# Transferability MATTRs: Towards Understanding Antecedents of Strategic Licensing *

Dafna F. Bearson
Harvard Business School

Maria P. Roche
Harvard Business School

March 18, 2025

**Abstract**

This paper examines the antecedents of firms' decisions to license intellectual property (IP). We propose a conceptual framework emphasizing two critical factors: (1) transferability, which we define as the ease of moving knowledge embodied in inventions across firm boundaries, and (2) relevance, representing the invention's importance to a firm's value creation and capture processes. Leveraging a novel proxy for transferability—the moving-average type-token ratio (MATTR), which measures the lexical diversity of patent descriptions—we analyze licensing agreements for U.S. patents granted from 1980 to 2015, integrating this data with patent- and firm-level characteristics. The results reveal that higher MATTR is associated with a greater likelihood of licensing, suggesting that firms strategically enhance transferability for potential licensees as they apply for a patent. Moreover, the relationship between transferability and licensing is moderated by the invention's economic and scientific value, and the degree of demand uncertainty in the industry. These findings contribute to the literature on innovation strategy and technology markets, highlighting the strategic selection of IP for licensing and the dynamic interplay of invention characteristics, firm strategy, and industry context in shaping licensing decisions.

**Keywords**: *Strategy, Technology, Licensing, Intellectual Property, Innovation*

*We thank Juan Alcácer and Shane Greenstein for their guidance and contributions. We also thank Patrick Clapp, Sazan Khalid, Madushani Rajapaksha, Christine Rivera, and Richie Zitomer for excellent research assistance. Many thanks to Bharat Anand, Bart Clarysse, Andrea Contigiani, Annamaria Conti, Javier Gimeno, Carlos Serrano, and Tarun Khanna for advice on drafts. Thank you, Jörg-Matthias Roche for sharing your expertise in Linguistics. This manuscript benefited from comments provided by participants at the Academy of Management, KU Leuven MSI, ATEC and the Harvard Business School Strategy Unit Seminar and Research Development Course. We gratefully acknowledge funding from the Harvard Business School Division of Research and Faculty Development. All errors and omissions are our own. Contact: dbearson@hbs.edu; mroche@hbs.edu

# 1 Introduction

Innovation, defined as the commercialization of invention (Schumpeter, 1934), is an important driver of business productivity and growth. Over the past several decades, the number of firms exchanging inventions in markets for technology (MFT) has increased at stunning rates (Kwon et al., 2022; Serrano, 2010). Existing studies offer valuable insights on technology markets broadly, but one necessary condition has remained understudied to date. Namely, to enable transactions in MFT, inventions must first be supplied—typically by an organization.[1]

There are different ways to supply inventions to MFT.[2] Our focus in this paper is on licensing as a mechanism for transferring inventions and its strategic implications. Licensing involves granting another party the rights to use, produce, or commercialize an invention while retaining ownership, enabling the licensor to maintain control over the invention and capture ongoing value through fees and royalties. This approach allows licensors to leverage the expertise and resources of licensees for downstream activities such as production and distribution. A substantial body of research explores various aspects of interfirm licensing, including the motivations for licensing (e.g., Shapiro (2000)), the timing of licensing agreements (Allain et al., 2016; Hegde & Luo, 2018), and the design of optimal licensing contracts (Hegde, 2014). One aspect that has fallen under the radar, however, is that inventions may differ in the extent to which they are licensable and that this may be a inherently strategic choice of the inventing firm.

In this study, we embrace this type of heterogeneity by asking what the antecedents of licensing inventions are. We thereby conceptualize the likelihood of licensing to depend on two key factors: 1) transferability, which operates at the invention-level and is defined as the ease with which knowledge embedded in an invention can be moved across firm boundaries,

[1]Individuals can also supply inventions in MFT. However, the focus of this paper is on companies that serve as suppliers in MFT.

[2]Two of the most salient ways are licensing and selling. Licensing and selling represent distinct mechanisms for transferring inventions, each with unique strategic implications. In contrast to licensing, selling an invention entails transferring ownership outright, which provides a one-time revenue stream but relinquishes control over future development and commercialization (Gans & Stern, 2003). The literature provides some guidance on when selling is chosen, such as when a firm lacks complementary assets to commercialize the invention effectively (Teece, 1986).

and 2) relevance, which operates at the firm-level, and is defined as the importance of the invention to the firm's value creation and value capture. Prior research on licensing emphasizes aspects of relevance, such as the degree of competition in the technology and product market (Arora et al., 2001; Moreira et al., 2020; Torrisi et al., 2016) and ownership and availability of complementary assets (Gans & Stern, 2003; Teece, 1986, 2006). This paper focuses on transferability. Critically, we propose that the degree to which an invention is transferable is a firm-level choice. We hypothesize that a firm will alter the transferability of a given invention by changing information frictions as a function of expected value of the focal intellectual property (IP) to the firm.

Transferability depends on factors such as the intelligibility of the invention by external organizations with domain expertise. Conditional on enforceable IP protections, transferability is enhanced when the knowledge underpinning the invention is organized to be understandable and reproducible by external parties. While prior research in factor markets highlights the inherent tradeability of technologies based on their tacit or codified nature (Barney, 1986; Dierickx & Cool, 1989), we argue that firms influence transferability through strategic decisions about how technologies are described. For instance, firms may leverage or create information frictions to shape external expectations about the value of a resource (Kankanhalli et al., 2024; Makadok & Barney, 2001), thereby affecting the costs and feasibility of technology transfer. Consequently, we posit that transferability is distinct from tradability; it not only reflects the inherent characteristics of the technology but also strategic choices about its presentation and the extent to which information frictions are purposefully reduced or heightened.

Using our conceptual model as guidance, we build a unique data set on the firm-license- and patent-license-level that combines novel data with existing and widely established sources to empirically examine the decision to license U.S. patents granted to public companies from 1980 to 2015. We analyze patent licensing agreements from the U.S. Securities and Exchange Commission's EDGAR data compiled by RoyaltyStat. To proxy for transferability, we introduce a measure from linguistics called the moving-average type-token ratio (MATTR). The MATTR is calculated by analyzing the semantic context of a text to assess lexical diversity. It computes the number of unique words divided by text length in a smoothly moving window. The MATTR proxies for the descriptiveness of the knowledge embodied in a patent and thus its intelligibility to outside organizations. In a sense, this measure captures the degree to which codified knowledge is described (i.e., the same codified knowledge may be written differently). Therefore, information frictions decrease with the MATTR. We calculate the MATTR for over 1.7 million patents, providing a new text-based empirical window into strategic patenting (Hall & Ziedonis, 2001). To assess the value of patents, we rely on data provided by Kogan et al. (2017) on the private economic value and the scientific value of patents. We supplement these data with information on firm characteristics from Compustat.

Naturally, in this context, identification concerns such as omitted variable bias, selection, and simultaneity abound. We take a multi-pronged approach to alleviate these concerns to the extent possible. First, our goal is to establish correlations, while using a host of fixed effects to keep constant industry, year and other trends that could impact the decision to license. In selected models, we further hold constant firm characteristics. Here our goal is to compare two technologies that are similar but differ along the transferability dimension. We supplement these analyses with text matching to identify similar underlying technologies. Finally, we implement a coarsened exact matching approach where we match on firm characteristics, including assets, R&D share, sales, employees, capital stock, and operating income.

Our results reveal noteworthy patterns about the features of licensed IP and attributes of inventing firms that may influence licensing decisions. We find that firms are 4 to 8 times more likely to license patents with higher MATTR. Importantly, we establish that the MATTR varies with the expected value of the patent. In addition, we exploit industry differences in the ability of firms to discern patent value (i.e., whether commercial potential is somewhat predicable at an early stage).[3] We show that there is a stronger relationship between the MATTR and licensing for firms in industrial settings where predictions about market demand can be made with some degree of certainty (Scott et al., 2020). Indeed, a higher MATTR is associated with

[3]Given that the patent exists, we assume that technological uncertainty has at least been resolved to some extent.

a fourfold increase in the likelihood of licensing in Biopharma, whereas in IT, the increase is only threefold.

Our findings carry important implications for at least three streams of literature. First, we contribute to the literature on MFT that facilitate the diffusion and efficient use of existing inventions (Arora et al., 2001; Banerjee & Siebert, 2017; Cassiman & Veugelers, 2006; Haucap et al., 2019). We advance this line of work by examining the antecedents of MFT and what drives firms to license their inventions. We focus on the decision to license as opposed to other channels for technology transfer, and its unique strategic implications.

Second, our findings speak to the broader firm innovation strategy literature. In this line of work, the assumption often prevails that firms can choose different ways to source inventions. One of the most widespread models suggests that firms have agency to decide whether to build, borrow, or buy inventions (Capron & Mitchell, 2012). We propose that, though this view is useful and warranted, firms may also need to assess the quality and type of inventions that are put up to "borrow" in MFT in the first place. If these are supplied based on calculations of the expected value to the inventing firm, this implies that inventions supplied to MFT represent a strategically highly selected type of invention—potentially not the providing firm's best. This is an important consideration as an increasing share of firms source inventions from external parties (Eklund & Kapoor, 2022; Hoang & Rothaermel, 2010). Furthermore, we illustrate critical boundary conditions rooted in the ability to predict market demand by industry.

Third, by introducing transferability as new terminology in the context of licensing and using MATTR as a proxy, we contribute to efforts to characterize the strategic nature of inventions embodied in patents (Datar et al., 2024; Kong et al., 2023). Our measure builds upon a growing body of work that uses text analysis to characterize important features of markets, such as companies' technology portfolio (Arts et al., 2023) or product market position (Hoberg & Phillips, 2016). It also expands upon other research that leverages patents to assess the scientific nature (Arora et al., 2018), scope (Lee, 2023), and product and process intensity (Ganglmair et al., 2022) of inventions. The MATTR opens up an important avenue for a deeper understanding of long-established patterns in patent trade, and provides granular insight into systematic differences across inventions.

