LUND UNIVERSITY

SCHOOL OF ECONOMICS AND MANAGEMENT

Tracking strategic patents through patent self-citations: An empirical study of the US semiconductor industry

Sten Almgren & Gunnar Hallberg
Bachelor Thesis
Department of Economics
Supervisor: Håkan Jerker Holm
May 2024

Abstract:

This thesis explores the impact of preemptive patenting strategies on technological competition in the context of the US semiconductor industry. Employing publicly available patent data from 49 US based semiconductor firms, strategic patents serving preemptive purposes are identified through self-citations to prior art. Patent-level data is analyzed to determine whether strategic patents deter future entry into the technological domain of the patenting firm. In addition, firm-level financial data is gathered to study potential determinants of strategic patenting. Our findings reveal that patents that are protected by clusters of strategic patents tend to be less cited by rival firms. Yet, we find no economically significant effect on the external citations received by the strategic patents themselves. Size and capital intensity of semiconductor firms are found to be linked to greater strategic patenting propensity, while R&D intensity has little impact. Our results further contribute to a greater understanding of possible ways in which patent data can be employed to track preemptive patenting behavior in industries.

Table of Content

Abstract 1
1. Introduction 2
2. Background 5
2.1 Measuring technological progress and performance through patent data 5
2.2 Strategic patenting 7
2.3 The semiconductor industry 9
3. Theoretical framework 10
Figure 1: Illustration of the citation linkages between protected-, strategic- and external patents 11
3.1 Testable predictions regarding the impact of strategic patenting 11
3.2 Determinants of strategic patenting 13
4. Sample selection and data 14
Table 1: Full sample descriptive statistics 16
5. Empirical analysis 18
5.1 Hypothesis 1 18
Table 2: Two-sample Wilcoxon rank-sum (Mann–Whitney) test 19
Figure 2: Distribution of data Wilcoxon rank-sum test, full sample 19
Figure 3: Distribution of data Spearman's rank correlation test, full sample 20
Table 3. Linear regression output 21
5.2 Hypothesis 2 22
Figure 4: Distribution of data Wilcoxon rank-sum test, full sample 22
Figure 5: Distribution of data Spearman's rank correlation test, full sample 23
5.3 Exploring potential determinants of strategic patenting 24
Figure 5: Distribution of strategic patent counts 1993 - 2002 25
Table 4: Regression output 28
6. Discussion 30
7. Conclusion 33
8. Appendix: 33
Figure 6: External forward citations and forward self-citation counts per patent 34
Figure 7: External forward citations and backward self-citation counts per patent 34
Table 5: Complete list of the semiconductor firms in the sample 34
9. References: 35

1. Introduction

There is growing concern in the industrial policy debate whether the increasing levels of concentration observed in many high-tech industries over the past decades, particularly in the US, have led to weakened competition. On one hand, concentration appears to be a natural and even efficient outcome in industries where technical competition is fierce and up front costs are high, often involving a "winner-takes-all" mechanism and leading to vast productivity differences between firms (see Autor et al., 2020; Haskel and Westlake, 2018). On the other hand, there are fears that firms that have been endowed with substantial market power are exploring various anticompetitive ways to preserve it. Decker et al. (2016) for instance document that business dynamism, measured as the difference in firm growth rates between industry entrants and incumbents, in the US has declined since 2000, driven partly by a decline in high growth startups. Furthermore, Akcigit and Ates (2023) conclude that the decline in business dynamism has primarily been driven by lower levels of knowledge diffusion from industry leaders to laggards and potential entrants. Viewed through the lens of creative destruction, higher levels of concentration in highly innovative industries should come as no surprise as firms actively seek to overtake their competitors by improving upon their technology. Yet, as pointed out by Haskel and Westlake (2018), as soon as firms obtain a technological lead, they have incentives to avoid being replaced themselves and to absorb any knowledge spillovers that could facilitate the innovative activities of competitors. Hence, firms growing large following a series of successful innovative investments are likely to increasingly rely on unproductive investments - such as lobbying, acquiring innovative startups, and preemptive patenting - to protect their current competitive advantage rather than investing into further advancing the technological frontier of their industry. In this study, we take a closer look at one of the strategies leading firms might adopt to preserve their market power: the strategic use of patents to preempt technological competition.

The patent system is designed to create a trade-off between granting short-term monopoly rights to new technologies and incentivizing innovative efforts. By conditioning patent grants on the publication of new ideas, there is an additional purpose of the system to stimulate the diffusion of new knowledge in the economy (Hall and Harfoff, 2012). However, in industries characterized by cumulative innovation, where new inventions largely build on the technologies that came before them, the patent system has been criticized for slowing down the rate of technological progress. When patents are granted too generously, the higher costs of pursuing follow-on innovations risks outweighing the net social benefits of protecting old technologies. In a survey, Cohen et al. (2000) found that in patent intensive industries, including telecommunications, semiconductors and pharmaceuticals, patents were paradoxically among the least emphasized mechanisms for appropriating the profits of one's inventions. Instead, patents were mostly used as a means to either indirectly increase the profitability of some core technologies by blocking rivals from developing related technologies, or as leverage in negotiations with other firms holding IP rights to complementary technologies. The foundational theoretical analysis of market structure and incentives to innovate when there is competition in R&D is laid out in Gilbert and Newbery (1982), stating that a monopolist (for our purposes: a patent holder) will always have greater incentives to develop substitute technologies than a potential entrant in order to preserve his monopoly power. The authors conclude that the real-world complexities of R&D competition limit the feasibility of preemptive strategies to exceptional circumstances.

Any attempt to empirically study the extent and effectiveness of preemptive patenting strategies faces the challenge of distinguishing innovative from preemptive efforts. For this reason, most empirical evidence on the matter has come from large-scale inventor surveys (e.g. Cohen et al., 2002; Giuri et al., 2006; Torrisi et al., 2016). Enabled by several ambitious efforts to standardize and structure extensive amounts of patent-office data over the past decades, we explore an alternative method of determining patenting motives through examination of the information contained in patent publications. Employing patendata from the USPTO, we hypothesize that patent self-citation counts found in patent prior art are indicative of the extent to which a given patent serves a preemptive purpose. In the empirical literature on patenting behavior, self-citations has been used to measure the degree to which a patent serves to deter competition (Akcigit and Ates, 2023), and as a proxy for the ability of firms to internalize the knowledge spillovers of their inventions (Hall, Jaffe and Trajtenberg, 2005). The empirical method presented in this study draws on extensive firm samples, making it appropriate for analyzing aggregate trends in industry patenting behavior. Our approach might thus prove to be a useful way to quantitatively test the qualitative findings in future survey based research on this topic.

The purpose of this thesis is to empirically examine the impact of preemptive patenting strategies on technological competition in the context of the US semiconductor industry. By distinguishing strategic patents serving preemptive purposes through patent self-citations, we first analyze patent-level data to establish whether patents with high shares of self-citations successfully deter future entry into the technological domain of the patenting firm. Next, we employ firm-level data to explore what firm characteristics are predictive of higher strategic patenting rates. For our theoretical framework, we adopt the early theoretical predictions in the literature (Gilbert and Newbery, 1982; Gilbert, 1987) in combination with more recent contributions by Ackigit et al. (2013) and Argente at el. (2020). Our data sample is made up of 49 US based publicly traded semiconductor firms. The semiconductor industry is known for exhibiting rapid technological progress and fierce technological competition, with many firms having a proven record of strategically using the patent system to their advantage. Thus, it is a suitable candidate to study our topic of interest. By assigning patent data from 1980 to 2006 to our firm sample and combining it with financial data, our final dataset covers 21,684 patents granted between 1998 and 2002.

The remainder of the study is organized as follows: *Section 2* contains a literature review of previous related academic studies. In *section 3*, the theoretical framework for our study is laid out. In *section 4*, the data sources we have used are presented and the final dataset is described. *Section 5* contains statistical tests of the predictions from *section 3*, the results of which are discussed in *section 6*. Finally, in *section 7* the main findings of the empirical analysis are summarized and areas in need of further academic research are identified.

2. Background

2.1 Measuring technological progress and performance through patent data

It is widely recognized that technological progress plays a crucial role in the economy as the main driver of long run productivity growth. Accordingly, substantial research efforts have been dedicated to identify the key mechanisms behind the innovative process. Empirical research on the topic has however been constrained by a lack of proper measures of innovation output and reliable ways of tracking the knowledge spillovers commonly assumed to arise as the byproduct of a successful invention. In light of this, patent data has been extensively employed by economists through the years as it remains one of few ways in which it is possible to empirically study the relationship between innovation and various economic conditions. A characteristic feature of patent data is that it contains detailed information about the scope of the technology being patented and about any adjacent technologies that it claims to be different from. To be eligible for patent protection, an invention must be novel, non-trivial and commercially viable. Due to legal requirements, any patented technology is obliged to publicly disclose two potentially valuable pieces of information which are detailed in the patent document created upon the granting of the patent. First, it must distinguish itself from any existing technologies that it relates to or claims to improve upon. This is done through citations to *prior art* (henceforth, *backward citations*) and represents the boundaries of the patentable claims made. Second, the inventor must specifically assert the technological novelty of his invention which is done through a number of claims, representing the scope of the intellectual property right (Hall, Jaffe and Trajtenberg, 2001).

