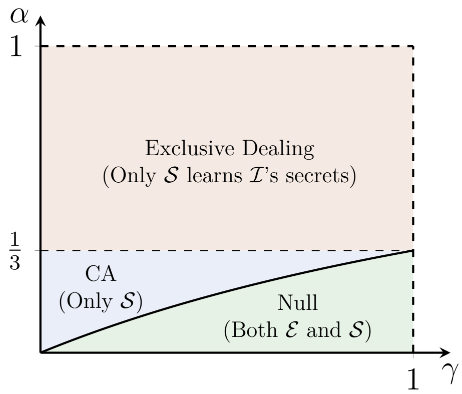
The model in Section 3.1 predicts the equilibrium vertical contract as a function of the competition level (α) and the supplier's bargaining position (γ). In this section, I study the welfare properties of the equilibrium and policies that can increase social welfare. Social welfare is defined as the sum of consumer and producer surpluses and is determined by the consumers' willingness to pay

$$SW(q_{\mathcal{I}}, q_{\mathcal{E}}) = \frac{1-\alpha}{2} \cdot (q_{\mathcal{I}} + q_{\mathcal{E}}) + \alpha \cdot \max\{q_{\mathcal{I}}, q_{\mathcal{E}}\}$$

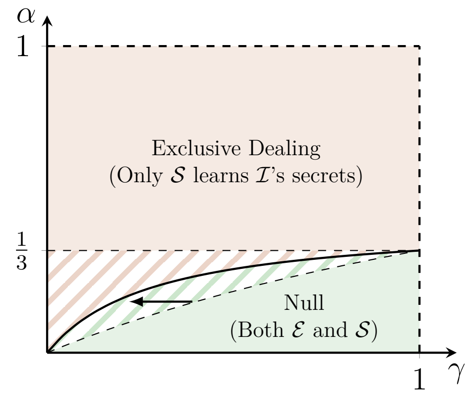
where $q_{\mathcal{I}}$ and $q_{\mathcal{E}}$ are qualities of products from firms $\mathcal{I}$ and $\mathcal{E}$, respectively.

The game in Section 3.1 has four possible outcomes with different welfare implications. First, both $\mathcal{E}$ and $\mathcal{S}$ learn $\mathcal{I}$'s private knowledge, and $\mathcal{E}$ uses a high-quality input from $\mathcal{S}$. This is a socially optimal outcome that gives the largest total welfare. It is achieved when $\mathcal{I}$ and $\mathcal{S}$ do not sign any contracts in the first stage (neither exclusive dealing nor a confidentiality agreement), and $\mathcal{I}$ shares knowledge with $\mathcal{S}$. The second outcome is when only $\mathcal{S}$ learns $\mathcal{I}$'s private knowledge, but $\mathcal{E}$ can still use $\mathcal{S}$'s high-quality input. This equilibrium has lower welfare because $\mathcal{E}$-product is of lower quality. This happens when $\mathcal{I}$ and $\mathcal{S}$ sign a confidentiality agreement. The third outcome is when only $\mathcal{S}$ learns $\mathcal{I}$'s private knowledge and $\mathcal{E}$ uses a low-quality input. This equilibrium gives lower welfare than the previous two cases because firm $\mathcal{E}$ works with an inferior input and without $\mathcal{I}$'s knowledge. This happens when $\mathcal{I}$ and $\mathcal{S}$ sign exclusive dealing. The final outcome is when nobody learns $\mathcal{I}$'s secrets, and both $\mathcal{I}$ and $\mathcal{E}$ work with an inferior input. However, this outcome is never an equilibrium if $\mathcal{I}$ and $\mathcal{S}$ can sign exclusive dealing. To sum up, the equilibrium outcomes (contracts) can be ranked in terms welfare as