This paper proceeds as follows. In Section 2, we develop the basic conceptual framework to guide the empirical predictions and interpretations of findings. Section 3 presents the empirical strategy and an overview of how the data were constructed. In Section 4 we describe and discuss the main results. We conclude in Section 5 with limitations, implications, and directions for future research.

# 2 Conceptual framework

## 2.1 Licensing

In this paper, we focus on licensing from the perspective of the firm that produces the invention that is being licensed out.[4] These boundary conditions are important to keep in mind throughout the paper. Licensing represents the transfer of certain aspects of control across firm boundaries, while the firm that produces the invention (the "licensor" throughout the paper) retains ownership. The user of the invention (the "licensee" throughout the paper) pays for the right to use, make, and sell the invention. In exchange for a fee and royalty payments, the licensee owns the final product and typically has control of downstream development decisions such as financing, testing, production, distribution, and marketing. A large literature provides critical insight on aspects of interfirm licensing such as the factors that motivate licensing (de Bettignies et al., 2023; Shapiro, 2000), the timing of licensing (Allain et al., 2016; Hegde & Luo, 2018), and the optimal structure of licensing contracts (Hegde, 2014).

Most of the work in this space takes for granted that inventions can be licensed. Put differently, the literature implicitly assumes that inventions are homogeneous in their propensity to be licensed. We relax this assumption by conceptualizing the likelihood of licensing as a function of transferability and relevance, and illustrate that there may, in fact, be selection of inventions into licensing transactions based on these factors. As we describe in what follows, transferability in particular may be strategically altered by the firm to influence information

[4]We recognize that licensing is a two-way transaction, where both the licensor and licensee play important roles. However, our focus is on the strategic motivations of the licensor, which has been relatively understudied compared to the demand-side.

frictions associated with exchanging an invention in MFT.

## 2.2 Transferability

We introduce *transferability* of an invention as a determinant of licensing. In this paper, transferability is defined as the ability to move knowledge embedded in an invention across firm boundaries. Conditional on being able to protect the invention through IP rights and enforcement,[5] transferability increases with the degree to which the invention is intelligible to an outside organization with expertise in the technological domain.[6] Information frictions surrounding the invention decline with transferability.

A large literature is dedicated to understanding the relationship between the nature of invention and its influence on strategic factor markets.[7] Dierickx and Cool (1989) observe that not "all resources are actually bought and sold" in factor markets (p. 1505). Indeed, resource stocks that are accumulated largely through tacit knowledge are nontradeable and, consequently, must either be "built" internally or acquired through M&As. On the opposite extreme, codification represents the commodification of knowledge into a form that is more readily exchanged and traded via factor markets. As such, tradability occurs naturally through characteristics of an invention that enable it to be codified and/or, perhaps more importantly, by organizing knowledge such that it is understandable to and reproducible by an outside organization with expertise in the technological domain.

The latter possibility is in large part overlooked in the literature on factor markets for tech-

[5]Strong and enforceable IP protections are important elements of transferability, which affects the firm's ability to profit from other firms' use of its invention. In settings with weak IP rights and enforcement, firms may have greater incentive to commercialize internally in order to profit from their investment in the invention (Zhao, 2006). We assume strong IP protections exist for the inventions that we discuss for the remainder of this paper.

[6]Transferability may also hinge on other factors such as geographic proximity between the licensor and licensee. We expect that geographic agglomeration (Alcácer & Chung, 2007) and proximity through micro-geographies (Roche, 2020; Roche et al., 2024) affect the costs of transferring and commercializing (Lerner et al., 2024) specialized knowledge—especially involving tacit know-how, which represent a critical component to enabling knowledge flows and transfer beyond the codifiable. Micro-geographies influence on licensing and technology transfer more generally is beyond the scope of this paper, but it is the subject of another paper that is a work in progress.

[7]Barney (1986) defines a strategic factor market as "a market where resources necessary to implement a strategy are acquired" (1231).

nology. Indeed, a strong assumption in this research is that the degree to which an invention is tradable depends on the underlying nature of the invention itself. We ascertain that firms that produce inventions face choices about how to describe them, which may render otherwise similar underlying inventions more or less transferrable. To this point, recent research suggests that information costs (Hegde & Luo, 2018) and the way in which inventions are described affect patent trade. For example, at one extreme, Arora et al. (2018) assert that technologies conceptualized in "scientific terms makes them easier to codify, reducing search costs for potential buyers, and enables buyers to evaluate and integrate inventions. This should reduce transaction costs that are thought to afflict trade in technology, as well as enhance the potential gains from trade" (p. 7192). At the opposite extreme, Guo et al. (2017) illustrate that by "using vague language to make strategic information less precise, a firm's managers may create ambiguity that hampers rivals' market entry decisions" (p. 2075). As Grant (1996) explains, "knowledge is revealed by its communication" (p. 111).[8] These choices about how companies describe inventions affect transferability, which is distinct from tradability in that it is not simply a function of the nature of the invention and the degree to which it can be codified. This decision is a strategic one that that affects the costs associated with technology transfer.

We focus on inventions that are tradable by definition—whereby it is technically feasible to fully codify the knowledge embedded in the invention. We do this in order to better isolate the role of variations in descriptions of otherwise similar, codifiable inventions. In other words, we are not interested in underlying differences in the nature of the invention, rather in the way that the invention itself is described.

If we accept that companies may describe similar technologies differently to influence transferability, then the salient questions become why and under what conditions. One possibility is that firms leverage information frictions for competitive advantage. Past scholarship shows that information frictions affect the ability to form expectations about the future value of a resource (Barney, 1986). Indeed, asymmetries in expectations about the future value of a

[8]Grant (1996) distinguishes between different types of knowledge. In particular, explicit knowledge represents knowing about facts and theories, while tacit knowledge represents know how. In this quote he is referring to explicit knowledge.

resource may provide competitive advantage—in particular arising from the accuracy of expectations (Makadok & Barney, 2001). In one example of public firms selectively disclosing information to influence market expectations, Kankanhalli et al. (2024) find that firms redact information associated with more valuable IP in license agreements, and that these nondisclosures serve as informative signals to the market. In another example, Hsu and Ziedonis (2013) show that patents serve as informative signals that reduce information frictions between investors and entrepreneurial firms in the semiconductor industry. These examples point to firms leveraging—and even creating—information frictions associated with their IP to influence expectations for outside organizations.

To summarize, a first order predictor of the likelihood of licensing is the transferability of the invention. We expect that licensing increases with transferability, as information is more readily available and costs associated with moving specialized knowledge across firm boundaries decline. *Specifically, we expect that the inventing firm's propensity to license an invention increases with the degree to which the invention is made transferable.*

## 2.3 Relevance

We term the importance of the invention to the firm's value creation and capture processes *relevance.* Relevance informs the optimal commercialization approach based on the expected value of invention to the firm and the firm's ability to capture value generated from the invention.

The existing literature focuses on licensing as a function of competitive forces (de Bettignies et al., 2023) and ownership of complementary assets (Gans & Stern, 2003; Teece, 1986).[9] Firms' position in the factor market affects the bargaining power associated with negotiating the terms of the license agreement, while position and level of differentiation in the product market influences the degree of risk associated with potential competitors using the invention for downstream development of products and processes. Licensing increases with the firm's power in the factor market and in product markets outside the scope of the firm or undifferentiated

[9] Birhanu and Gambardella (2024) examine the non-economic forces that influence the likelihood of licensing. In particular, they find that family firms are less likely to license and more likely to commercialize internally because of higher preference for control than non-family firms.

product markets in which the firm competes (Arora et al., 2001; Baker et al., 2008).

Licensing allows companies to accrue higher returns to R&D if they do not own the resources—including the knowledge or expertise—to exploit the full potential of their invention (Degnan, 1999). Indeed, the gains from trade associated with technology reallocation is a leading explanation for technology markets (Serrano, 2010, 2018). This results in a division between firms that specialize in innovating versus typically larger firms that have the capital and other resources required for commercialization. Less research examines the role of large firms that both invest in innovating and own complementary assets for commercialization, yet choose to license. Given that value capture capabilities are not differentiating factors among these companies (i.e., resource constraints and capabilities are less likely to impede in-house commercialization), we expect that licensing is driven by the expected value creation from the invention. Put differently, for these companies, licensing is less motivated by constraints on ability to commercialize the invention in-house; rather, licensing occurs because more value is created through outside organizations' use of the invention.[10]

We expect the relationship between value creation and licensing to be nonlinear. At one extreme, companies license highly valuable inventions. Indeed, licensing has been used for standard setting, whereby any firm seeking to produce a compliant product must obtain a license to avoid facing infringement action (Shapiro, 2000).[11] At another extreme, licensing has been used as a way to profit from lower value inventions that are otherwise underutilized or shelved (Basir et al., 2012; Moore, 2005). In expectation, firms license their most valuable inventions from which more value is created through licensing than internal commercialization. Alternatively, companies license their least valuable technologies that represent low competitive

[10]Both value creation and value capture capabilities are key inputs to licensing. However, we focus on large, innovative firms and, therefore, emphasize the outsized role of value creation. This focus aligns with our empirical setting, where we examine large public companies—a relatively homogeneous group in terms of their value capture capabilities.