There is of course large variation in technological impact and importance of patents, implying that mere patent counts in isolation might be an inadequate measure of inventive output. Recognizing this, Trajtenberg (1990) was one of the first studies to combine patent counts with patent citation data to measure the innovative performance of firms. In a small sample of computed tomography patents, it was documented that patent counts weighted by *forward citations* (i.e. the number of citations received from subsequent patents) was strongly associated with the economic value of the patented inventions. Using firm-level data, Hall, Jaffe and Trajtenberg (2005) study the relationship between R&D spending (innovative input) and citation weighted patent counts (innovative output) and find that a higher number of forward citations corresponds to higher stock market valuations of the accumulated R&D stocks of firms, representing their stock of knowledge. They further find that self-citations (citations received from subsequent patents that belong to the same firm) are valued the highest, suggesting that self-citations reflect successful appropriation of the profits- and internalization of the spillovers generated by a firm's innovative efforts. The interpretation of patent citations as an indicator of patent value and technical significance fits conveniently into the view of innovation as a cumulative process of creative destruction. By citing another patent, a firm declares that the cited innovation either constitutes a crucial complementary input to its own innovation, or alternatively, that it represents the previous technological frontier in an area that the citing firm now improves upon. Thus, highly cited patents can to some extent be interpreted as more groundbreaking discoveries that have paved the way for new lines of applied research. Correspondingly, patent citation data has been used by researchers to track the knowledge spillovers of inventions (Jaffe, Trajtenberg and Henderson, 1993). Moreover, assuming that firms actively allocate their R&D efforts to technological areas where the expected profits are highest, forward citations should be indicative of the economic value of inventions.

Although patent citations to prior art have been favored by many empirical researchers, they do exhibit certain limitations. During the patent application process, both patent attorneys and patent examiners are free to add additional citations to ensure the patentability of the invention. Evidence indicates that non-patent-applicant citations can introduce significant noise to the data. Alcácer et al. (2009) find that examiners on average account for a significant amount of total citations and that the degree of examiner interference particularly relates to certain patent characteristics. Their data suggests that high shares of examiner citations are skewed towards patents with low numbers of applicant citations stemming from industries where IP ownership is fragmented, where patents more frequently serve strategic purposes and where lower private values are typically placed on any single patent. Additionally, the share of examiner citations tended to be greater in foreign patent applications, reflecting national differences in the patent application process. Moser et al. (2018) find that patent attorneys on average exhibit bias towards citing a small group of early patents within a technological field, but conclude that attorneys generally contribute to distorting citation data far less than patent examiners. Further, Hall, Jaffe and Trajtenberg (2005) point out that self-citations might reflect applicant self-bias due to informational differences determining how well firms are able to relate inventions to other internal- versus external technologies. Several studies have explored the usefulness of alternative sources of patent information in estimating patent value. For instance, the number of claims, reflecting the scope of patented technologies, has been found to be a strong predictor of patent quality (Lanjouw and Schankerman, 2004). Von Graevenitz et al. (2013) use the number of citations to non-patent literature within technological fields (e.g. scientific publications) to estimate technological opportunity, and Harhoff et al. (2003) find that such references are predictive of patent quality in highly science based fields. Citation data has also been employed to measure the range of the technological foundation- and impact of patented inventions. Jaffe, Trajtenberg and Henderson (1997), construct two concentration indexes of the distribution of citations to different technological fields: one for forward citations (*"generality"*) and one for backward citations (*"originality"*).

2.2 Strategic patenting

The patent system seeks to stimulate R&D investment by promising exclusionary rights to whomever is first to produce an invention. However, where inventions are sequential, strong patent protection could impede the rate of technological progress by raising the costs of pursuing follow-on innovation (Schotchmer, 1991). In such circumstances, patenting in itself can prove to be a useful strategic tool for firms as it affords them an opportunity to restrict the technological access of rivals. Thus, whenever strategic patenting practices are present, there is good reason to question the accuracy of methods relying on patent data to measure innovative output. In a survey of European, American and Japanese patent applicants, Torrisi et al. (2016) found that 40% of patents were not commercially used, out of which 67% were filed solely to block other patents. The term *patent thicket* is commonly used to describe contexts where patent rights overlap and IP ownership is fragmented, implying that firms require access to large sets of patented external technologies in production. Patent thickets often arise in industries where innovation is cumulative and firms develop complementary technologies (Shapiro, 2001). In a study of the semiconductor industry, Ziedonis (2004) finds that firms aggressively expand their patent portfolios in response to patent thickets to avoid being held up by the owners of patented complementary technologies. She argues that fragmentation of patent rights can increase the transaction costs of engaging in increasingly complex licensing agreements, driving firms to instead use patents as bargaining chips in negotiations with competitors. Noel and Schankerman (2013) present evidence from the software industry that patent thickets further negatively impact the market value of patents and R&D investments.

The literature highlights different motives for strategic patenting when firms compete for substitute technologies. Gilbert and Newbery (1982) show that in a basic setting where an incumbent with monopoly power in the product market, derived from proprietary rights to a superior technology, and a potential entrant compete to develop a single substitute technology, the incumbent will always be willing to invest more than the entrant to ensure that his monopoly power is preserved. Note that the incumbent’s higher incentives to invest does not depend on the characteristics of the technology in question. In fact, the analysis stays intact even when the two compete for ownership of an inferior technology, where the incumbent would simply avoid putting the technology to use if he acquires it. Rather, the incumbent’s incentive to invest in R&D is determined by the difference in monopoly profits and the incumbent's share of duopoly profits if a competitor enters the product market (i.e. the opportunity cost of losing the patent race). Since monopoly profits are always weakly greater than aggregate duopoly profits, the incumbent will place a higher value on the patent than the entrant. In practice however, preemptive patenting strategies are limited by the complexity and uncertainty involved in the R&D process. The authors stress that the likelihood for successful preemption decreases in the degree to which it is possible for entrants to invent around the incumbent's patent. Although successful preemption implies sustained concentration in the product market, as Gilbert (1987) points out, if the incumbent fails to keep up higher R&D intensity than the aspiring entrants, this is unlikely to occur. An important exemption exists when the R&D process exhibits increasing returns. If the incumbent’s R&D productivity increases in response to higher rates of investment, then over time he will be able to lower his R&D intensity without attracting opportunistic entrants.

Akcigit et al. (2013) argue that many attempts to measure patent values through forward citations neglect the possibility that the owners of the most valuable patents have incentives to limit external access to their technologies. Using data from non-practicing entities, an inverted-U relationship between forward citations and private patent values was identified, suggesting that higher value patents are more often subject to preemptive patenting strategies. Akcigit and Ates (2023) further propose that the number of backward self-citations might be a useful way to distinguish preemptive patents, reflecting an attempt to build a thicket around core technologies. Argente et al. (2020) suggest that incentives to pursue preemptive strategies increase in firm size as larger firms generally own more valuable stocks of intangible assets and are less eager to develop new technologies that risk cannibalizing on the profits of their old technologies. Galasso and Schankerman (2015) correspondingly find that in instances where patents of large firms have been invalidated, a substantial increase in citations from smaller firms to the focal patent generally followed. Survey evidence in Guiri et al. (2006) documents that 40% of patents belonging to large firms were left unused, compared to 18% for smaller firms. The authors point out that larger firms might benefit from scale economies in the patent application process, leading them to more frequently patent minor inventions yielding comparatively lower returns.

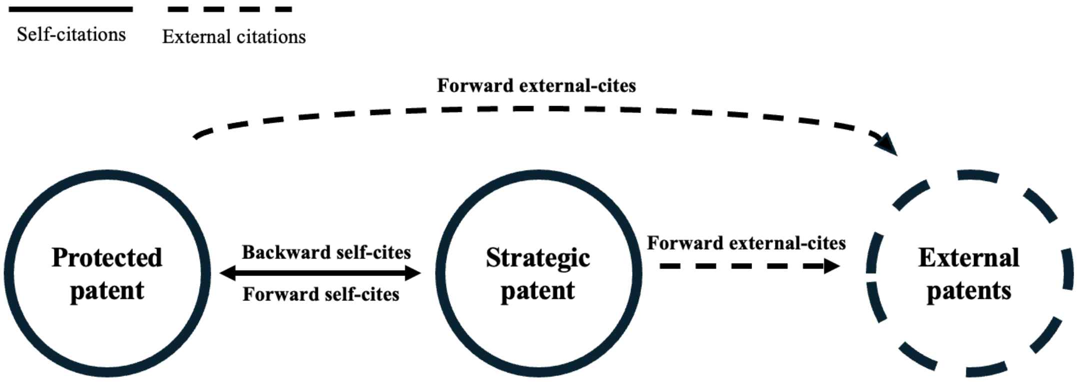
2.3 The semiconductor industry

Due to high-levels of technical specialization, intense R&D competition, and a cumulative process of technological progress, many researchers have studied the behavior of semiconductor firms to empirically test various hypotheses on innovation. Semiconductor firms are widely recognized to be among the most patent intensive (Cohen et al. 2000; Bessen and Maskin, 2009) and tend to have exceptionally high patent citation rates (Hall, Jaffe and Trajtenberg, 2005). Previous studies focusing on patenting in response to hold-up risks in the semiconductor industry (Hall and Ziedonis, 2001; Ziedonis, 2004) have found that patents play an integral role in firm strategy. Texas Instruments for instance is well known for having been one of the most prolific exploiters of the patent system historically (Pepall et al., 2013). The results in Hall and Ziedonis (2001) also allude to important differences in patenting behavior between semiconductor design firms and manufacturing firms. The authors show that the strengthening of patent rights during the 1980s facilitated industry wide specialization and the vertical disintegration between design- and manufacturing firms. Thus, stricter patent protection seems to have been a necessary condition for design firms, relying mainly on stocks of intangible assets, to be able to operate as independent commercial entities. The semiconductor industry exhibits certain features that, consistent with Gilbert (1987), makes it susceptible to preemptive patenting practices: 1) due to the sequential nature of innovation, patents can be used to fence in significant markets, 2) R&D expenditure is an important component of firm strategy, and 3) the range of patented technologies appears to be narrowly contained within the industry. To the best of our knowledge, no empirical study has yet been conducted to explore the presence and extent of preemptive patenting in the semiconductor industry through the use of publicly available patent data.