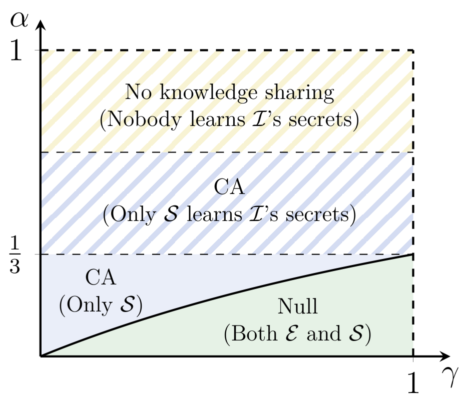
$$\text{Null} \succ_w \text{CA} \succ_w \text{Exclusive} \succ_w \text{No Knowledge Sharing}$$



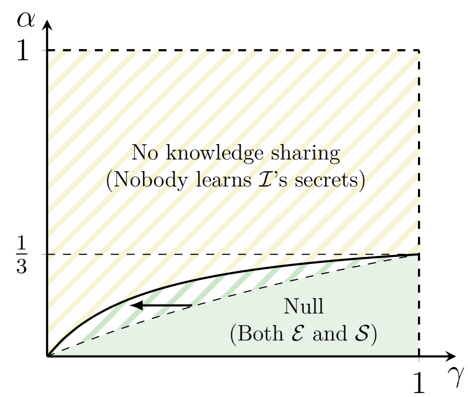
(a) No Restrictions



(b) Ban on CAs



(c) Ban on Exclusive Dealing



(d) Ban on Both

Figure D10: Equilibrium Under Restrictions on Vertical Contracts

Panel (a) shows the equilibrium without any policy interventions (Proposition 1). Panels (b), (c), and (d) show the equilibrium under the restrictions on the use of confidentiality agreements, exclusive dealing, and both, respectively. Hatched areas show the regions where the equilibrium has changed due to a policy regulation. Colors reflect an equilibrium that has replaced the initial one.

Policy. Proposition 1 shows that the socially optimal equilibrium arises only under a low level of competition (low α) and high quality of the supplier's input (γ). I discuss benefits and costs of three potential policy interventions: the restrictions on the use of confidentiality agreements, the restrictions on the use of exclusive contracts, and the restrictions on both. The equilibrium in each of these cases is described in Proposition 2.

Proposition 2. *Consider the model with a common supplier who can sell inputs to two downstream firms from Section 3.1. With restrictions on the use of confidentiality agreements, the equilibrium contract between $\mathcal{I}$ and $\mathcal{S}$ is*

$$\begin{cases} \text{Exclusive Dealing} & \text{if } \alpha > \frac{\gamma}{4(1-\frac{\lambda_0}{\lambda})+\gamma(4\frac{\lambda_0}{\lambda}-1)} \\ \text{Null} & \text{if } \alpha \leqslant \frac{\gamma}{4(1-\frac{\lambda_0}{\lambda})+\gamma(4\frac{\lambda_0}{\lambda}-1)} \end{cases} \tag{D.7}$$

Under exclusive dealing only $\mathcal{S}$ learns $\mathcal{I}$'s secrets, and without any contract between $\mathcal{I}$ and $\mathcal{S}$ both $\mathcal{S}$ and $\mathcal{E}$ learn $\mathcal{I}$'s secrets.

With restrictions on the use of exclusive dealing, if $\alpha \cdot \left(4\frac{\lambda_0}{\lambda} - 1\right) > 1$, then $\mathcal{I}$ will not share its secrets with $\mathcal{S}$ regardless of whether they have a confidentiality agreement or not. If $\alpha \cdot \left(4\frac{\lambda_0}{\lambda} - 1\right) \leqslant 1$, the equilibrium contract is

$$\begin{cases} \text{Confidentiality Agreement} & \text{if } \alpha > \min\left\{\frac{\gamma}{\gamma+2}, \frac{\gamma}{4(1-\frac{\lambda_0}{\lambda})+\gamma(4\frac{\lambda_0}{\lambda}-1)}\right\} \\ \text{Null} & \text{if } \alpha \leqslant \min\left\{\frac{\gamma}{\gamma+2}, \frac{\gamma}{4(1-\frac{\lambda_0}{\lambda})+\gamma(4\frac{\lambda_0}{\lambda}-1)}\right\} \end{cases} \tag{D.8}$$

Under a confidentiality agreement only $\mathcal{S}$ learns $\mathcal{I}$'s secrets, and without any contract between $\mathcal{I}$ and $\mathcal{S}$ both $\mathcal{S}$ and $\mathcal{E}$ learn $\mathcal{I}$'s secrets.