[11]One concern may be that licensing is driven by litigation. Galasso et al. (2013) examine the relationship between patent trade and litigation. They find that under 5% of traded patents are litigated. On average, trade reduces the likelihood of litigation—especially when the buyer has a large patent portfolio. They also do not find evidence of a large role of patent assertion entities (i.e., patent trolls), which are companies that acquire patents in order to generate revenue by pursuing infringement action or through licensing rather than through developing or manufacturing the underlying invention, although they caution against generalizing this conclusion based on individually-owned patents to corporate patent transactions.

risk and would otherwise create little to no value for the firm.

Considering relevance and transferability together, we expect that transferability increases if an invention is more valuable to the firm via licensing.[12],[13] Specifically, if low and high value inventions generate higher expected payoffs from licensing compared to strictly internal commercialization and transferability is a function of the expected payoff from licensing, then low and high value inventions will be made more transferable. *We therefore expect that the inventing firm will alter the transferability of an invention by changing information frictions as a function of expected value of the invention to the firm.*

## 2.4 Expectations

As firms increase transferability, search costs for potential licensees decline and licensors shape expectations about the potential returns to their invention, thereby altering information frictions associated with trading IP. As a consequence, transferability may influence both search for and selection of licensed inventions.

Figure 1 is a simple depiction of our conceptual framework. We expect that if transferability and relevance are low, then it is unlikely the firm will license because of high transaction costs and relatively low expected gains from trade. If transferability is low and relevance is high, it is also unlikely that the inventing firm will license because it would represent a potential competitive risk. If transferability and relevance are high, there are two possibilities. In some cases, firms may choose not to license due to competitive risk (e.g., patent for a compound key to a blockbuster drug). However, if the firm has reason to believe that they can generate substantial value and/or potentially set an industry standard, then they may choose to license (e.g., patent for a discovery such as CRISPR or monoclonal antibodies).[14] If transferability is

[12]For example, in the telecommunications industry, Toh and Miller (2017) argue that firms are more likely to disclose to standard setting organizations they have complementary technologies, as the firm stands to profit from nondisclosed complementary technologies as well through the disclosure.

[13]Licensing and selling are two mechanisms to commercialize IP via technology markets—each with different strategic implications. We focus on licensing and its relation to transferability, although we are building upon this framework as it applies to selling in other work (Bearson, 2025).

[14]These differences are particularly important when considering the way that licensing strategies vary across industries. For example, in semiconductors, where patents are more incremental and fast-paced, cross licensing across companies is more common. In biotech and pharmaceuticals, where inventions are more breakthrough

high and relevance is low, we expect the firm to license given that they have little to gain from the invention and have the opportunity to earn fees and royalties off of other firms' use of it without competitive risk.

[Figure 1 about here]

## 3 Data

To examine our predictions and validate our model, we build a unique dataset consisting of both novel and established data sources at the patent-level. The variables in our dataset are divided into the three main components of our conceptual framework: licensed and non-licensed inventions, transferability, and relevance. For a description of all the variables used for estimation and their original source, please refer to the Appendix Table A1.

We begin with a database of patents belonging to publicly traded companies from Kogan et al. (2017). We keep patents from the updated data that were issued between 1980 and 2015, which include nearly 1.8 million unique observations with measures of the market and scientific value of each patent. To characterize the inventions in our sample, we link patent abstract, technology class, and patent family information from PatStat.[15] Next, we merge our database of licensed patents. Licensed patents are identified through public companies' 8-K and 10-K filings to the U.S. SEC collected by RoyaltyStat.[16] There are 3,454 patents licensed by 596 unique firms between 1994 and 2015.

We collect data on company characteristics from Compustat. Specifically, we use the CRSP/Compustat Merged - Fundamentals Annual database. We retrieve measurements of total assets (*at*), employees (*emp*), sector (*gsector*), sales (*sale*), expenditure on R&D (*xrd*),

and fundamental, exclusive licensing is more common.

[15] We match the majority of patents from the Kogan et al. (2017) data to PatStat. However, approximately thirty-five thousand patents do not have associated entries in PatStat. Although we do not find any meaningful correlations between missingness and our patent- or firm-level variables, we collect the missing abstracts from Google Patents. Patent number US5787796 is US5839352A in Google Patents. We are unable to find data for US4326083.

[16] RoyaltyStat LLC offers a proprietary database of IP licensing agreements. They source these data from the universe of filings with exhibits to the U.S. SEC and update the database daily. They identify license agreements using a natural language processing algorithm, which achieves high accuracy, and through keyword identification (e.g., "license", "royalty", "granting license", "granting rights", etc.). The team manually checks filings when there is ambiguity in classification.

operating income adjusted for depreciation (*opincar*), and capital stock (*cstk*). We collect these variables from January 1980 through December 2015. We drop observations from the Kogan et al. (2017) database that do not match with Compustat data.[17]

We use the Compustat variables to construct proxies for relevance. For the purpose of this study, value creation capabilities are measured as R&D expenditure as a share of gross sales. Value creation is measured as the average 1) private economic and 2) scientific value of firms' patents. Naturally, these two measures differ in the extent to which they capture expectations versus realized value and commercial versus scientific potential. Value capture is measured by replicating the price cost margin from Aghion et al. (2005): operating profits net of depreciation, provisions less financial cost of capital divided by sales, where the cost of capital is assumed to be 0.085 for all firms and time periods. The intuition is that firms with higher price cost margins have more market power—representing the ability to capture a larger share of value created from their patents.

Our final sample for econometric analysis consists of an unbalanced panel of 7,746 firms. These are public companies that have at least one patent granted from 1980 to 2015. They own 1,764,464 patents in total.

## 3.1 Moving-average type-token ratio (MATTR)

A major empirical contribution of our paper is the introduction of a novel proxy for transferability, the moving-average type-token ratio (MATTR).[18] The MATTR is an established measure of lexical diversity in computational linguistics (Covington & McFall, 2010). It has been shown to be simple but useful measure of descriptiveness through the "range and variety of vocabulary" used in a text (Kyle, 2019; McCarthy & Jarvis, 2007) and correlated with other

[17] If a firm's main activity is to license technologies to downstream manufacturers, then their Standard Industrial Classification (SIC) class will be 6794—patent owners and lessors. Only 4 companies in our data fall into this SIC code, and we exclude them from our analysis.

[18] The ideal measure of transferability would estimate the information costs associated with a text. This is infeasible in practice—in part because information costs depend not only on the text itself but also subjectively on the reader. By measuring the lexical diversity of a text, the MATTR provides insight into how descriptive a text is. While it is imperfect, it provides insight into how much information is provided in a text unbiased by the content or the length of the text.

similar measures of text richness (Shaib et al., 2024). These measures represent the degree to which a text is informative through writing quality and competence (Kong et al., 2023). The descriptiveness of a text increases—and information frictions decline—with the MATTR.

The MATTR is based on the type-token ratio (TTR), which is calculated by dividing the number of unique words by the full text length. The TTR estimates the share of the words in a text that are unique. However, this simple calculation is biased by text length. For example, a word is more likely to appear twice in a longer text—potentially biasing the TTR downward as text length increases. Therefore, in our analysis, we use the MATTR, which is a more robust measure that calculates the TTR across a smooth moving range. We calculate the MATTR for each patent abstract in our sample using smooth moving ranges of 10, 30, and 50 words. Figure 2 provides the formulas for the TTR and MATTR, and an illustrative example of the mechanics of calculating the TTR and MATTR(30) on a patent abstract.

[Figure 2 about here]

We draw from prior research to validate the use of the MATTR. Kong et al. (2023) use a variety of measurements from computational linguistics to measure the disclosure of U.S. patent applications in nanotechnology, batteries, and electricity. The authors hypothesize that disclosure levels will be higher for university patents compared to corporate patents that are otherwise similar. This is because universities and corporations follow different business models for patenting, whereby the former focuses on generating income through licensing and the latter focuses on in-house commercialization. In addition to the TTR, the study uses Gunning Fog, Kincaid, Flesch, average age of acquisition of words (AoA), dependent clauses to total clauses ratio, content word overlaps, mean length of T-unit (MLT), and the proper noun ratio to measure the disclosure function of patents. All the measures, combined with Principal Component Analysis, are consistent in predicting that university patents are associated with more disclosure than corporate patents. Relatedly, Arts et al. (2025) use Gunning Fog to control for detail and clarity in a scientific paper. In another study, Shaib et al. (2024) compare measures of lexical diversity generated by large language models, including the Self-repetition score, Hypergeometric Distribution D, and compression ratios, and present high correlations between these measures and the MATTR.[19] These studies validate the MATTR by demonstrating its relationship to other measures of lexical diversity and support its interpretation as a measure of information availability.

To further validate our measure and identify potential boundary conditions, we provide a random sample of patents to experts in linguistics. Their assessment coincides with our approach: the MATTR is consistently higher for the patents that the experts tag as more descriptive. Taken together, this reinforces that the MATTR provides insight into how descriptively a text is written and how easily it may be understood—particularly by a reader with domain expertise—thereby influencing search costs, mitigating or heightening information frictions, and shaping expectations about value.

Our measure is constructed using publicly available data from PatStat, making it replicable and scalable for innovation research.[20] While past research focuses on strategic reasons for patenting (Grindley & Teece, 1997; Hall & Ziedonis, 2001; Torrisi et al., 2016), the nature of the strategy embodied in the patent is more difficult to characterize and observe. Existing measures that exploit text analysis seek to characterize the relationship to the firms' and competitors' technology portfolio (Arts et al., 2023) and relative product market position (Kwon et al., 2022). In contrast, our measure allows us to capture strategic characteristics of the patent while being agnostic to the technology, market, and competitive environment in which it is applied. Our measure opens a new window into the characteristics of an invention that impact its transferability.

## 3.2 Data limitations

The construction of our database presents several challenges. A company that files an 8-K or 10-K to the U.S. SEC is not always the licensor. Although cases where the filer is not the

[19] Additional text-based measures of readability are discussed in the Appendix of Hengel (2022).