3. Theoretical framework

Following Akcigit et al. (2013) we adopt the view that firms are guided by two conflicting interests when deciding on how to allocate their R&D resources. On one hand, firms are interested in developing *productive* inventions of the highest possible economic value (both product- and process innovations). Such efforts contribute to advancing the technological frontier, generating knowledge spillovers and profitable opportunities to develop additional technical applications that can be exploited by competitors. Firms that have managed to produce and patent a valuable invention might therefore opt to allocate resources towards protecting those technologies to avoid an outcome where shares of their monopoly profits fall into the hands of rival firms. Such efforts are deemed *preemptive* and do not lead to any technological advancement. Attempts to preemptively patent substitute technologies before potential entrants are however limited by the possibility that there exists several ways to invent around the patent in question. In that case, preemption can be a costly endeavor, especially if it is unlikely the firm itself is able to identify all potential applications of its invention. Gilbert and Newbery (1982) argue that an alternative strategy to deter entry is to explore ways of making it more costly for rivals to develop substitute technologies. If the expected profits to potential entrants are reduced, inventors considering entering the patent race might suddenly find it more worthwhile to redirect their R&D resources elsewhere. Akcigit et al. (2013) recognize that firms can deliberately create dense thickets of related and overlapping *strategic* patents to expand the scope of patented core technologies. As the claims of the focal patent are effectively extended in many new directions, the R&D costs of entrants will presumably rise in proportion to the size of the thicket as they have to put greater efforts into distinguishing their technology from prior art. Additionally, the higher risk of incurring litigation costs from inadvertently infringing on adjacent patents will likely discourage rivals - especially small firms with less financial strength - from developing technologies in the domain of the incumbent firm. Assuming that the strategic patents constituting the thicket are intimately linked, one should expect self-citations to make up a substantial amount of their prior art. Likewise, shares of self-citations are expected to be higher in denser thickets. For this reason, we predict that highly self-citing patents are more likely to serve preemptive purposes.

Figure 1: Illustration of the citation linkages between protected-, strategic- and external patents



3.1 Testable predictions regarding the impact of strategic patenting

We argue for two primary reasons why strategic patents themselves will tend to receive fewer external citations. The first mechanism relates to the nature of the technology underlying a strategic patent. A primary focus when pursuing preemptive patenting should be that each individual strategic patent is cost effective to produce. Thus, it is reasonable to assume that strategic patents will represent incremental improvements on existing technologies that require little R&D effort to develop. Accordingly, even absent a wider thicket, external inventors will find little reason to cite such patents. Second, since successful execution of strategic patenting will deter entry, fewer substitute technologies will as a consequence be developed by rivals. Assuming that the patents of potential entrants would have to cite several of the strategically clustered patents, lower entry rates (i.e. fewer external follow-on inventions) should lead to fewer external citations received by the strategic patents. We predict that the magnitude of both these effects grow in the share of patent self-citations, indicating denser thickets. Hence, we are able to formulate our first testable hypothesis:

- **Hypothesis 1:** Patents with high shares of backward self-citations, representing strategic patents, will tend to receive fewer external forward citations.

Our first hypothesis naturally extends to the impact of strategic patenting on the citations received by the protected patent. Because the very purpose of preemption is to deter follow-on inventions from appropriating the monopoly profits of specific technologies, one should expect the presence of strategic patents to contribute to fewer external citations received by the protected patent as well. Akcigit et al. (2013) illustrate this effect by showing that the number of forward citations and the private value of patents follow an inverted-U relationship. The authors argue that at some level of patent value, IP owners will shift their focus towards appropriating the spillovers of their invention and maximize its private value by preempting rival entry. This creates a negative relationship between forward citations and the degree to which a patent is protected. Since protected patents are assumed to receive the highest numbers of forward self-citations, indicating the presence of a preemptive thicket, we can state our next hypothesis as follows:

- **Hypothesis 2:** Patents with high shares of forward self-citations, representing protected patents, will tend to receive fewer external forward citations.

3.2 Determinants of strategic patenting

In addition to hypothesis 1 and 2, we are interested in exploring what firm characteristics might hold general predictive power of the propensity to produce strategic patents. Studying the effects of strategic patenting on product market competition, Argente et al. (2020) demonstrate that incentives to produce productive- versus strategic patents vary across the firm size distribution. The authors emphasize two reasons why larger firms achieve higher returns from strategic patenting than smaller firms. First, in technologically competitive industries, firm success in the product market is often attributable to ownership of certain key technologies. Consistent with the logic in Gilbert and Newbery (1982), incentives to preemptively patent depend on the size of the product market that the patented technology is able to fence in. Thus, where product market power is attained through ownership of specific technologies, larger firms generally have larger patent values to protect. Second, Argente et al. (2020) point out that incentives to develop productive technologies decrease in firm size due to higher risks of introducing inventions that cannibalize on the profits of existing inventions. In the literature, this effect is commonly referred to as the replacement effect: when incumbents introduce new technologies, the new monopoly profits gained will to some extent replace the monopoly profits of their old technologies (Pepall et al., 2013). Argente et al. (2020) further present empirical evidence that patents of large firms are less likely to result in product introductions. In their sample, these patents were also less novel, received fewer forward citations and had higher shares of self-citations. All this suggests that smaller firms will tend to primarily allocate their R&D resources towards developing productive patents, while larger firms will tend to pursue strategic patenting more aggressively.

Previous studies of the semiconductor industry have found that R&D- and capital intensity are important determinants of patenting (Hall and Ziedonis, 2001; Ziedonis, 2004). R&D investments are commonly assumed to be associated with inventive input, suggesting that it should be strongly correlated with the number of productive patents produced by firms. The impact of R&D intensity on strategic patenting is however less clear. If strategic patents represent technically insignificant inventions, then the R&D investments of firms will have little impact on the number of strategic patents produced. If, however, successful preemption requires that the incumbent firm has to sustain higher levels of R&D investment than potential entrants, as suggested by Gilbert (1987), then R&D intensity will likely be a significant determinant of strategic patenting. Evidence that capital intensive firms, measured by the book value of property, plant and equipment per employee, tend to patent more frequently has been linked to higher costs of being held-up due to large technology specific sunk investments. Semiconductor manufacturing firms in particular make investments in production capacity that often require access to patented technologies of other firms. Expanding the patent portfolio is one way to mitigate the risk of ending up in disputes with the owners of complementary patents (Hall and Ziedonis, 2001). Yet, where sunk investments in machinery are tied to technologies patented by the firm itself, it is also plausible that capital intensive firms will be more prone to pursue strategic patenting compared to other firms. Specifically, experiencing competition from patented substitute technologies of rival firms could possibly be more costly for incumbents with large technology specific sunk capital investments. Thus, there is reason to suspect that capital intensity might be correlated with strategic patenting as well.

In the following sections, we attempt to test our theoretical predictions against patent data from the semiconductor industry. To distinguish strategic patents, we construct a measure of shares of self-citations per patent. Before moving on to the empirical analysis, there are however some limitations to our approach that are worth highlighting. First of all, clusters of patents with high self-citation shares are not necessarily always the result of anticompetitive conduct. These clusters could also indicate successful appropriation of the spillovers of inventions, stemming from superior innovative capabilities of some firms. Because the effects on technical competition in both cases are similar, it is difficult to distinguish one type of conduct from the other in the data. Next, regarding our second prediction, although strategic patent clusters are expected to negatively impact the number of future citations from substitute technologies, the same cannot be said about future citations from complementary technologies. In particular, a dominant position in one part of the semiconductor value chain, e.g. within chip design, could possibly lead to more manufacturing firms adapting their technologies to accommodate that specific design in production. Thus, there could be two opposing effects present regarding the impact of preemption on external forward citations, rendering the ultimate effect less clear cut.

4. Sample selection and data

Similar to Hall and Ziedonis (2001) and Ziedonis (2004), we define our relevant market as US-based, publicly traded firms whose principal line of business is *semiconductors and related devices* as indicated by the assigned four digit SIC code 3674. These selection criteria are necessary to ensure that the financial data we intend to use later on (revenue, R&D expenditure, PP&E, etc.) are mostly tied to semiconductor related areas, and that patents are primarily applied for within the US legislative system. This however means that several relevant diversified firms with semiconductor branches and international semiconductor firms are excluded from our sample. Using CAPITAL IQ, we begin by screening for current publicly traded US based firms with SIC code 3674 that were listed during the period 1998-2002. We then conduct wide ranging internet searches to identify additional firms that were publicly traded 1998-2002 but that have been acquired, gone private or gone out of business in later years. Although we cannot claim to have gathered an exhaustive list of the entire universe of relevant firms, comparing our firm sample size to previous studies (Hall and Ziedonis, 2001; Ziedonis, 2004), it appears that we are not too far off.

The NBER Patent Data Project (Hall, Jaffe and Trajtenberg, 2001) offers one of the most comprehensive and easily accessible collections of patent publication information of US companies. The original dataset which contained around 3 million utility patents gathered from the USPTO stretching from 1975 to 1999 has been updated several times and now extends through 2006. One of the main benefits of the NBER patent database has been to address the challenge of assigning patents to firms on a large scale. Firms patent under a variety of names, both their own and through subsidiaries, and the USPTO does not have a consistent identifier for patenting entities. The fact that firms frequently undergo changes of names and ownership further complicates the task of assigning consolidated patent stocks to firms that are accurate over time. The DISCERN database (Arora, Belzon and Sheer, 2021) reconstructs the NBER datafiles by correcting firm-patent matches which were originally either incorrectly matched or omitted, accounting for about 20% of the NBER dataset. The DISCERN files assign patents from 1980 to 2015 to US headquartered firms which enter the dataset once they are publicly traded and have at least one patent granted by the USPTO. Patents are assigned to the *ultimate owners*, accounting for ownership structures, and names are standardized to keep track of name changes. Using the DISCERN database, we proceed by assigning patent stocks, including all patents granted from 1980-2006, to our list of semiconductor firms. The resulting dataset is then matched to the NBER data files through the unique patent number assigned to each patent by the USPTO. The following patent specific measures are obtained: 1) number of backward- and forward citations (NBER), 2) originality- and generality indices (NBER), 3) number of claims (NBER), 4) number of citations to non-patent-literature (DISCERN), 5) patent application year and grant year (DISCERN).[1]

The NBER database further contains the patent numbers of forward citing patents. Building on the NBER database, the Harvard Dataverse files (Bhaven, 2011) contains a corresponding dataset for backward citations. We match these datasets to the patent stocks of each firm to calculate the number of forward- and backward self-citations of each patent in our sample. To account for the truncation effect that arises because many patents will continue to receive forward citations long after the final year of the dataset (2006), we decide to delimit our sample to patents granted in the period 1998-2002. We restrict our sample to firms that have at least one patent with a backward self-citation during this time period. Further, all patents in our sample are required to have complete patent information according to the variables 1-5 above. The final step is to match relevant financial data to the sample firms. The following data is collected through FACTSET and CAPITAL IQ: 1) annual net revenue, 2) annual R&D expenditure, 3) net book asset value of property, plant and equipment, 4) average market capitalization per year, 5) book value of equity, 6) number of full time employees per year.