With restrictions on the use of confidentiality agreements and exclusive dealing, both $\mathcal{S}$ and $\mathcal{E}$ learn $\mathcal{I}$'s secrets if

$$\alpha \leqslant \frac{\gamma}{4(1-\frac{\lambda_0}{\lambda})+\gamma(4\frac{\lambda_0}{\lambda}-1)} \tag{D.9}$$

and nobody learns $\mathcal{I}$'s secrets otherwise.

The socially optimal regulation of confidentiality agreements and exclusive dealing is

$$\begin{cases} \text{Ban on CA} & \text{if } \alpha \leqslant \frac{\gamma}{4(1-\frac{\lambda_0}{\lambda})+\gamma(4\frac{\lambda_0}{\lambda}-1)} \\ \text{Ban on Exclusive Dealing} & \text{if } \alpha \leqslant \frac{\gamma}{4(1-\frac{\lambda_0}{\lambda})+\gamma(4\frac{\lambda_0}{\lambda}-1)} \text{ and } \alpha \cdot \left(4\frac{\lambda_0}{\lambda} - 1\right) \leqslant 1 \\ \text{No Restrictions} & \text{if } \alpha \cdot \left(4\frac{\lambda_0}{\lambda} - 1\right) > 1 \end{cases} \tag{D.10}$$

Proof. With restrictions on confidentiality agreements, $\mathcal{I}$ and $\mathcal{S}$ have two options: sign exclusive dealing or do not sign any contracts. Under exclusive dealing, $\mathcal{S}$ cannot gain from knowledge sharing with $\mathcal{E}$ because it does not sell an input to it. I assume that in this case it does not share $\mathcal{I}$'s secrets with $\mathcal{E}$.

Joint surplus of $\mathcal{I}$ and $\mathcal{S}$ is higher without a contract relative to exclusive dealing if

$$\frac{1-\alpha}{2}\lambda + \frac{1-\alpha}{2}\gamma\lambda \geqslant \frac{1-\alpha}{2}\lambda + \alpha\left(\lambda - (1-\gamma)\lambda_0\right) \Leftrightarrow \alpha \leqslant \frac{\gamma}{2\left(1-\frac{\lambda_0}{\lambda}\right) + \gamma\left(2\frac{\lambda_0}{\lambda}+1\right)}$$

Without a contract, $\mathcal{I}$ has incentives to share its secrets with $\mathcal{S}$ if condition (D.9) is satisfied. The derivation is the same as in Proposition 1. Since

$$\frac{\gamma}{4(1-\frac{\lambda_0}{\lambda})+\gamma(4\frac{\lambda_0}{\lambda}-1)} \leqslant \frac{\gamma}{2\left(1-\frac{\lambda_0}{\lambda}\right) + \gamma\left(2\frac{\lambda_0}{\lambda}+1\right)} \text{ for any } \gamma \in [0,1]$$

$\mathcal{I}$ and $\mathcal{S}$ sign exclusive dealing only when (D.9) is not satisfied. This leads to the equilibrium contract given in (D.7).

With restrictions on exclusive dealing, $\mathcal{I}$ and $\mathcal{S}$ have two options: sign a confidentiality agreement or do not sign any contracts. For $\alpha \leqslant \frac{1}{3}$, the equilibrium contract is the same as in Proposition 1 because these two options dominate exclusive dealing. For $\alpha > \frac{1}{3}$, $\mathcal{I}$ does not share its secrets with $\mathcal{S}$ without any contract. Under a confidentiality agreement it shares them only it its profit under sharing (D.2) is greater than the profit without knowledge sharing (D.4). This condition can be simplified to

$$\alpha \cdot \left(4\frac{\lambda_0}{\lambda} - 1\right) \leqslant 1 \tag{D.11}$$

If (D.11) is satisfied, then the equilibrium contract is described in (D.8). Otherwise, $\mathcal{I}$ does not share its secrets with $\mathcal{S}$.