[20] We use Python Linguistic Analysis Tools (pylats) to preprocess the text and retrieve the tokens. We then use the Tool for the Automatic Analysis of Lexical Diversity (TAALED) to calculate the TTR and the MATTR(10), MATTR(30), and MATTR(50) using the tokens obtained from pylats.

licensor are atypical, in order to verify our match to firm characteristics from Compustat, we algorithmically check for accuracy. We do this by initially comparing the text similarity of the filing company to the SEC to the Compustat firm name. Here we find that 80% of the company names match, and we manually check the remaining company names. After correcting matches, we drop remaining observations that do not properly match, often because they were either individuals or universities.

Public companies are required by law to report licensing transactions to the SEC that exceed a "materiality threshold." According to the US Securities Acts of 1933 and 1934, companies are required to disclose agreements "not made in the ordinary course of business" that contain information that would significantly change the total information available in an investor's decision (Munter, 2022). In practice, companies and auditors often adopt a quantitative reporting threshold of 5% of net income (Vorhies, 2005). But because the materiality threshold is not precisely defined by law, it is largely determined by company discretion. As a result, our sample represents companies' most valuable patent licensing—transactions that the company chooses to report voluntarily—which is a fraction of all licensing transactions. Surveys documenting the commercial applications of patents suggest that licensing could be much more common (Cohen et al., 2000; Sadao & Walsh, 2009; Torrisi et al., 2016). Our results should be interpreted cautiously as presenting the most critical IP relative to the universe of licensed patents.

Licensing is a mechanism to transfer IP, technology, and knowhow beyond patents. For example, Kapacinskaite et al. (2024) examine trade secret licensing. We recognize that patents are a partial measure of invention with many limitations, although they are associated with meaningful innovation and business outcomes (Farre-Mensa et al., 2020). We restrict our analysis to patents due to data limitations quantifying the value of other types of IP that may be included in a license agreement.

# 4 Results & discussion

## 4.1 Summary statistics

In Table 1 Column 1, we report the industrial composition of the companies in the sample. The composition of public firms that produce inventions is different from the composition of public firms overall, with higher concentration of companies in industries that patent more (MSCI, 2022). Among the companies in the sample, 22% are in Biopharma and other health related industries and 28% are in Information Technology (IT). Column 2 reports the share of licensing firms by industry, showing that licensing has a larger influence in technology transfer in Biopharma than in any other industry. Nearly one-quarter of biopharma firms transact through licensing. Consistent with this, Column 3 shows that the majority of license agreements are in the Biopharma (64%) sector followed by IT (15%). Table 2 describes our main sample of firms. The average expenditure on R&D is 11% of sales. Mean sales are $2,260 million and assets are $3,807 million. The price cost margin is 7% on average. Productivity, defined as revenue per employee, is $216 thousand. All of the means are highly skewed by outliers.

[Table 1 & Table 2 about here]

Table 3 reports summary statistics for the patents in our sample. We find a mean market value of $11.45 million and a scientific value of 20 citations, in a given year. The mean TTR and MATTR ranges from 0.52 to 0.93 depending on the measure used. We use the most conservative measure—the MATTR(50)—for our main analyses. We observe 0.1% of patents licensed. We estimate that firms have an average of 38 patents, and 8% of firms license at least one patent. Appendix Table A2 decomposes the summary statistics for our key variables by licensing.

[Table 3 about here]

## 4.2 Examining transferability

We define transferability as the ease with which knowledge embedded in an invention can be moved across firm boundaries. We propose this as a critical notion because it extends tradability. Tradability is thought to be determined by the underlying features of the knowledge embedded in the invention. At opposite extremes, there are low costs associated with exchanging codifiable knowledge compared to relatively high costs associated with exchanging tacit knowledge. This is because codification represents the commodification of knowledge, which allows it to be traded or exchanged, whereas tacit knowledge remains in a form for which trade or exchange is less technically feasible.[21] However, critical to this paper, not all codified knowledge is described in the same way. Instead of using a binary view, we propose a continuous one. We propose that companies may, in fact, influence the degree of transferability of codified knowledge based on their desire to license. This implies that transferability becomes a firm-level choice rather than a function of the inherent characteristics of invention. If this were the case, then we would expect to see transferability of similar underlying inventions vary systematically with likelihood of licensing.

To test this idea, we examine one form of invention that is entirely codifiable—patents. Patents are specified to provide a "description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same" (USPTO, 2024b). In particular, the patent abstract is used to "efficiently serve as a scanning tool for the purposes of searching in the particular art" (USPTO, 2024a), making it especially important for informing outside organizations searching for relevant inventions. In practice, however, many patents use vague or opaque language in their descriptions offering limited technical information that could help predict the ultimate scope of the claims to an outside organization (Datar et al., 2024; Seymore, 2009). Indeed, Anand and Khanna (2000) assert that "patenting activity offers tangible evidence of the ambiguity in description of technologies" (p. 126). We leverage these variations in language—proxied by the MATTR—to measure the degree to which a patent is transferable. Then, we ask if there are systematic differences in the

[21]In a different context, Nelson and Winter (1982) explain that "to be able to do something, and at the same time be unable to explain how it is done, is more than a logical possibility—it is a common situation" (p. 76). They assert that much "operational knowledge remains tacit because it cannot be articulated fast enough, because it is impossible to articulate all that is necessary to a successful performance, and because language cannot simultaneously serve to describe relationships and characterize the things related" (p. 81-82).

MATTR between patents that are licensed and patents that are not.

#### 4.2.1 Empirical challenges

What is the best way to measure the relationship between transferability and likelihood of licensing? In an ideal world, twin patents would exist and would be randomly assigned to companies. The twins would differ on their written descriptiveness: one written such that it is highly transferable and one written such that it is difficult to transfer. We could then directly measure how high transferability affects the likelihood of licensing compared to a twin patent with low transferability. Through randomization, companies would not choose to license based on underlying differences in patents that make them more or less tradable or on firm characteristics (e.g., resources, capabilities, etc.), nor would the licensing decision affect the transferability of the patents, allowing us to cleanly estimate the effect of transferability on licensing. Although randomization solves this sort of identification challenge, it is infeasible in the real world due to an array of social and economic costs. Nonetheless, this thought experiment highlights the major threats to identification our study faces: omitted variable bias, simultaneity, and selection.

#### 4.2.2 Empirical Approach

To address these threats to identification, we apply several complementary approaches: fixed effects, firm controls, and coarsened exact matching (CEM). We include year and industry fixed effects to control for unobserved, time-invariant characteristics that could influence likelihood of licensing, making it more likely that the variation in licensing decisions is driven by transferability rather than static year- or industry-specific factors. We include firm fixed effects in select models. We add firm controls for observable characteristics that could bias the estimate between transferability and likelihood of licensing, such as R&D intensity and firm size. Meanwhile, CEM balances the treatment and control groups on observed covariates, reducing the risk that differences in the decision to license are due to firm characteristics that vary across firms. We perform CEM on firm characteristics, including assets, R&D share, sales, capital stock, operating income, price cost margin, and sector. Balance results for our CEM process are presented in Table 4. By combining a host of fixed effects with firm controls and CEM, we aim to address both unobserved and observed sources of bias, improving the robustness of identification of the relationship, and enabling us to better isolate the role of patent transferability itself. Naturally, our empirical design remains correlational and we refrain from making causal claims.

[Table 4 about here]

Our main specification is represented in Equation 1:

$$\begin{aligned} \text{Likelihood to license}_{itj} = \beta_0 + \beta_1 \text{Transferability}_i \\ + \mathrm{C}_{itj} + \alpha_t + \gamma_j + \epsilon_{itj} \end{aligned} \tag{1}$$

where *Likelihood to license*$_{itj}$ represents the dependent variable, which is the probability that patent $i$ is licensed in year $t$ and industry $j$. *Transferability*$_i$ is proxied by the MATTR of patent $i$. $\mathrm{C}_{itj}$ is a vector of controls that includes firm-level characteristics associated with the firm owning patent $i$ in year $t$ and industry $j$. $\alpha_t$ represents year fixed effects, and $\gamma_j$ represents industry fixed effects. Finally, $\epsilon_{itj}$ is the error term, capturing unobserved factors affecting the likelihood of licensing.

### 4.2.3 MATTR analysis

Does the MATTR vary with likelihood of licensing? Figure 3 presents descriptive results of the distribution of the MATTR across the patents in our sample. The distribution of licensed patents is shifted to the right, meaning that the average MATTR is higher for licensed patents. In order to better isolate the relationship between MATTR and likelihood of licensing, we estimate a logit regression model of likelihood of licensing on the MATTR using our matched sample. We divide the MATTR into 10 quantiles and examine the probability of licensing relative to the median. The results in Figure 4 show that licensing increases linearly with the MATTR. Relative to the median, companies are more likely to license patents that have higher MATTR and less likely to license patents that have lower MATTR. Patents with higher MATTR may have a higher probability of being licensed because they are easier for the licensor to value and describe the range of business applications, and it could be because there are lower search costs and higher gains from trade for the licensee.

[Figure 3 & Figure 4 about here]

Table 5 displays the results from our main specification represented by Equation 1 using linear regression model of likelihood of licensing on MATTR. Columns 1 to 5 estimate a positive and statically significant relationship between the MATTR and likelihood of licensing, ranging from 0.012 to 0.023 (representing a 4 to 8 times increase from the baseline of 0.3%). The magnitude of the coefficient decreases as we include year and industry fixed effects. This decrease demonstrates that the MATTR is unlikely random and that it is influenced, in part, by time and industry trends that are positively correlated with the probability of licensing. These results confirm what we learned from speaking to patent experts: that the way patents are written is influenced by technology, industry, and time related factors. The coefficient on the MATTR decreases in magnitude and loses statistical significance when including firm fixed effects, suggesting that the relationship between MATTR and licensing is primarily influenced by differences across firms rather than within-firm variation. This provides additional evidence that the MATTR is a firm-level strategic choice, influenced by firm-specific factors that also shape licensing behavior. We include controls for R&D share and price cost margin in Appendix Table A3, and the size of the coefficient increases, suggesting that the results in Table 5 represent a lower bound.