[1] The following files were used to create our final patent dataset:

- NBER Patent Data Project (https://sites.google.com/site/patentdataproject/Home?authuser=0): *cite76_06*, *ori_gen_76_06*, *pay_76_06*.
- DISCERN (https://zenodo.org/records/4320782): *DISCERN_Panel_Data_1980_2015, corp_NPL_cite_per_year_firm_80_15*.
- Harvard Dataverse (https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/16412): *ucites.*

All monetary values are transformed to 2010 USD using a US consumer price index[2] to ensure that the observations are comparable over time.

Our final dataset contains 21,684 patents granted 1998-2002 that are assigned to 49 US based semiconductor firms. A full list of our firm sample can be found in the appendix. *Table 1* provides descriptive statistics of our data.

Table 1: Full sample descriptive statistics

Variable	Mean	Median	SE	Min	Max
Total forward citations	12	7	15	2	235
Total backward citations	13	9	16	2	316
Total external forward citations	10	6	13	0	235
Total external backward citations	11	8	15	0	316
Total forward self-citations	2	0	6	0	104
Total backward self-citations	2	1	4	0	119
Patents granted per company annually	83	13	184	1	726
Patent stock per firm [1)]	443	65	1,037	5	5,013

1) Only includes patents granted between 1998 and 2002.

When measuring backward self-citation shares to identify strategic patents, one apparent issue with the full sample is the presence of patents making few backward citations to other patents overall. The share of backward self-citations becomes particularly sensitive in cases where the total number of backward citations is low. For example, a patent making 2 total backward citations, out of which one is a self-citation will have a 50% share of backward self-citations. Because we argue that successful preemption requires thickets of patents protecting some core technology, it does not seem reasonable to count such patents as "strategic". An additional reason why these observations might introduce noise to the data is related to higher risks of patent-examiner interference. As pointed out by Alcácer et al. (2009), patent examiner citations, which are less clear how to interpret, tend to be most prevalent in patents with few to none applicant citations. According to our definition of strategic patents, we argue it is necessary to account for patents with few overall backward citations as they are less likely to relate to the phenomenon we are interested in exploring. It

[2] Collected from the World Bank DataBank. Link: https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG?locations=US.

is however necessary not to delimit the sample too strictly to avoid excluding an excessive amount of "non-strategic patents". Thus, we create a backward citation adjusted sample where patents making less than 5 backward citations in total are excluded. As a result, the sample shrinks from 21,683 observations to 17,067. The mean and median of external forward citations is unchanged compared to the non-adjusted sample. The standard deviation increases from 13 to 14 which is reasonable given the smaller sample. The mean, median and standard deviation of backward self citations are all unchanged to the non-adjusted sample. In *section 5.1*, tests for hypothesis 1 are run both using the full sample and the backward citation adjusted sample.

Similarly, when measuring forward self-citation shares to identify protected patents, we want to account for the fact that patents receiving fewer forward citations overall might mechanically obtain higher self-citation shares. Because we wish to interpret high forward self-citation shares as indicative of a preemptive effort requiring a thicket of strategic patents to be successful (see *section 3*), it is necessary to impose a lower bound on the number of forward citations per patent in our sample. However, we want to avoid excluding an excessive amount of the patents we deem "non-protected" in the process. Therefore, a forward citation adjusted sample is created where patents receiving fewer than 5 total forward citations are excluded. As a result, the sample shrinks from 21,683 to 15,198 observations. The mean of external forward citations increases to 13 compared to 10 in the non-adjusted sample. The median increases from 6 to 9 and the standard deviation increases from 13 to 15. The forward self-citation mean increases from 2 in the non-adjusted sample to 3 in the adjusted sample. The median increases from 0 to 1 and the standard deviation increases from 6 to 7. In *section 5.2*, tests for hypothesis 2 are run both using the full sample and the forward citation adjusted sample.

5. Empirical analysis

5.1 Hypothesis 1

For hypothesis 1, we are interested in determining whether strategic patents generally receive fewer external forward citations than non-strategic patents. The sample mean of external forward citations per patent is 10 and the sample median is 5, indicating a skewed citation distribution towards zero. Because forward citation counts are possibly not normally distributed, we begin by performing a non-parametric Wilcoxon rank-sum test to study if there is a difference in the number of external forward citations received by strategic and non-strategic patents. To do so, we define strategic patents as patents with a share of backward self-citations larger than or equal to 30%, accounting for 17% of all patents in the full sample. This can be compared to the estimate in Torrisi et al. (2016) that about 25% of patents are used to block other patents. The mean share of backward self-citations in our sub-sample of strategic patents is 0.50 with a median of 0.45. Non-strategic patents are defined as patents with a share of backward self-citations below 30%. The mean share of backward self-citations in our sub-sample of non-strategic patents is 0.06 with a median of 0.

In *table 2*, the full sample Wilcoxon rank-sum test shows that there is a significant but small difference in the number of external forward citations received by strategic and non-strategic patents (test 1). There is a 53.4% probability that a non-strategic receives a higher number of external forward citations than a strategic patent. Test 2 in *table 2* displays the backward citation adjusted sample Wilcoxon rank-sum test results. After removing patents with fewer than 5 total backward citations from the sample, the probability that a non-strategic patent receives a higher number of external forward citations remains significant and increases to 53.6%.

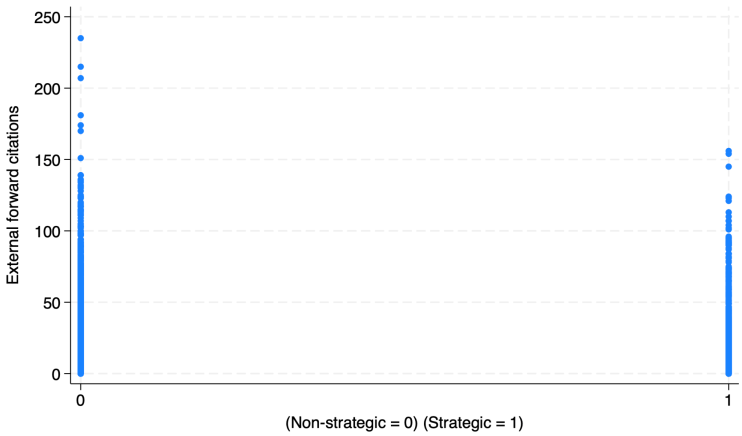
Table 2: Two-sample Wilcoxon rank-sum (Mann–Whitney) test

H0: forward citations(non-strategic) = forward citations(strategic)

	Test 1	**Test 2**	**Test 3**	**Test 4**
Observations:	21,683	17,067	21,683	15,198
Strategic patents	3,601	2,646	5,508	3,570
Non-strategic patents	18,082	14,421	16,175	11,628
Prob > |z| :	(0,000) ***	(0,000) ***	(0,000) ***	(0,000) ***
Prob: forward citations(non-strategic) > forward citations(strategic):	0.534	0.536	0.791	0.830

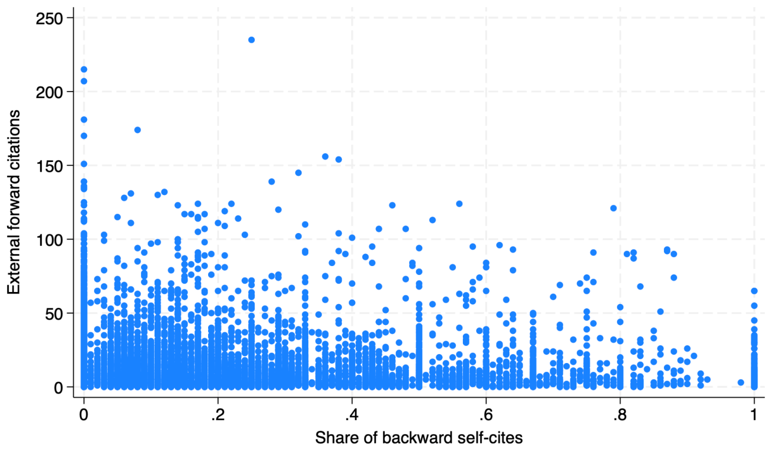
Table description: Test 1 shows the results of the full sample test for hypothesis 1. Test 2 shows the results of the backward citation adjusted sample results for hypothesis 1. Test 3 shows the results of the full sample test for hypothesis 2. Test 4 shows the results of the forward citation adjusted sample results for hypothesis 2.

Figure 2: Distribution of data Wilcoxon rank-sum test, full sample



Next, we examine the correlation between higher shares of backward self-citations and the number of external forward citations received per patent using the non-paramteric Spearman's rank correlation test. For the full sample, a small but significant negative correlation between higher shares of backward self-citations and external forward citations of -0.0532 is identified. This means that patents with higher shares of backward self-citations are expected to receive slightly fewer future citations from the patents of other firms. Performing the same test on the backward citation adjusted sample, we obtain a stronger significant negative correlation of -0.0686, consistent with the results of the Wilcoxon rank-sum test.