Finally, with restrictions on both exclusive dealing and confidentiality agreements, $\mathcal{I}$ compares its profit under knowledge sharing (D.3) relative to the profit without sharing (D.4). This condition is given in (D.9).

The analysis above leads to the optimal regulation in (D.10).

□

Panel (b) of Figure D10 shows the equilibrium under the restriction on the use of confidentiality agreements (CA). In this case, only an exclusive contract is available to $\mathcal{I}$ and $\mathcal{S}$ in the first stage. Hatched areas in Panel (b) show the regions (α, γ) where the equilibrium has changed due to a policy regulation. Colors reflect an equilibrium that has replaced the initial one. If the initial equilibrium is inefficient due to a confidentiality agreement, the ban on CA can achieve a socially optimal outcome if $\mathcal{I}$ would continue sharing its secrets with $\mathcal{S}$ without any contract (condition (3.6)). Otherwise, $\mathcal{I}$ and $\mathcal{S}$ sign exclusive dealing that gives lower social welfare relative to the initial equilibrium because $\mathcal{E}$ does not get access to $\mathcal{I}$'s secrets and works with an inferior input.

Panel (c) shows the equilibrium under the restriction on the use of exclusive contracts. If the initial equilibrium is inefficient due to exclusive dealing ($\alpha > \frac{1}{3}$), the ban on exclusive contracts can never achieve the socially optimal outcome, but it can increase welfare by providing firm $\mathcal{E}$ the access to the high-quality input from $\mathcal{S}$. However, for sufficiently high level of competition (α) this ban might lead to lower welfare because $\mathcal{I}$ refuses to share its secrets with $\mathcal{S}$.

Panel (d) shows the equilibrium under the restriction on both confidentiality agreements and exclusive contracts. In this case, $\mathcal{I}$ shares knowledge with $\mathcal{S}$ only if the competition (α) is low enough and the gain in the input quality (γ) is high enough (condition (3.6)). Otherwise, $\mathcal{I}$ does share its secrets with anybody, and this gives the lowest possible welfare.

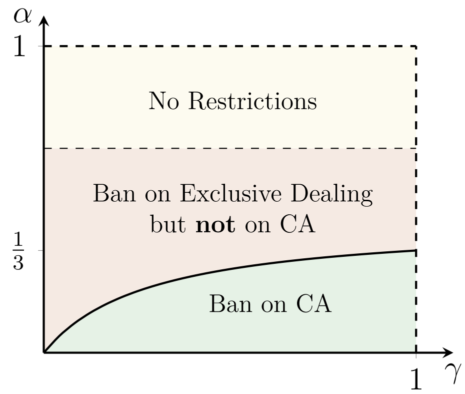


Figure D11: Optimal policy as a function of (α, γ)

This figure illustrates the optimal restrictions on confidentiality and exclusive contracts as a function of (α, γ). Parameter α measures the degree of competition between $\mathcal{I}$ and $\mathcal{E}$, and parameter γ reflects the value of $\mathcal{S}$'s input relative to $\mathcal{I}$'s and $\mathcal{E}$'s outside options.

Summary. Optimal Policy

Figure D11 illustrates the optimal set of restrictions on confidentiality and exclusive contracts in the model with a common supplier from Section 3.1. The precise characterization is given in equation (D.10). If the competition between $\mathcal{I}$ and $\mathcal{E}$ is high, the restriction on exclusive dealing can only decrease welfare because $\mathcal{I}$ will not share its knowledge with $\mathcal{S}$. For the intermediate level of competition, the ban on exclusive dealing can increase welfare because it allows $\mathcal{E}$ to get access to a high-quality input from $\mathcal{S}$. However, the planner should leave confidentiality agreements to ensure that $\mathcal{I}$ continues to share its knowledge with $\mathcal{S}$. If the competition is low, the restriction on confidentiality agreements will increase welfare only when the input from $\mathcal{S}$ is sufficiently valuable relative to $\mathcal{I}$'s and $\mathcal{E}$'s outside options (high γ).