[Table 5 about here]

#### 4.2.4 Is it the MATTR or inherently different technology?

Our results show that licensed patents have higher MATTR. We argue this is because companies write their patents differently to influence transferability, and thereby, likelihood of licensing. But it could also be the case that licensed patents have higher MATTR because of their underlying technological nature that makes them easier to describe by design, making them straightforward to license.

We address this concern in two ways. First, we analyze keywords based on the term frequency-inverse document frequency (TF-IDF) for the set of Biopharma patents because they represent a large share of the licensed patents. We do not observe any significant differences in these keywords based on if the patent was licensed or not licensed, suggesting that these patents are comparable based on their substantive content.[22] Second, we re-estimate Equation 1 on a subset of patents that have over 50% similarity based on our TF-IDF text analysis matching. This subset includes approximately 500 patent-pairs with over 50% similarity. This analysis is at the patent-level on the subset without CEM, with robust standard errors clustered by patent-pair and firm. Table 6 shows that when we match on the underlying technology, the coefficients on MATTR significantly increase in magnitude. This is likely a result of reducing noise associated with industry differences in licensing and underlying technological differences that may be downward biasing the estimate in the full sample. As expected, the reduced sample size decreases the power of estimates. Taken together, these two exercises provide additional evidence that the underlying technologies in the patents are not driving the relationship, rather the way that companies choose to describe their patent is what varies and meaningfully influences licensing.

[Table 6 about here]

#### 4.2.5 MATTR and relevance

We next examine how the interaction between the MATTR and our firm-level proxies for relevance influence likelihood of licensing. Based on our discussion in Section 2.3, we expect that companies that specialize in value creation will be more likely to license to other firms

[22] We use the SciSpaCy model "en_core_sci_sm" to identify the keywords and remove common words such as "the", "and", etc., using a list of stop words that comes with the model. We also lemmatize the words, which reduces words to their base form. This means that word derivatives (e.g., "electrical" vs. "electrically") are counted as the same word.

better positioned to exploit the commercial potential of their patent—and that the MATTR should strengthen this relationship. Conversely, we expect that companies better positioned to capture value from their patents in-house will be less likely to license—regardless of the MATTR. Table 7 tests these predictions using linear regression models with multiple levels of fixed effects standard errors clustered by firm. We use the unmatched sample because we are interested in variations in firm-level characteristics that may affect the likelihood of licensing. We proxy for value creation capabilities using R&D intensity, and for value capture capabilities using the price cost margin and sales per employee.

The first column estimates that the interaction between an above median MATTR and R&D intensity is positively correlated with licensing, suggesting that licensing increases in companies that invest in innovation that also make their inventions more transferable. In fact, higher MATTR without high R&D intensity is not associated with greater likelihood of licensing. Consistent with the idea that companies better positioned to internally commercialize their patents are less likely to license, the second and third columns estimate that companies with higher price cost margin and higher productivity are less likely to license patents even if they have higher MATTR. These results suggest that there are important firm-level factors that interact with transferability to influence the likelihood of licensing.

[Table 7 about here]

## 4.3 Toward a mechanism: The ability to predict and capture value, expected payoff, and the likelihood to license

Thus far, we establish that transferability proxied by the MATTR is positively correlated with likelihood of licensing. What remains to be understood is if firms decide to make their invention more transferable based on the expected payoffs of the patent to the firm. Specifically, how is transferability influenced by the ability of the firm that produces the patent to predict value and expected payoffs of an invention? To investigate this, in the following analysis we take advantage of heterogeneity in our sample of firms in two ways. First, we establish that the

MATTR varies with the value of the patent, suggesting that the degree of transferability is related to the value of the patent to the firm. Second, we exploit industrial differences in the ability to discern patent value based on the market uncertainty that firms in those industries face (i.e., whether patent value is observable at an early stage or not). We show that there is only a relationship between the MATTR and patent value for firms in industrial settings facing lower uncertainty—specifically lower market uncertainty, since this will influence value creation and capture potential—suggesting that the ability to predict patent value influences how transferable the firm makes the patent.

#### 4.3.1 Economic and scientific value of patents

We examine how the MATTR varies with two established measures of patent value: private economic value and scientific value of granted patents—which have been shown to be useful proxies for patent value to the firm (Hall et al., 2005; Kogan et al., 2017; Nicholas, 2008). In Figure 5, we observe a positive relationship between the MATTR and private economic value of a firm's patent, suggesting that the MATTR increases with higher economic values. Figure 6 presents an inverse U-shaped relationship between MATTR and scientific value. Patents with moderate scientific value tend to have higher MATTR, while those with both low and high scientific value exhibit lower MATTR values. Given the difference of what these two measures of value proxy, the more precisely estimated relationship with economic value seems reasonable since this measure assesses anticipated value. Appendix Table A4 estimates the relationship between the MATTR and the private economic and scientific value using a linear regression model and confirms that the two measures of patent value are separate constructs since the significance and magnitude of the coefficients remain comparable upon inclusion of both variables across models.

[Figure 5 & Figure 6 about here]

We next seek to assess how patent value interacts with the MATTR to impact the likelihood of licensing.[23] To examine this, we define a dummy variable for if the MATTR is above the median level (= 0.65) and interact it with our proxies for patent value. This allows us to investigate how patent description interacts with patent value and its influence on likelihood of licensing. Table 8, Columns 1 and 2 estimates that as the MATTR increases so does the likelihood of licensing when patent economic value is low, but as patent economic value increases firms are less likely to license. The patents that are "somewhat" economically valuable (between the 25th to 75th percentile of economic values—between \$0.5 to \$10 million) and have above median MATTR are most likely to be licensed. Private economic value represents the expected payoffs of a patent to the firm, and the analysis suggests that companies are reluctant to license patents that are more valuable, regardless of how they are written.

[Table 8 about here]

Columns 3 and 4 perform the same estimation, but instead use scientific value. The likelihood of licensing decreases with the MATTR when scientific value is low. But it increases when MATTR is above median and scientific value is high, and higher MATTR interacted with scientific value produces a larger effect. The interaction is largest for highly scientifically valuable patents (above the 75th percentile representing over 18 citations) and have above median MATTR. As scientific value represents the potential for follow-on invention rooted in a particular patent, a higher MATTR as scientific value increases might suggest that companies are aware that more value can be captured from licensing and IP that builds on their potentially more "foundational" patent(s) relative to exclusive internal commercialization—and therefore write patents to facilitate commercialization through licensing.

#### 4.3.2 Industry heterogeneity

The ability to assess the value of a patent to the firm in large part depends on levels of market uncertainty.[24] Our assumption is that there will always be some underlying source of

[23] Appendix Table A5 shows that the relationship between licensing and economic value and scientific value is nonlinear. Licensing decreases for more economically valuable patents, while it increases for more scientifically valuable patents.

[24] There may be other factors that affect the ability to assess the value of a patent to the firm, such as technological complexity, domain expertise with similar technology, strategic fit with core business, etc. However,

uncertainty, such as the outcome of a given patent commercialization approach for the firm. But certain industries are associated with lower levels of market uncertainty, and therefore, relatively more effective evaluation of value and potential payoffs of a licensing approach to commercialization.

Following Scott et al. (2020), two industries appear useful to contrast: Biopharma and IT. As shown, investors are better able to assess the value of Biopharma ventures than those in IT. In industries like Biopharma, where regulatory approval and clinical trials serve as clear milestones, the valuation of a patent is more predictable, which reduces uncertainty for both licensors and licensees. As a result, Biopharma firms may be better able to estimate future revenues and risks, as well as market demand, associated with the innovation. They may know early enough to manipulate the text of the patents to make those they would like to license out more transferable. In contrast, in the IT sector, the value of an invention is less predictable due to rapidly evolving markets, shorter product life cycles, and the potential for disruptive innovations. This uncertainty may make it more difficult to evaluate a patent, especially early on. Firms might fear an undervaluing or missing out on opportunities by locking in agreements too early. For example, a software company may worry that licensing its technology prematurely could limit its ability to pivot or scale, unlike a Biopharma firm that can rely on more defined market pathways. This unpredictability in IT may lead firms to either delay licensing or seek more flexible, contingent agreements that allow them to retain some control over future developments.

Using this logic, we predict that the expected payoffs from licensing a patent are more easily anticipated at the time the patent application is filed in Biopharma than in IT. As a consequence, we should observe a stronger relationship with our measure of transferability and licensing in Biopharma compared to IT, and this relationship should hinge on the expected value of the invention. Given that the scientific value of an invention is likely to be less conflated by other signals such as the economic value of the invention, we also expect the relationship to be

we argue that demand-driven uncertainty is the key underlying factor that influences the ability to predict value in a way that is particularly meaningful for determining commercialization approach.

less noisy in the former than the latter.

Table 9 estimates the a linear regression model of likelihood of license agreement on MATTR. Column 1 estimates a positive relationship between MATTR and likelihood of licensing. Column 2 shows an even stronger relationship between the MATTR in Biopharma, where relatively lower uncertainty entails more effective evaluation of patent value and optimal commercialization approach. In contrast, Column 3 estimates a smaller positive relationship between the MATTR and likelihood of licensing in IT, where greater uncertainty makes predictions of patent value noisier.[25] All models include time fixed effects and standard errors are clustered at the firm-level.