Figure 3: Distribution of data Spearman's rank correlation test, full sample



As a final step, we run a linear regression to obtain an estimate of the marginal effect of higher shares of backward self-citations, indicating denser thickets, on the expected number of external forward citations received by a patent. Our dependent variable is the number of external forward citations received in the time period between a patent's grant date and 2006. Our explanatory variable is the share of backward self-citations per patent. Using patent specific data from the NBER patent database, we include three control variables that previous studies have found to predict the technological impact of an invention:

- The generality index: indicative of the range of subsequent applications of the patent measured as the distribution of forward citations stemming from different technical fields. Values close or equal to 1 indicate high generality and are associated with general purpose technologies (Jaffe, Trajtenberg and Henderson, 1997).
- Number of claims: indicative of the scope of a patent.
- Number of non-patent literature (NPL) backward citations: indicative of technological opportunity.

We considered including the originality index as well but dropped it due to the risk of multicollinearity given a correlation with the generality index of 0.56. We specify the linear regression equation as follows:

$$y_i = \beta_1 + \beta_2 x_{share-bself-cites} + \beta_3 x_{generality} + \beta_4 x_{claims} + \beta_5 x_{NPL-cites}$$

Table 3. Linear regression output

	Model 1	Model 2	Model 3	Model 4
Constant	8.232	8.361	10.267	14.067
	(0.000) ***	(0.000) ***	(0.000) ***	(0.000) ***
Share self-cites	-0.122	-0.040	-12.942	-18.158
	(0.793)	(0.946)	(0.000) ***	(0.000) ***
Generality	0.534	0.978	0.345	1.222
	(0.053) *	(0.006) ***	(0.193)	(0.000) ***
#Claims	0.052	0.050	0.072	0.064
	(0.000) ***	(0.000) ***	(0.000) ***	(0.000) ***
#NPL cites	0.146	0.138	0.167	0.177
	(0.000) ***	(0.000) ***	(0.000) ***	(0.000) ***
R-squared	0.012	0.012	0.081	0.109
Adj. R-squared	0.012	0.012	0.081	0.109

Table description: The dependent variable is the number of external forward citations received per patent. Model 1 shows the results of the full sample regression for hypothesis 1. Model 2 shows the results of the adjusted sample results for hypothesis 1. Model 3 shows the results of the full sample regression for hypothesis 2. Model 4 shows the results of the adjusted sample results for hypothesis 2.

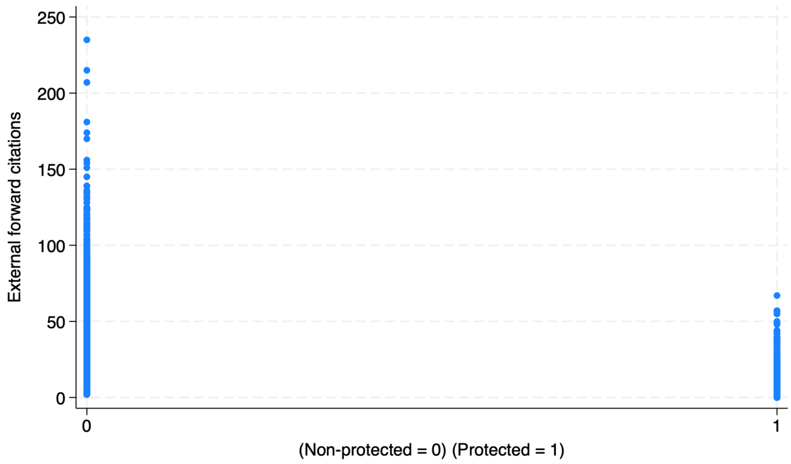
With p-values of 0.793 in model 1 and 0.946 in model 2, the regressions show insignificant coefficients for the impact of higher shares of backward self-citations on the dependent variable. Even though the coefficients indicate negative correlation, this is not possible to statistically prove with the linear regression.

5.2 Hypothesis 2

For hypothesis 2, we wish to determine whether patents that are protected by thickets of strategic patents tend to receive fewer external citations than other, non-protected patents. Similar to *section 5.1*, we begin by performing non-parametric tests to examine the relationship between the share of *forward* self-citations per patent and the number of external forward citations received. We define *protected patents* as patents with shares of forward self-citations above or equal to 30%, accounting for about 25% of all patents in the full sample. The mean share of forward self-citations in our sub-sample of protected patents is

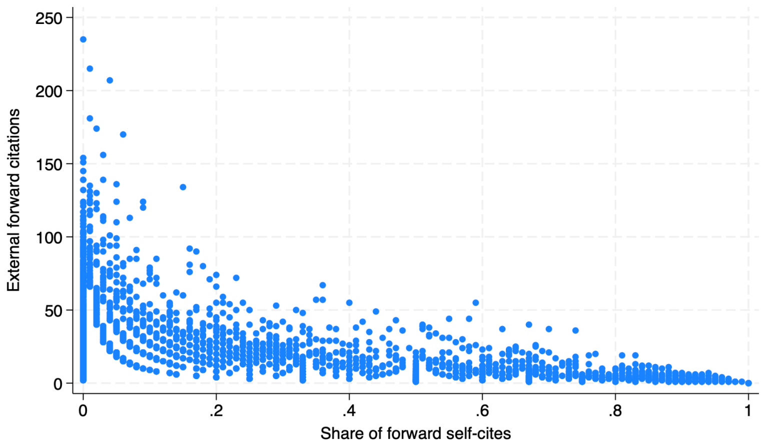
0.60 with a median of 0.50. Non-protected patents are defined as patents with a share of forward self-citations below 30%. The mean share of forward self-citations in our sub-sample of non-protected patents is 0.05 with a median of 0. In *table 2*, the full sample Wilcoxon rank test shows that there is a significant difference in the number of external forward citations received by protected versus non-protected patents (test 3). There is a 79.1% probability that a non-protected patent has a higher number of external forward citations than a protected patent. Test 4 in *table 2* displays the forward citation adjusted sample Wilcoxon rank-sum test results. The exclusion of patents with fewer than 5 total forward citations results in a stronger significant test result as the probability that a non-protected patent receives more external forward citations than a protected patent increases to 83%.

Figure 4: Distribution of data Wilcoxon rank-sum test, full sample



A Spearman's rank correlation test is performed to examine the correlation between higher shares of forward self-citations and the number of external forward self-citations received per patent. For the full sample, a significant negative correlation of -0.2809 is determined. For the forward citation adjusted sample, the significant negative correlation increases to -0.415, consistent with the results from the Wilcoxon rank-sum tests.

Figure 5: Distribution of data Spearman's rank correlation test, full sample



To obtain estimates of the marginal effect of higher shares of forward self-citations, reflecting more protected patents, on the expected number of external forward citations received by a patent, we run a linear regression. Our dependent variable is the number of external forward citations received in the time period between a patent's grant date and 2006. Our explanatory variable is the share of forward self-citations per patent. Generality indices, number of claims, and number of NPL backward citations are included as control variables. We specify the linear regression equation as follows:

$$y_i = \beta_1 + \beta_2 x_{share-fself-cites} + \beta_3 x_{generality} + \beta_4 x_{claims} + \beta_5 x_{NPL-cites}$$

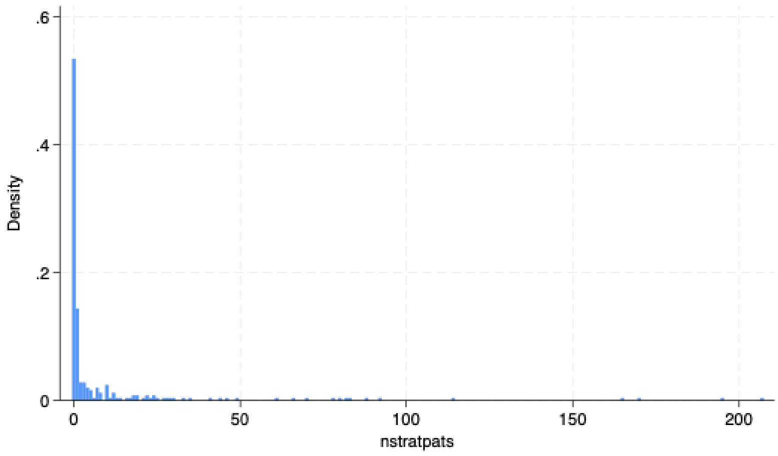
In *table 4*, the full sample results (model 3) show that higher shares of forward self-citations per patent has a statistically significant negative effect on the number of external forward citations received. The estimated coefficient is -12.9 with a p-value of 0.000. This means that for every 10% increase of the share of forward self-citations, the expected number of forward citations received from patents of other firms decreases with 1.29. This effect is even stronger for the adjusted sample in model 4 where the coefficient is -18.2 with a p-value of 0.000. This means that for every 10% increase of forward self citations relative to total forward citations, the expected number of forward citations from patents of other firms decreases with 1.82.

Collectively, our results indicate that protected patents do receive fewer citations from patents of other firms, suggesting that semiconductor firms generally succeed in protecting specific patented technologies from external competition by creating dense thickets of strategic patents.

5.3 Exploring potential determinants of strategic patenting

Earlier studies identify several semiconductor firm characteristics that can explain firm differences in patenting rates. In this section, we perform an exploratory study of what firm characteristics determine strategic patenting rates in the semiconductor industry. We begin by estimating a strategic patent production function for the purpose of regressing strategic patent counts against firm-level data gathered from FACTSET and CAPITAL IQ. Strategic patents are defined as patents with shares of backward self-citations above or equal to 30%. Firms are assigned strategic patent counts for each year they applied for at least one patent that was later granted between 1998 and 2002. Note that we assign counts by the *application* year of the patents. Since there exists a lag between the time of application and the time of grant (usually around two years), we are interested in relating firm financial data to the year of application rather than to the grant year. The resulting panel dataset contains 255 observations where annual strategic patent counts 1993-2002 are assigned to our 49 semiconductor firms. We exclude patents with fewer than 5 total backward citations (see *section 4*) leading the number of annual observations to drop to 251. The time period 1998-2001 covers 70% of observations in the dataset, while 1993 and 2002 only account for 4 and 10 observations, respectively. The final sample contains 16,268 patents, out of which 2,491 are strategic.