Notice that for the social planner who does not know the details of the economy $(\alpha, \gamma, \lambda_0/\lambda)$ it is never optimal to simultaneously ban both the exclusive and confidentiality contracts. If the economy is such that the restrictions on vertical contracts can lead to the first best outcome, the ban on confidentiality agreement will be sufficient. Otherwise, the simultaneous restriction on both types of contracts will lead to the equilibrium with no knowledge sharing.

From a policy perspective, the framework with a common customer who can buy inputs from several suppliers is different from the environment with a common supplier who can sell inputs to multiple downstream firms. In the model from Section D.5, a customer who can buy inputs from several suppliers signs a confidentiality agreement or exclusive dealing with the incumbent supplier in order to create incentives for this supplier to share its private knowledge. They do not use these contracts to increase their total surplus by excluding another supplier. Therefore, the restrictions on vertical contracts can only decrease welfare because firms will stop sharing knowledge with each other.

The results in this section should be interpreted with caution because they are based on the highly stylized framework. The central idea is that contractual arrangements between firms affect the diffusion of knowledge in the economy, and there is a room for policy interventions. The analysis of a more general environment is a fruitful area for future research.[38]

D.5 Two Suppliers and a Common Customer

Suppose an incumbent input supplier ($\mathcal{I}$) and a buyer ($\mathcal{B}$) expect the entry of a new supplier ($\mathcal{E}$). To produce a final product, $\mathcal{B}$ needs three inputs to perform a continuum of measure 1 of tasks. A share $\frac{1-\alpha}{2}$ of tasks requires one unit of an input from $\mathcal{I}$, and a share $\frac{1-\alpha}{2}$ requires an input from $\mathcal{E}$. A share $\alpha \in (0,1)$ of tasks requires an additional third input from either $\mathcal{I}$ or $\mathcal{E}$. $\mathcal{B}$'s gross profits are the following

$$\Pi_{\mathcal{B}} = U_{\mathcal{I}} \cdot \frac{1-\alpha}{2} q_{\mathcal{I}} + U_{\mathcal{E}} \cdot \frac{1-\alpha}{2} q_{\mathcal{E}} + U \cdot \alpha \tilde{q}$$

where $q_{\mathcal{I}}$ and $q_{\mathcal{E}}$ are input qualities of $\mathcal{I}$ and $\mathcal{E}$, respectively; $U_{\mathcal{I}} = 1$ ($U_{\mathcal{E}} = 1$) if $\mathcal{B}$ buys a specialized input for $\frac{1-\alpha}{2}$ tasks from $\mathcal{I}$ ($\mathcal{E}$) and $U_{\mathcal{I}} = 0$ ($U_{\mathcal{E}} = 0$) otherwise; $U = 1$ and $\tilde{q} = q_{\mathcal{I}}$ ($\tilde{q} = q_{\mathcal{E}}$) if $\mathcal{B}$ buys a third input for α tasks from $\mathcal{I}$ ($\mathcal{E}$) and $U = 0$ otherwise.

$\mathcal{I}$ has private knowledge about a valuable technology. If $\mathcal{B}$ learns $\mathcal{I}$'s secrets, it can better accommodate $\mathcal{I}$'s input in its production and the input's quality to $\mathcal{B}$ is $q_{\mathcal{I}} = 1$. Otherwise, $\mathcal{I}$'s input quality is $q_{\mathcal{I}} = 1 - \gamma$ where $\gamma \in [0,1]$. There are also two ways how $\mathcal{B}$ can use $\mathcal{I}$'s secrets to improve $\mathcal{E}$'s input in the final production. First, without sharing $\mathcal{I}$'s secrets with $\mathcal{E}$ it can use them to better accommodate $\mathcal{E}$'s input in the production and improve it from $q_{\mathcal{E}} = (1-\gamma)\lambda_0$ to $q_{\mathcal{E}} = \lambda_1$ where $\lambda_0 \leqslant \lambda_1 \leqslant 1$. Second, $\mathcal{B}$ can also share $\mathcal{I}$'s secrets with $\mathcal{E}$ to get the input quality $q_{\mathcal{E}} = 1$.