[Table 9 about here]

#### 4.3.3 Information frictions: Potential boundary conditions in space

Thus far, we argue that firms manipulate information frictions associated with their patents based on the desired commercialization approach. Firms make their patents more transferable if they seek to license. As evidence of this, we estimate that licensing increases with the MATTR—and that this effect varies with patent value and industry. We further test this mechanism by examining how geography moderates the relationship between the MATTR and licensing.

Geographic proximity lowers the cost of transferring specialized knowledge (Alcácer & Chung, 2007; Roche, 2020). Given the ability to interact locally, firms can potentially observe the use of invention and gain insight into its applications before licensing. As a result, we expect the MATTR to be more important for firms with a greater distance between them.

[25] Licensing—as well as the structure of licensing contracts—may in part be contingent on differences in appropriability across industries. For example, Anand and Khanna (2000) find that exclusive licensing is higher in the Chemicals industry, which is characterized by strong intellectual property rights, compared to Computers and Electronics, where intellectual property rights are more difficult to enforce due to the incremental and iterative innovations and the ability to clearly articulate the know-how embodied in the underlying technologies. In our setting, the relationship between the MATTR and intellectual property rights is ambiguous, although we expect that firms might leverage variations in language more in industries with less enforceable intellectual property rights as a way to protect their inventions. However, Table 9 finds a weaker relationship between the MATTR and licensing in IT—an industry characterized by weaker intellectual property rights compared to Biopharma—suggesting that this alternative explanation is unlikely.

Put differently, because these firms face higher information frictions via geographic distance, the MATTR should play a more important role in lowering information costs and facilitating licensing.

Furthermore, there are important differences in information availability across industries (Cohen et al., 2002). In Biopharma, information is relatively standardized and widely shared through the publication of clinical trials, scientific papers, and detailed documentation required by regulatory agencies. As a result, critical knowledge is distributed across companies, regardless of location. In contrast, IT lacks mandatory disclosure of technical documentation, software code, and hardware design, potentially leading to a more uneven distribution of information. This makes proximity more important as a mechanism for acquiring information. In other words, location-based information frictions are likely higher in IT, making the role of the MATTR more pronounced compared to Biopharma.

To examine this, we restrict our sample to the firms that license. We then calculate the distances between licensors and the licensees. In particular, we expect that the MATTR will have less influence in micro-geographies where knowledge is readily spread through word-of-mouth, meetings, and social interactions (Roche et al., 2024)—and for the effect to be stronger in IT. As follows, we construct an indicator for licensor-licensee pairs that are more than 5 kilometers apart, and estimate its relationship with the MATTR. Table 10 presents the estimation results. Consistent with our prediction, the coefficient on the MATTR increases with distance, although it is not statistically significant; it is higher for licensor-licensee pairs that are further apart relative to licensor-licensee pairs that are within 5 kilometers of one another, and the effect is even stronger for firms in the IT industry.

[Table 10 about here]

## 5 Conclusion

This study investigates the strategic determinants of IP licensing decisions in MFT, with a particular focus on the role of transferability—the ability to convert knowledge embodied in an invention across firm boundaries—and its interaction with relevance—the importance of an invention for firm value creation and value capture. Our findings highlight that firms appear to be strategically adjusting the extent to which their patents are transferable, which we proxy using a new measure called MATTR. Higher MATTR scores, indicating increased lexical diversity (in other words, transferability) are positively associated with the likelihood of a patent being licensed. This relationship is most pronounced in Biopharma, an industry with relatively greater predictability in patent value via market demand, underscoring further critical contextual nuances of licensing strategies.

We demonstrate that transferability is likely a strategic choice of reducing information frictions shaped by firms to align with their commercialization goals. The interplay between transferability and relevance reveals a nuanced decision-making process where firms weigh the anticipated value of a patent against competitive risks. The differential impacts of private economic and scientific value on licensing likelihood further emphasize the complex trade-offs firms navigate when determining their optimal invention transfer strategies, and industry heterogeneity analysis reinforces these dynamics. Moreover, the heightened influence of transferability for licensors and licensees that are not in geographic proximity further supports its role in reducing information frictions.

Our findings make at least three contributions. First, we advance the understanding of MFT by identifying the antecedents of licensing decisions, emphasizing the unique strategic implications of licensing (Allain et al., 2016; Hegde, 2014; Hegde & Luo, 2018). Second, we enrich the firm innovation strategy literature (Gans & Stern, 2003) by highlighting supply-side considerations of invention. This oft-overlooked perspective underscores the importance of industry-specific boundary conditions, and the selective nature of IP supplied to these markets. Third, by introducing MATTR as a robust proxy for transferability, our study opens new avenues for exploring the textual and strategic dimensions of patents in invention dissemination and commercialization research (Arora et al., 2018; Arts et al., 2023; Ganglmair et al., 2022; Hoberg & Phillips, 2016; Lee, 2023).

While this study provides novel insights into the antecedents of licensing in MFT, several limitations warrant attention. First, the reliance on MATTR as a proxy for transferability focuses on linguistic diversity and may not fully capture other critical dimensions, such as technological complexity, contextual relevance, or enforcement challenges associated with IP. Second, the study centers exclusively on licensing and does not explore how findings might generalize to other mechanisms of invention transfer, such as outright sales or alliances. Although this may potentially limit the scope of implications, we are able to provide more precise guidance for the decision to license. Third, while fixed effects, firm controls, and coarsened exact matching are employed to address endogeneity concerns, potential unobserved confounders related to firm-specific strategies, industry dynamics, or technological lifecycle stages may still remain. Future work may focus effort in providing more rigorous identification. Finally, our reliance on patent-based data, though extensive, and our focus on U.S. firms, may exclude inventions not formally patented or those with minimal textual descriptions, which could introduce selection bias. This points to the need for further research to validate and expand our insights across different contexts and mechanisms of technology transfer. Particularly the role of micro-geography and institutions appear a fruitful avenue for research (Gross & Roche, 2024; Marinoni & Roche, 2025; Roche et al., 2024). We hope that future work will continue to explore the more dynamic interplay between transferability, complementary assets, and geographic proximity to provide a more holistic understanding of licensing strategies.

## References


Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2005). Competition and Innovation: An Inverted-U Relationship*. *The Quarterly Journal of Economics*, *120*(2), 701–728.

Alcácer, J., & Chung, W. (2007). Location Strategies and Knowledge Spillovers. *Management Science*, *53*(5), 760–776.

Allain, M.-L., Henry, E., & Kyle, M. (2016). Competition and the Efficiency of Markets for Technology. *Management Science*, *62*(4), 1000–1019.

Anand, B. N., & Khanna, T. (2000). The Structure of Licensing Contracts. *The Journal of Industrial Economics*, *48*(1), 103–135.

Arora, A., Belenzon, S., & Patacconi, A. (2018). The decline of science in corporate R&D. *Strategic Management Journal*, *39*(1), 3–32.

Arora, A., Fosfuri, A., & Gambardella, A. (2001). *Markets for technology: The economics of innovation and corporate strategy*. MIT Press.

Arts, S., Cassiman, B., & Hou, J. (2023). Position and Differentiation of Firms in Technology Space. *Management Science*, mnsc.2023.00282.

Arts, S., Flemming, L., & Veretennik, E. (2025). Harnessing academic science for corporate technology: The role of interpersonal networks and brokers. *Working Paper*.

Baker, G. P., Gibbons, R., & Murphy, K. J. (2008). Strategic alliances: Bridges between "islands of conscious power". *Journal of the Japanese and International Economies*, *22*(2), 146–163.

Banerjee, T., & Siebert, R. (2017). Dynamic impact of uncertainty on R&D cooperation formation and research performance: Evidence from the bio-pharmaceutical industry. *Research Policy*, *46*(7), 1255–1271.

Barney, J. B. (1986). Strategic Factor Markets: Expectations, Luck, and Business Strategy. *Management Science (1986-1998)*, *32*(10), 1231–1241.

Basir, N., Beyhaghi, M., & Mohammadi, A. (2012). The Fate of Patents: An Exploratory Analysis of Patents as IPO Signals of Reputational Advantage. *Academy of Management Proceedings*, *2012*(1), 11689.

Bearson, D. (2025). Keep, license, sell: Examining the antecedents of strategic intellectual property transfer. *Working Paper*.

Birhanu, A., & Gambardella, A. (2024). To commercialize inside or outside of the firm: Behavioral considerations in patent exploitation by family firms. *Strategic Management Journal*, *45*(5), 875–901.

Capron, L., & Mitchell, W. (2012). *Build, borrow, or buy: Solving the growth dilemma*. Harvard Business Review Press.

Cassiman, B., & Veugelers, R. (2006). In Search of Complementarity in Innovation Strategy: Internal R&D and External Knowledge Acquisition. *Management Science*, *52*(1), 68–82.

Cohen, W., Nelson, R., & Walsh, J. (2000, February). *Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)* (tech. rep. No. w7552). National Bureau of Economic Research. Cambridge, MA.

Cohen, W., Nelson, R., & Walsh, J. (2002). The influence of public research on industrial rd. *Management Science*, 1–23.

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian Knot: The Moving-Average Type–Token Ratio (MATTR). *Journal of Quantitative Linguistics*, *17*(2), 94–100.

Datar, A., Amore, M. D., & Fosfuri, A. (2024). Express: Strategic patent disclosure: Unraveling the influence of temporal preferences. *Strategic Organization*.

de Bettignies, J.-E., Liu, H. F., Robinson, D. T., & Gainulline, B. (2023). Competition and Innovation in Markets for Technology. *Management Science*, *69*(8), 4753–4773.