Figure 5: Distribution of strategic patent counts 1993 - 2002



The strategic patent count variable *nstratpats* has a mean value of 9.92 and a median of 0, indicating a skewed distribution towards zero with 53% of observations having a strategic patent count equal to 0. The distribution of *nstratpats* is illustrated in *figure 5*. To account for the high share of 0-values and the non-normal distribution of our variable of interest, we follow Hall and Ziedonis (2001), Ziedonis (2004) and Noel and Schankerman (2013) in adopting a Poisson based regression model, first proposed in Hausman, Hall and Griliches (1984). The Poisson distribution is appropriate to apply to open ended count data as it describes the probability of the number of times a specific event occurs over a large number of observations. Inference based on the Poisson distribution requires that the population mean equals the population variance of the random variable (Moore et al,. 2020). We find that the standard deviation is roughly three times the size of the sample mean of our dependent variable, displaying clear signs of overdispersion (i.e. the population variance is greater than the population mean). Cameron and Trivedi's test for overdispersion further rejected the pure Poisson regression model. Thus, we follow Ziedonis (2004) and adopt a negative binomial regression model, a more appropriate Poisson based model in cases of overdispersion. In similar panel-datasets to our, previous studies have found presence of autocorrelation in the residuals of some variables within firm observations. For this reason we use clustered standard errors, correcting for eventual autocorrelation and heteroskedasticity in the sample (Ziedonis, 2004). Instead of OLS, coefficients are estimated using the maximum-likelihood estimation method.

Hausman, Hall and Griliches (1984) argue that the expected number of patents firms produce each year can be specified as an exponential function of R&D investments and other relevant firm characteristics. The authors suggest that R&D investments are most appropriately measured through contemporary R&D expenditure due to the high correlation of firm R&D expenditures over time. We specify the strategic patent production function, where the number of successful strategic patent applications per year (*nstratpats*) is our dependent variable, as an exponential function of the following variables:

- Firm size measured as the logarithm of net revenues in million 2010 USD (*logrev*). The variable is predicted to be positively correlated with strategic patent propensity. The sample median revenue is 2,795 with a median of 743.
- The logarithm of firm price-to-book ratio (*logpb*), measured as the average stock market capitalization of firms per year over their book value of equity. The variable is intended to proxy for differences in accumulated patent stock values between firms in the sample. Although firm size is assumed to be positively associated with patent values, the variable is included to capture some of the variation in patent stock private

values that is not explained by firm size. The rationale behind the variable is that firms typically only capitalize the costs associated with obtaining patents, implying that the true market value of a firm's patent stock is left out of the balance sheet. Assuming that accumulated patent stocks significantly impact the expected future profits of semiconductor firms, we argue that variations in the private values of patent stocks will, to some extent, be reflected in the variation of price-to-book multiples of our semiconductor firms. Thus, *logpb* is expected to be positively associated with strategic patent propensity. We however recognize that price-to-book ratios can exhibit considerable noise as they are sensitive to factors that are not related to patent values. To account for this, we argue that price-to-book ratios above 20x (17 observations) constitute outliers and a dummy variable (*high_pb*) is created to separate the impact of these observations from the rest of the sample. The sample mean price-to-book ratio is 8.81 while the median is 4.52.

- Capital intensity, measured as the book value of property, plant and equipment (PP&E) in million 2010 USD, normalized by full-time employees in hundreds (*logppenorm*).
- R&D intensity, measured as contemporary R&D expenditures in million 2010 USD, normalized by full-time employees in hundreds (*logrdnorm*).
- Year dummies are included to account for industry wide effects related to the economic climate, shifts in patent policy, stock market swings, and truncation. To avoid spreading the observations thin over too many dummies, each dummy covers two- or three years, except the baseline year 2000 which has 43 observations.
- In accordance with Hausman, Hall and Griliches (1984), we finally include a year specific strategic patenting mean measured as the logarithm of the average strategic patent count for all observations in a given year (*logavg_stratpats*).

We run a pooled negative binomial regression with clustered standard errors as our base model, the results of which are presented in *table 4* under model 1. Coefficients are presented in incidence rate ratios (IRRs) which should be interpreted as the multiplicative change in the expected count of the dependent variable *nstratpats* for marginal increase in the independent variable. As theory would predict, the results show that revenue is a strong predictor of how many strategic patents a firm will produce in a given year with an estimated IRR of 10.62 and a p-value of 0.000. It also appears that capital intensity positively influences strategic patent propensity. Thus, our results indicate that semiconductor firms with large technology specific sunk investments (e.g. manufacturing firms) are more inclined to pursue strategic patenting. A particularly interesting result is that R&D intensity is found not to be a significant predictor of our dependent variable, in contrast to the findings of previous studies of patenting propensity in the semiconductor industry. This could potentially relate to the idea that the technologies underlying strategic patents require less R&D effort to develop, rendering R&D investments to become a less important component of the strategic patent production function. The regression produces an insignificant estimated impact of *logpb*. This could imply that price-to-book ratios exhibit too much noise to be a reliable indicator of the accumulated patent stock values of firms.

To check the robustness of our results, we run the regression on a cross-sectional dataset where the variable values for each year a firm is present in the sample are averaged. The results of this regression are presented in model 2 where robust standard errors are used to account for potential heteroskedasticity. The main difference from the base model is that the significance of the estimated impact of R&D intensity improves to a p-value of 0.054 while the p-value of the capital intensity variable increases to 0.099. The estimated IRRs of the two variables remain similar however.

Table 4: Regression output

	Base model	**Robustness check**	**Base Model**	**Robustness check**
	Model 1	**Model 2**	**Model 3**	**Model 4**
Dependent variable:	*nstratpats*	*avg_nstratpats*	*share_stratpats*	*avgshare_stratpats*
Constant	0.000	0.000		
	(0.000) ***	(0.000) ***		
Year specific mean	14.042			
	(0.000)***			
Revenue	10.624	11.973	0.057	0.057
	(0.000) ***	(0.000) ***	(0.099)*	(0.046)**
Capital intensity	6.252	6.449	0.235	0.221
	(0.018) **	(0.099) *	(0.267)	(0.348)
R&D intensity	2.408	3.510	-0.107	-0.199

	(0.170)	(0.054) *	(0.591)	(0.419)
Price-to-book	1.222	1.076	0.161	-0.077
	(0.781)	(0.939)	(0.330)	(0.876)
Dummies:				
High pb	0.097	0.068	-0.286	-0.375
	(0.064)*	(0.134)	(0.179)	(0.147)
1993-1995	1.372		-0.000	
	(0.539)		(0.992)	
1996-1997	0.960		0.013	
	(0.919)		(0.055)	
1998-1999	1.280		0.069	
	(0.493)		(0.209)	
2001-2002	1.043		0.021	
	(0.910)		(0.633)	
Observations	*251*	*49*	*251*	*49*
Log-likelihood	-505.714	-79.602	-71.237	-71.237
Chi-squared	208.27	46.12	18.44	18.44
	(0.000)	(0.000)	(0.030)	(0.030)

Table description: Columns show regression results for model 1 (pooled negative binomial), model 2 (cross sectional negative binomial of average variable values per firm 1993-2002), model 3 (pooled fractional logit regression) and model 4 (cross-sectional fractional logit regression of average variable values per firm 1993-2002). In model 1 and 2, level to each variable name are the ML-estimated IRRs for each independent variable and in parentheses are the p-values. In model 3 and 4, level to each variable name are the quasi-likelihood estimated elasticities for each independent variable and in parentheses are the p-values. Model 1 and 3 use clustered standard errors, correcting for potential heteroskedasticity and autocorrelation within firm observations. Model 2 and 4 use robust, heteroskedastic-consistent standard errors. Log-likelihood and Chi-squared statistics indicate model fit.

While the negative binomial regressions in *table 4* indicate that firm size is a strong predictor of the expected number of strategic patents a firm will produce per year, it does not tell us anything about how many strategic patents larger firms are expected to produce in relation to the total number of patents produced. To examine whether firms increasingly redirect R&D resources from productive- to strategic patenting as they grow, we regress the share of strategic patents over all patents produced per year (*share_stratpats*) against our explanatory variables. Since we have a continuous fractional outcome for our dependent variable, bounded between 0 and 1, we need a regression model that will consistently estimate the mean of *share_stratpats* conditional on the independent variables within this closed interval.

The fractional logistic regression is used by Papke and Woolridge (2008) in a panel of test-pass-rates and does not require any specific distribution to calculate consistent estimators. As our base model we run a pooled fractional logistic regression with clustered standard errors, the results of which are presented in *table 4* under model 3. For our independent variables, we include 1) annual revenue (*rev*), 2) normalized book value of property plant and equipment (*ppenorm*), 3) normalized annual R&D expenditures (*rdnorm*), 4) average price-to-book ratio per year (*pb*). We include the same dummy variables as in model 1. The coefficients in model 3 and 4 are presented as elasticities and should be interpreted as the percentage change in the dependent variable given a marginal increase in the independent variable.