The timing and contracting are the same as in Section 3.1. First, $\mathcal{I}$ and $\mathcal{B}$ have a Nash Bargaining with 50/50 split of the surplus over exclusive dealing ($\mathcal{B}$ only buys an input from $\mathcal{I}$) and a confidentiality agreement ($\mathcal{B}$ does not communicate with $\mathcal{E}$ about $\mathcal{I}$'s secrets). Second, $\mathcal{I}$ decides whether to share its secrets with $\mathcal{B}$, and then they have a Nash Bargaining with 50/50 split of the surplus over the input price. Third, $\mathcal{B}$ decides whether to share $\mathcal{I}$'s secrets with $\mathcal{E}$ (if it is not restricted by a confidentiality agreement), and makes a take-it-or-leave-it offer about the input price (if it is not restricted by exclusive dealing).

The equilibrium is summarized in Proposition 3 and in Figure D12. Some predictions are similar to the model with a common supplier. In particular, lower competition (α) and greater importance of knowledge sharing (γ) lead to the equilibrium where all firms learn $\mathcal{I}$'s secrets.

[38] See Whinston (2006, ch. 4) for the comprehensive review of the literature on exclusionary vertical contracts.

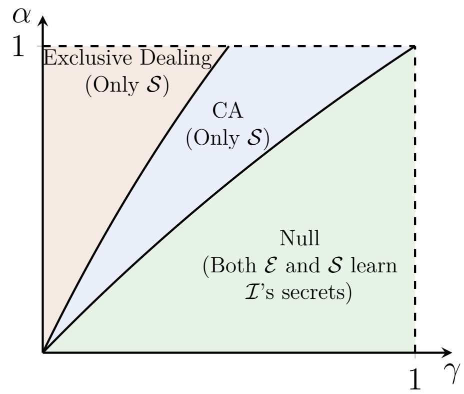


Figure D12: Equilibrium in the Model with Two Suppliers and a Common Customer

This figures illustrates the equilibrium contract between $\mathcal{I}$ and $\mathcal{B}$ described in Proposition 3 in Appendix **??**. Parameter α measures the degree of competition between $\mathcal{I}$ and $\mathcal{E}$, and parameter γ reflects the value of $\mathcal{I}$'s input in $\mathcal{B}$'s production.

However, there are differences between common suppliers and customers in incentives to sign exclusive dealing and confidentiality agreements. The common supplier might want to maximize the difference between product qualities of its customers because they compete with each other. So, it might prefer to sign exclusive dealing to create a commitment that it will not sell inputs to $\mathcal{E}$ and will not share $\mathcal{I}$'s secrets with $\mathcal{E}$. On the other hand, profits a common customer are increasing in the quality of inputs from all its suppliers. So, the common customer ($\mathcal{B}$) does not have incentives to create such a commitment, and it signs exclusive dealing or a confidentiality agreement only to encourage a supplier to share private knowledge with it.

Proposition 3. *The equilibrium contract in the model from Section D.5 is the following*

$$Contract = \begin{cases} \textit{Exclusive with CA} & \textit{if } \alpha > \frac{\gamma}{2(\lambda_1-\lambda_0)+\gamma(2\lambda_0-1)} \\ \textit{CA only} & \textit{if } \frac{\gamma}{2(1-\lambda_0)+\gamma(2\lambda_0-1)} < \alpha \leqslant \frac{\gamma}{2(\lambda_1-\lambda_0)+\gamma(2\lambda_0-1)} \\ \textit{No Contract} & \textit{if } \alpha \leqslant \frac{\gamma}{2(1-\lambda_0)+\gamma(2\lambda_0-1)} \end{cases} \tag{D.12}$$

Firm $\mathcal{I}$ always shares knowledge with $\mathcal{B}$, and $\mathcal{B}$ shares $\mathcal{I}$'s secrets with $\mathcal{E}$ only when $\mathcal{I}$ and $\mathcal{B}$ do not sign any contract in the first stage.