Degnan, S. A. (1999). The Licensing Payoff from U.S. R&D. *Research-Technology Management*, *42*(2), 22–25.

Dierickx, I., & Cool, K. (1989). Asset Stock Accumulation and Sustainability of Competitive Advantage. *Management Science*, *35*(12), 1504–1511.

Eklund, J., & Kapoor, R. (2022). Mind the gaps: How organization design shapes the sourcing of inventions. *Organization Science*, 1251–1699.

Farre-Mensa, J., Hegde, D., & Ljungqvist, A. (2020). What Is a Patent Worth? Evidence from the U.S. Patent "Lottery". *The Journal of Finance*, *75*(2), 639–682.

Galasso, A., Schankerman, M., & Serrano, C. J. (2013). Trading and enforcing patent rights. *The RAND Journal of Economics*, *44*(2), 275–312.

Ganglmair, B., Robinson, W. K., & Seeligson, M. (2022). The Rise of Process Claims: Evidence From a Century of U.S. Patents.

Gans, J. S., & Stern, S. (2003). The product market and the market for "ideas": Commercialization strategies for technology entrepreneurs. *Research Policy*.

Grant, R. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, *17*, 109–122.

Grindley, P. C., & Teece, D. J. (1997). Managing intellectual capital: Licensing and cross-licensing in semiconductors and electronics. *California Management Review*, *39*(2), 8–41.

Gross, D. P., & Roche, M. P. (2024). Coordinated R&D programs and the creation of new industries. *Harvard Business School Strategy Unit Working Paper*, (24-027), 24–027.

Guo, W., Yu, T., & Gimeno, J. (2017). Language and competition: Communication vagueness, interpretation difficulties, and market entry. *Academy of Management Journal*, *60*, 2073–2098.

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *The Rand Journal of Economics*, *36*(1), 16–38.

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), 101–128.

Haucap, J., Rasch, A., & Stiebale, J. (2019). How mergers affect innovation: Theory and evidence. *International Journal of Industrial Organization*, *63*, 283–325.

Hegde, D. (2014). Tacit Knowledge and the Structure of License Contracts: Evidence from the Biomedical Industry. *Journal of Economics & Management Strategy*, *23*(3), 568–600.

Hegde, D., & Luo, H. (2018). Patent Publication and the Market for Ideas. *Management Science*, *64*(2), 652–672.

Hengel, E. (2022). Publishing while female: Are women held to higher standards? evidence from peer review. *The Economic Journal*, 2951–2991.

Hoang, H., & Rothaermel, F. (2010). Leveraging internal and external experience: Exploration, exploitation, and rd project performance. *Strategic Management Journal*, 734–758.

Hoberg, G., & Phillips, G. (2016). Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy*, *124*(5), 1423–1465.

Hsu, D. H., & Ziedonis, R. H. (2013). Resources as dual sources of advantage: Implications for valuing entrepreneurial-firm patents. *Strategic Management Journal*, *34*(7), 761–781.

Kankanhalli, G., Kwan, A., & Merkley, K. (2024). The Paradox of Innovation Nondisclosure: Evidence from Licensing Contracts. *American Economic Journal: Applied Economics*, *16*(4), 220–256.

Kapacinskaite, A., Nandakumar Vrajesh, R., & Cunningham, C. Markets for trade secrets. In: In *Strategic management society 44th annual conference*. 2024.

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

Kong, N., Dulleck, U., Jaffe, A., Sun, S., & Vajjala, S. (2023). Linguistic metrics for patent disclosure: Evidence from university versus corporate patents. *Research Policy*, *52*(2).

Kwon, J. H., Park, H. D., & Deng, S. (2022). When Do Firms Trade Patents? *Organization Science*, *33*(3), 1212–1231.

Kyle, K. (2019). Measuring Lexical Richness. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (1st ed., pp. 454–476). Routledge.

Lee, H. (2023). The heterogeneous effects of patent scope on licensing propensity. *Research Policy*, *52*(3), 104696.

Lerner, J., Manley, H. J., Stein, C., & Williams, H. L. (2024, January). The Wandering Scholars: Understanding the Heterogeneity of University Commercialization.

Makadok, R., & Barney, J. B. (2001). Strategic Factor Market Intelligence: An Application of Information Economics to Strategy Formulation and Competitor Intelligence. *Management Science*, *47*(12), 1621–1638.

Marinoni, A., & Roche, M. P. (2025). You've Got Mail! The Late 19th Century US Postal Service Expansion, Firm Creation, and Firm Performance. *Management Science*.

McCarthy, P., & Jarvis, S. (2007). Vocd: A theoretical and empirical evaluation. *Language Testing*, *24*(4), 459–488.

Moore, K. (2005). Worthless patents. *Berkeley Technology Law Journal*, 1521–1552.

Moreira, S., Klueter, T. M., & Tasselli, S. (2020). Competition, Technology Licensing-in, and Innovation. *Organization Science*, *31*(4), 1012–1036.

MSCI. (2022). The Regional Breakdown of Market Sectors Over Time.

Munter, P. (2022). Assessing Materiality: Focusing on the Reasonable Investor When Evaluating Errors.

Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Belknap Press of Harvard University Press.

Nicholas, T. (2008). Does Innovation Cause Stock Market Runups? Evidence from the Great Crash. *American Economic Review*, *98*(4), 1370–1396.

Roche, M. P. (2020). Taking Innovation to the Streets: Microgeography, Physical Structure, and Innovation. *The Review of Economics and Statistics*, *102*(5), 912–928.

Roche, M. P., Oettl, A., & Catalini, C. (2024). (Co-)Working in Close Proximity: Knowledge Spillovers and Social Interactions. *Management Science*.

Sadao, N., & Walsh, J. P. (2009). Commercialization and Other Uses of Patents in Japan and the U.S.: Major findings from the RIETI-Georgia Tech inventor survey. *IDEAS Working Paper Series from RePEc*.

Schumpeter, J. A. (1934). *The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle*. Transaction Books.

Scott, E. L., Shu, P., & Lubynsky, R. M. (2020). Entrepreneurial Uncertainty and Expert Evaluation: An Empirical Analysis. *Management Science*, *66*(3), 1278–1299.

Serrano, C. J. (2010). The dynamics of the transfer and renewal of patents. *The RAND Journal of Economics*, *41*(4), 686–708.

Serrano, C. J. (2018). Estimating the Gains from Trade in the Market for Patent Rights. *International Economic Review*, *59*(4), 1877–1904.

Seymore, S. B. (2009). The Teaching Function of Patents. *Notre Dame Law Review*, *85*(2), 621–670.

Shaib, C., Barrow, J., Sun, J., Siu, A. F., Wallace, B. C., & Nenkova, A. (2024). Standardizing the Measurement of Text Diversity: A Tool and a Comparative Analysis of Scores.

Shapiro, C. (2000). Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting. *Innovation Policy and the Economy*, *1*, 119–150.

Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy*, *15*(6), 285–305.

Teece, D. J. (2006). Reflections on "Profiting from Innovation". *Research Policy*, *35*(8), 1131–1146.

Toh, P. K., & Miller, C. D. (2017). Pawn to Save a Chariot, or Drawbridge Into the Fort? Firms' Disclosure During Standard Setting and Complementary Technologies Within Ecosystems. *Strategic Management Journal*, *38*(11), 2213–2236.

Torrisi, S., Gambardella, A., Giuri, P., Harhoff, D., Hoisl, K., & Mariani, M. (2016). Used, blocking and sleeping patents: Empirical evidence from a large-scale inventor survey. *Research Policy*, *45*(7), 1374–1385.

USPTO. (2024a). 37 CFR 1.438 The abstract.

USPTO. (2024b). Manual of Patent Examining Procedure.

Vorhies, J. (2005). The New Importance of Materiality. *Journal of Accountancy*.

Zhao, M. (2006). Conducting R&D in Countries with Weak Intellectual Property Rights Protection. *Management Science*, *52*(8), 1185–1199.

# Figures

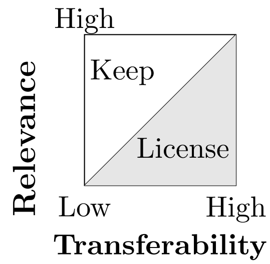


Figure 1: Conceptual framework

**Patent No. 5,554,453:**

*A fuel cell system employing a gasifier for generating fuel gas for the fuel cell of the fuel cell system and in which heat for the gasifier is derived from the anode exhaust gas of the fuel cell.*

$$\mathrm{TTR} = \frac{\#types}{\#tokens} = \frac{20}{38} \approx 0.526$$

$$\mathrm{MATTR}(L) = \frac{1}{N - L + 1} \sum_{i=1}^{N-L+1} \mathrm{TTR}_i$$

$$\mathrm{MATTR}(30) = \frac{1}{38 - 30 + 1} \sum_{i=1}^{9} \mathrm{TTR}_i$$

$$\mathrm{MATTR}(30) = \frac{1}{9} \begin{pmatrix} 0.60 + 0.60 + 0.63 + 0.67 + 0.67 + \\ 0.63 + 0.60 + 0.60 + 0.60 \end{pmatrix} \approx 0.622$$

Figure 2: Example calculation of the Type-Token Ratio (TTR) and the Moving Average Type-Token Ratio (MATTR) on patent abstract

*Notes:* $N$ is the total number of tokens in the text. $L$ is the fixed window size (i.e., the number of words in each sliding window). $i$ is index of each sliding window, ranging from 1 to $N - L + 1$, which is the total number of windows. $\mathrm{TTR}_i$ is the share of unique words in window $i$ of size $L$. We calculate the TTR, MATTR(10), MATTR(30), and MATTR(50) for all of the patent abstracts in our sample. The TTR is the simplest measure, but may be biased by the length of the text. The MATTR corrects for this by calculating the TTR incrementally over a smooth moving range.