The results of model 3 indicate that our explanatory variables generally predict the number of strategic patents that are produced in relation to the total number of produced patents poorly. Due to uncertainty regarding the correct specification of the function, we also run a beta regression against the same variables and obtain almost identical results. The beta regression follows a beta distribution where the dependent variable can take any value in an open bounded interval between 0 and 1. Thus, for this regression, we transformed all 0-values to 0.001 (*share_stratpats* never takes the value 1 in our sample). Similar to model 2, we run the regression on cross-sectional data of the variable averages per firm with robust standard errors in model 4, generating a significantly positive effect of firm size with a p-value of 0.046. The estimated elasticity of revenue however turns out to be small, consistent with the results from the base model. Thus, the main finding from the fractional outcome regression is that, while greater firm size appears to be strongly associated with higher numbers of strategic patents produced, it does not seem to lead firms to produce productive patents at any lower rates. Instead, larger semiconductor firms appear to exhibit higher rates of patenting overall.

6. Discussion

The purpose of this thesis is to empirically examine the impact of preemptive patenting practices on technological competition in a sample of 49 US based semiconductor firms. Using patent self-citation shares to identify strategic patents, our research question is approached through two main perspectives. First, we examine whether strategic patents successfully preempt technical competition by analyzing patent-level citation data (*sections*

5.1 and *5.2*). We then perform an exploratory study of several variables from the literature that could potentially determine firm differences in strategic patenting rates (*section 5.3*).

It was initially proposed that patents with high shares of backward self-citations, representing strategic patents, will tend to receive fewer external forward citations. This prediction was based on the theoretical assumption that the technologies underlying strategic patents generally require less R&D investment to develop and, as a result, exhibit less technological significance. Further, since strategic patent thickets serve to deter rivals from developing adjacent technologies, fewer external follow-on inventions should also contribute to fewer external citations received by each strategic patent belonging to the thicket. Consistent with theory, both the Wilcoxon rank sum test and the Spearman's rank correlation test determine a negative correlation between high shares of backward self-citations and the number of external forward citations received. Both tests however suggest that these differences are, in general, very small. The linear regression model, generating insignificant coefficient estimates, further illustrate this relationship. Counter to our theoretical prediction, our results suggest that on average there is no economically significant difference in the number of external citations received by strategic and non-strategic patents. We proceed by running the same tests to determine the relationship between external forward citations and high shares of forward self-citations, reflecting protected patents, in our sample. For this specification, all test results suggest a substantial general negative relationship between the two variables which closely aligns to our theoretical prediction from *section 3.1*.

Altogether, results in *sections 5.1* and *5.2* show that, while clusters of strategic patents do tend to limit the number of external citations to the protected patent, the number of external citations received by the strategic patents themselves appear to be almost completely unaffected. One interpretation of this finding is that preemptive patenting efforts to deter entry into specific technological domains are generally not executed with complete success. Even though firms are able to protect specific technologies from being copied by rivals through the use of strategic patent thickets, over time it appears that entry into the general technological domain is often inevitable. Alternatively, the results could indicate that successful preemption does not exclusively involve strategic patenting of technologically insignificant and incremental inventions. Our finding that patents belonging to clusters of similar patents from the same firm generally receive the same number of external citations as other patents might imply that not all clustered patents are purely strategic. Instead, it is plausible that firms that have produced a valuable invention will, in addition, focus more of their R&D resources to developing productive follow-on inventions in that specific area. As argued in Gilbert (1987), successful preemption in most cases requires sustained high levels of intense product development to deter potential entrants from participating in the patent race. The implication that firms employ both strategic and productive patents to protect and reinforce their product market position suggests that preemptive patenting strategies in practice involve multiple layers of complexity that cannot be fully understood through solely analyzing strategic patenting.

The later interpretation of our results also relates to the possibility that a substantial number of patented technologies in the sample reflect successful internalization of the spillovers from specific inventions. If some of the highly self-citing patents in fact reflect productive follow-on inventions of the same firm stemming from superior innovative capabilities with a specific field, then rival inventors generally will be as inclined to cite these patents as they would any other. If clusters of highly self-citing patents mainly reflect productive follow-on inventions from the same firm, the usefulness of self-citations as a general indicator of preemptive efforts can be seriously questioned. As discussed in *section 3*, neither can we tell what influence citations from patents representing complementary technologies - e.g. from firms in different parts of the semiconductor value-chain - has on the data. In future studies, it would thus be interesting to assign citations to different types of firms to identify and separate these effects.

In *section 5.3* we employ two different regression models to study the relationship between some potentially relevant semiconductor firm characteristics and strategic patenting propensity. There is a widespread concern in the industrial policy literature that large high-tech firms are increasingly relying on unproductive investments to reinforce their market power, to the detriment of smaller competitors. Although we find evidence that larger semiconductor firms in our sample do produce significantly greater amounts of strategic patents than the smaller firms, we find no evidence that they do so at the expense of developing fewer productive patents. In other words, semiconductor firms, on average, appear to allocate R&D resources towards producing productive- and strategic patents in a roughly constant proportion as they grow. In estimating a strategic patent production function, we further find evidence that capital intensity, measured as the normalized book asset value of property plant and equipment per firm, is positively associated with strategic patent propensity. This echoes the findings in Hall and Ziedonis (2001) and Ziedonis (2004) that semiconductor firms with large technology specific sunk investments tend to patent more frequently to mitigate hold-up risks. Our finding suggests that this relationship also includes a higher propensity among capital intensive firms to produce strategic patents for preemptive purposes. Thus, the results might reflect higher expected costs associated with competition from patented substitute technologies among firms with large technology specific capital investments. Moreover, we find that R&D intensity, which typically is viewed as one of the strongest predictors of patenting, is a weak predictor of the number of strategic patents produced. This result offers validation to the view that the technologies underlying strategic patents in general require less R&D investment to develop than productive technologies. Finally, the results indicate that price-to-book ratios are a noisy measure of differences in private values of the accumulated patent stocks of semiconductor firms. Thus, we recommend future studies similar to ours to explore ways of measuring patent stock values more directly.

A final point worth discussing is to what extent our assumptions regarding what defines a strategic patent might have impacted our results. Strategic patents cannot be objectively defined as the extent to which a given patent serves preemptive purposes is not disclosed by the patent applicant. In this study, we distinguish strategic patents through the share of backward self-citations made to prior art. Patents with shares of backward self-citation above or equal to 30% are defined as strategic, while patents with shares of forward self-citations above or equal to 30% are defined as protected. By making these delimitations, we by no means claim to perfectly identify and separate strategic patents from non-strategic patents, even if it would be possible to make such exact distinctions. From the standpoint of the existing literature on strategic patenting and our empirical results, we however do believe that we have been able to capture a substantial share of strategic patenting practices in our sample through these measurements. The primary motivation of using self-citation shares rather than self-citation counts is to account for differences in citation rates between firms. However, from *figures 6* and *7* in the appendix, it appears that the general negative relationship between self-citations and external citations is robust to an alternative measurement of strategic- and protected patents through self-citation counts per patent. Our understanding of the most accurate method of using citation data to study strategic patenting would nonetheless benefit from further research exploring the pros and cons of using self-citation shares compared to self-citation counts in studies similar to ours.

7. Conclusion

This thesis has been committed to empirically studying preemptive patenting practices in the US semiconductor industry during the time period 1998-2002. By analyzing publicly available patent data from a sample of 49 US based semiconductor firms, we have attempted to identify strategic patents by measuring the share of backward self-citations made to prior art. Protected patents are correspondingly identified by measuring the share of forward self-citations. Several statistical tests were then performed to determine the impact of strategic patenting on technological competition. Our results indicate that while clusters of strategic patents significantly limit the number of external citations received by the protected patent, there appears to be no economically significant effect on the number of external citations received by the strategic patents themselves. In addition, we conducted an exploratory study of potential determinants of firm propensity to produce strategic patents. Our main findings suggest that firm size is an important predictor of strategic patenting. We however find no evidence that firms tend to produce more strategic patents in relation to non-strategic patents as they grow. We further find that capital intensive firms are more prone to strategically patent, while R&D intensity is a weak predictor of strategic patenting. Our study does not control for eventual citation linkages between different types of semiconductor firms, e.g. design- and manufacturing firms. To obtain a better understanding of preemptive patenting behavior in the semiconductor industry, we suggest that future studies on the topic explore this issue closer to see how accounting for firm type differences could impact our results. It is also suggested that further research is required to determine the most accurate method for using self-citation data to track strategic patenting.

Appendix:

Figure 6: External forward citations and forward self-citation counts per patent

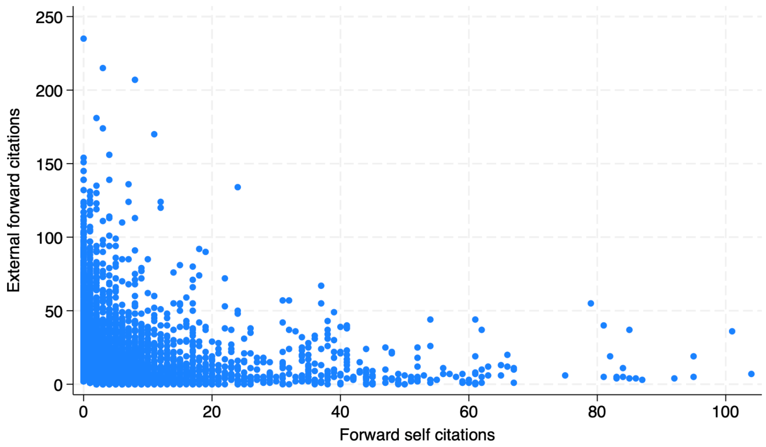


Figure 7: External forward citations and backward self-citation counts per patent