Proof. Joint surplus of $\mathcal{I}$ and $\mathcal{B}$ under different contracts is the following

$$\text{Exclusive, CA:} \qquad TS(ex=1, ca=1) = \frac{1-\alpha}{2} + \alpha \tag{D.13}$$

$$\text{CA Only:} \qquad TS(ex=0, ca=1) = \frac{1-\alpha}{2} + \alpha + \frac{1-\alpha}{2}(1-\gamma)\lambda_1 \tag{D.14}$$

$$\text{No Contract:} \qquad TS(ex=0, ca=0) = \frac{1-\alpha}{2} + \alpha + \frac{1-\alpha}{2} = 1 \tag{D.15}$$

$\mathcal{I}$ and $\mathcal{B}$ sign a contract that maximizes their joint surplus subject to the constraint that $\mathcal{I}$ has incentives to share its secrets with $\mathcal{B}$. If $\mathcal{I}$ shares its secrets with $\mathcal{B}$, then its profit under different contracts will be the following

$$\begin{aligned}
\pi_{\mathcal{I}}(ex=1,ca=1) &= \frac{1}{2}\cdot\left(\frac{1-\alpha}{2}+\alpha\right) = \frac{1+\alpha}{4} \\
\pi_{\mathcal{I}}(ex=0,ca=1) &= \frac{1}{2}\cdot\left(\frac{1-\alpha}{2}+\alpha-\alpha\lambda_1\right) = \frac{1+\alpha}{4}-\frac{\alpha}{2}\lambda_1 \\
\pi_{\mathcal{I}}(ex=0,ca=0) &= \frac{1}{2}\cdot\left(\frac{1-\alpha}{2}+\alpha-\alpha\right) = \frac{1+\alpha}{4}-\frac{\alpha}{2}
\end{aligned}$$

If $\mathcal{I}$ does not share its secrets with $\mathcal{B}$, then its profit will be

$$\pi_{\mathcal{I}}\,(\text{No Sharing}) = \frac{1}{2}\left(\frac{1-\alpha}{2}+\alpha\cdot(1-\lambda_0)\right)\cdot(1-\gamma) = \frac{1+\alpha}{4}(1-\gamma)-\frac{\alpha}{2}\lambda_0(1-\gamma)$$

Under exclusive dealing with a confidentiality agreement, $\mathcal{I}$ always has incentives to share its secrets with $\mathcal{B}$. Under a confidentiality agreement only, it shares its secrets if

$$\frac{1+\alpha}{4}\gamma \geqslant \frac{\alpha}{2}\left(\lambda_1-(1-\gamma)\lambda_0\right) \Leftrightarrow \alpha \leqslant \frac{\gamma}{2(\lambda_1-\lambda_0)+\gamma(2\lambda_0-1)} \tag{D.16}$$

Without any contract in the first stage, $\mathcal{I}$ shares its secrets with $\mathcal{B}$ if

$$\frac{1+\alpha}{4}\gamma \geqslant \frac{\alpha}{2}\left(1-(1-\gamma)\lambda_0\right) \Leftrightarrow \alpha \leqslant \frac{\gamma}{2(1-\lambda_0)+\gamma(2\lambda_0-1)} \tag{D.17}$$

The joint surplus of $\mathcal{I}$ and $\mathcal{B}$ is always higher with less contractual restrictions. Therefore, they sign a contract so that $\mathcal{I}$ has incentives to share its secrets with $\mathcal{B}$. The combination of (D.16) and (D.17) gives the contract in (D.12). □