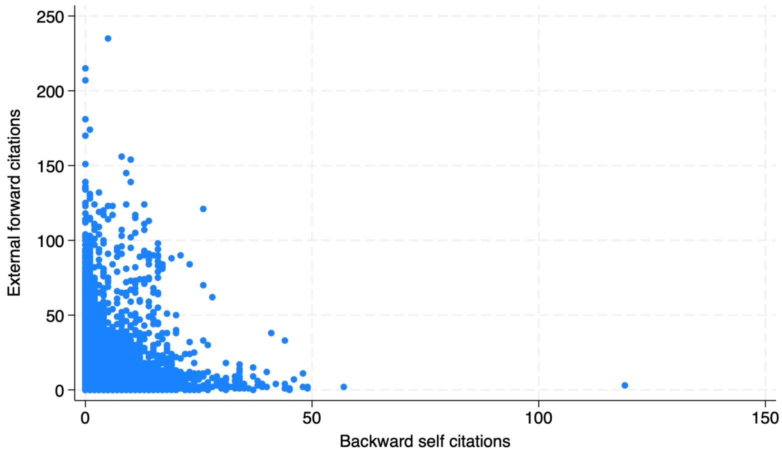


Table 5: Complete list of the semiconductor firms in the sample

Advanced Micro Devices, Inc.	Transmeta Corporation	Texas Instruments, Inc.
Amkor Technology, Inc.	Ramtron International Corporation	Wolfspeed, Inc.
Analog Devices, Inc.	Pericom Semiconductor Corporation	Conexant Systems, Inc.
Cirrus Logic, Inc.	TriQuint, Inc.	PMC-Sierra
Intel Corporation	R F Micro Devices	LSI Logic Corporation
Kopin Corporation	National Semiconductor Corporation	Integrated Device Technology
Lattice Semiconductor Corporation	Silicon Image, Inc.	Silicon Storage Technology
LINEAR Technology	Altera Corporation	International Rectifier Corporation
Microchip Technology, Inc.	Xilinx, Inc.	Electro Scientific Industries

Micron Technology, Inc.	NeoMagic Corporation	Xicor, Inc.
Monolithic System Technology, Inc.	NVIDIA Corporation	Semitool, Inc.
Burr-Brown Corporation	Plug Power Inc.	Sipex Corporation
Intersil Corporation	Power Integrations, Inc.	Vitesse Semiconductor Corporation
Maxim Integrated products	QUALCOMM, Inc.	Exar Corporation
Atmel Corporation	QuickLogic Corporation	TranSwitch Corporation
Cypress Semiconductor Corporation	Semtech Corporation	
Broadcom, Inc.	Silicon Laboratories Inc.	

References:

Akcigit, U., Abrams, D. S., & Grennan, J. (2013) ‘Patent value and citations: Creative destruction or strategic disruption?’, *NBER Working Paper No. 19647*, National Bureau of Economic Research, Cambridge, MA.

Akcigit, U. & Ates, S.T. (2023) ‘*What Happened to US Business Dynamism?*’, Journal of Political Economy, 131(8), pp. 2059–2124. Available at: https://doi.org/10.1086/724289.

Alcácer, J., Gittelman, M. & Sampat, B. (2009) ‘Applicant and examiner citations in U.S. patents: An overview and analysis’, *Research Policy*, 38(2), pp. 415–427. Available at: https://doi.org/10.1016/j.respol.2008.12.001.

Argente, D., Baslandze, S., Hanley, D., & Moreira, S. (2020) ‘Patents to Products: Product Innovation and Firm Dynamics’ *CEPR Discussion Paper No. DP14692*. Available at: https://doi.org/10.2139/ssrn.3577811.

Arora, A., Belenzon, S. & Sheer, L. (2021) ‘Matching patents to compustat firms, 1980–2015: Dynamic reassignment, name changes, and ownership structures’, *Research Policy*, 50(5), pp. 104217. Available at: https://doi.org/10.1016/j.respol.2021.104217.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., & Van Reenen, J. (2020) ‘The fall of the labor share and the rise of superstar firms’, *The Quarterly Journal of Economics*, 135(2), pp. 645-709.

Bessen, J., & Maskin, E. (2009) ‘Sequential innovation, patents, and imitation’, *The RAND Journal of Economics*, 40(4), pp. 611-635. Available at: https://doi.org/10.1111/j.1756-2171.2009.00081.x.

Bhaven, S. (2011) ‘USPTO Patent and Citation Data’, *Harvard Dataverse*. Available at: https://doi.org/10.7910/DVN/SJPHLG.

Cohen, W., Nelson, R. & Walsh, J. (2000) ‘Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)’, *NBER Working Paper No. 7552*, National Bureau of Economic Research, Cambridge, MA. Available at: https://doi.org/10.3386/w7552.

Decker, R. A., Haltiwanger, J., Jarmin, R. S., & Miranda, J. (2016) ‘Where has all the skewness gone? The decline in high-growth (young) firms in the US’, *European Economic Review*, 86, pp. 4-23.

Galasso, A. and Schankerman, M. (2015) ‘Patents and Cumulative Innovation: Causal Evidence from the Courts’, *The Quarterly Journal of Economics*, 130(1), pp. 317–369. Available at: https://doi.org/10.1093/qje/qju029.

Gilbert, R. J., & Newbery, D. M. (1982) ‘Preemptive patenting and the persistence of monopoly’, *The American Economic Review*, 72(3), pp. 514-526. Available at: http://www.jstor.org/stable/1831552.

Gilbert, R. J. (1987) ‘Patents, sleeping patents, and entry deterrence’, *Journal of Reprints for Antitrust Law and Economics*, 17(2), pp. 205-270.

Giuri, P. *et al.* (2006) ‘Everything you always wanted to know about inventors (but never asked): Evidence from the PatVal-EU survey’, *CEPR Discussion Paper No. 5752*.

Hall, B.H. & Harhoff, D. (2012) ‘Recent Research on the Economics of Patents’, *Annual Review of Economics*, 4(1), pp. 541–565. Available at: https://doi.org/10.1146/annurev-economics-080511-111008.

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2001) ‘The NBER patent citation data file: Lessons, insights and methodological tools’, *NBER Working Paper No. 8498*, National Bureau of Economic Research, Cambridge, MA.

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005), ‘Market Value and Patent Citations’, *The RAND Journal of Economics*, 36(1), pp. 16–38. Available at: http://www.jstor.org/stable/1593752.

Hall, B.H. & Ziedonis, R.H. (2001) ‘The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995’, *The RAND Journal of Economics*, 32(1), pp. 101–128. Available at: https://doi.org/10.2307/2696400.

Harhoff, D., Scherer, F.M. & Vopel, K. (2003) ‘Citations, family size, opposition and the value of patent rights’, *Research Policy*, 32(8), pp. 1343–1363. Available at: https://doi.org/10.1016/S0048-7333(02)00124-5.

Haskel, J., & Westlake, S. (2018) ‘Capitalism without capital: The rise of the intangible economy’, *Princeton University Press*.

Hausman, J., Hall, B. H., & Griliches, Z. (1984) ‘Econometric Models for Count Data with an Application to the Patents-R & D Relationship’, *Econometrica*, 52(4), pp. 909–938. Available at: https://doi.org/10.2307/1911191.

Jaffe, A., Trajtenberg, M., & Henderson, R. (1993) ‘Geographic localization of knowledge spillovers as evidenced by patent citations’, *The Quarterly Journal of Economics*, 108(3), 577-598.

Jaffe, A., Trajtenberg, M., & Henderson, R. (1997) ‘University Versus Corporate Patents: A Window On The Basicness Of Invention’, *Economics of Innovation and New Technology*, 5(1), pp. 19-50. Available at: https://doi.org/10.1080/10438599700000006

Lanjouw, J.O. & Schankerman, M. (2004) ‘Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators’, *The Economic Journal*, 114(495), pp. 441–465. Available at: https://doi.org/10.1111/j.1468-0297.2004.00216.x.

Moore, D. S., McCabe, G. P., Alwan, L. C., & Craig, B. A. (2020) ‘The practice of statistics for business and economics’, 5th Edition, WH Freeman and Company.

Moser, P., Ohmstedt, J. & Rhode, P.W. (2018) ‘Patent Citations—An Analysis of Quality Differences and Citing Practices in Hybrid Corn’, *Management Science*, 64(4), pp. 1926–1940. Available at: https://doi.org/10.1287/mnsc.2016.2688.

Noel, M. and Schankerman, M. (2013) ‘Strategic Patenting and Software Innovation’, *The Journal of Industrial Economics*, 61(3), pp. 481–520. Available at: https://doi.org/10.1111/joie.12024.

Papke, L.E. & Wooldridge, J.M. (2008) ‘Panel data methods for fractional response variables with an application to test pass rates’, *Journal of Econometrics*, 145(1), pp. 121–133. Available at: https://doi.org/10.1016/j.jeconom.2008.05.009.

Pepall, L., Richards, D. & Norman, G. (2013) ‘Industrial Organization: Contemporary Theory and Empirical Applications’, 5th Edition, Wiley.

Scotchmer, S. (1991) ‘Standing on the Shoulders of Giants: Cumulative Research and the Patent Law’, *Journal of Economic Perspectives*, 5 (1), pp. 29-41.

Shapiro, C. (2000) ‘Navigating the patent thicket: Cross licenses, patent pools, and standard setting’, *Innovation policy and the economy*, 1, pp. 119-150.

Torrisi, S. *et al.* (2016) ‘Used, blocking and sleeping patents: Empirical evidence from a large-scale inventor survey’, *Research Policy*, 45(7), pp. 1374–1385. Available at: https://doi.org/10.1016/j.respol.2016.03.021.

Trajtenberg, M. (1990) ‘A Penny for Your Quotes: Patent Citations and the Value of Innovations’, *The RAND Journal of Economics*, 21(1), pp. 172. Available at: https://www.jstor.org/stable/2555502.

Von Graevenitz, G., Wagner, S. & Harhoff, D. (2013) ‘Incidence and Growth of Patent Thickets: The Impact of Technological Opportunities and Complexity’, *The Journal of Industrial Economics*, 61(3), pp. 521–563. Available at: https://doi.org/10.1111/joie.12032.

Ziedonis, R.H. (2004) ‘Don’t Fence Me In: Fragmented Markets for Technology and the Patent Acquisition Strategies of Firms’, *Management Science*, 50(6), pp. 804–820. Available at: https://doi.org/10.1287/mnsc.1040.0208.