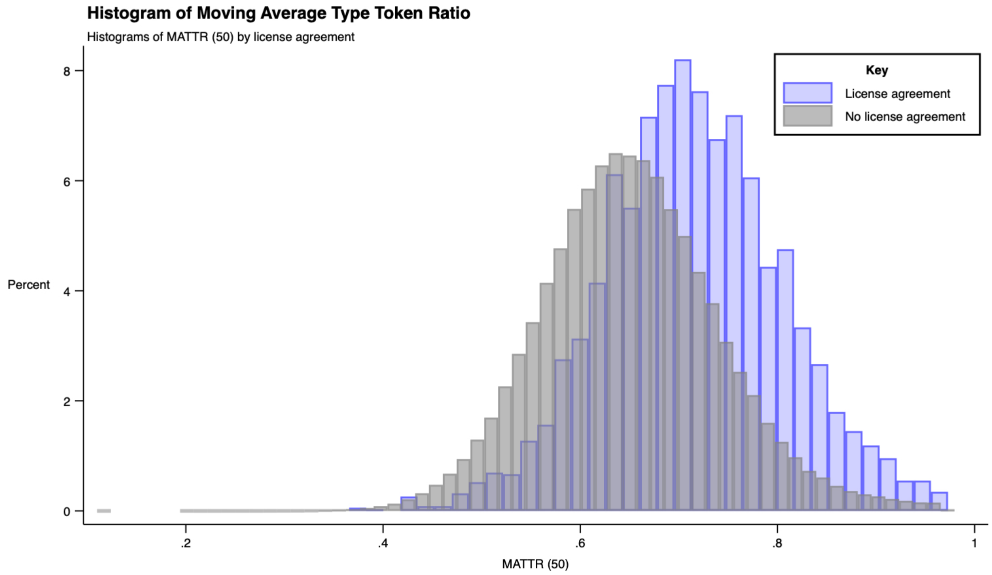


Figure 3: Histogram of moving-average type-token ratio

*Notes:* This figure illustrates the distribution of the MATTR (50) in our sample. The unit of observation is the patent. A MATTR is calculated for each patent based on text analysis of the abstract and ranges from 0 to 1.

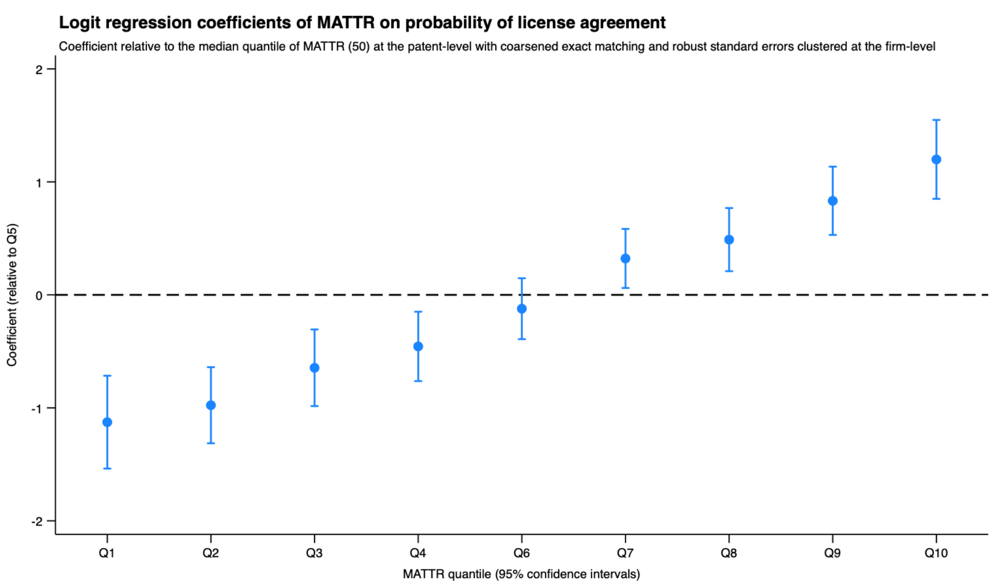


Figure 4: Logit regression coefficients of MATTR on probability of license agreement

*Notes:* This figure plots coefficient estimates and 95% confidence intervals from a logit regression in which the dependent variable indicates whether a patent was licensed. The unit of observation is the patent. We use our matched sample and standard errors are robust and clustered by firm. The coefficients plotted consist of the relationship between patents at a given quantile relative to the median quantile. There are 10 quantiles total. All coefficients are estimated from the same regression.

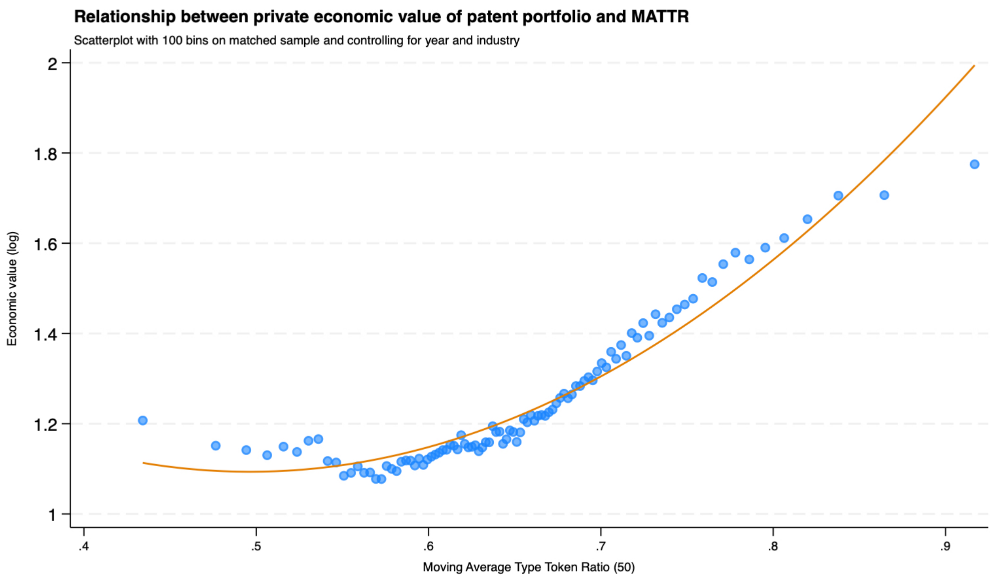


Figure 5: MATTR increases with economic value of the patent

*Notes:* This figure is a binned scatterplot of the relationship between the log economic value of patents and the MATTR. The unit of observation is the patent. We use the matched sample and control for year and industry. There are 100 bins. The orange line is the quadratic fit between log economic value and the MATTR.

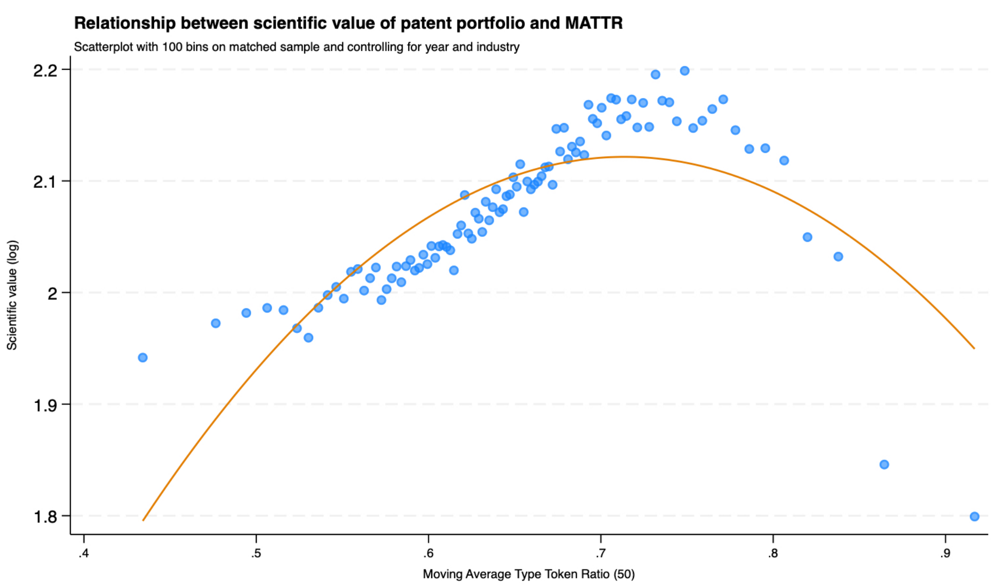


Figure 6: MATTR rises with scientific value of the patent and then declines

*Notes:* This figure is a binned scatterplot of the relationship between the log scientific value of patents and the MATTR. The unit of observation is the patent. We use the matched sample and control for year and industry. There are 100 bins. The orange line is the quadratic fit between log scientific value and the MATTR.

# Tables

Table 1: Industrial composition of the sample

| | Obs. | (1) | (2) | (3) |
|---|---|---|---|---|
| Energy | 245 | 3.29 | 4.90 | 2.01 |
| Materials | 498 | 6.68 | 4.62 | 3.86 |
| Industrials | 1197 | 16.06 | 3.34 | 6.71 |
| Consumer discretionary | 923 | 12.38 | 1.95 | 3.02 |
| Consumer staples | 261 | 3.50 | 6.90 | 3.02 |
| Health care | 1652 | 22.16 | 23.24 | 64.43 |
| Financials | 268 | 3.59 | 1.49 | 0.67 |
| IT | 2075 | 27.83 | 4.43 | 15.44 |
| Communication services | 183 | 2.45 | 2.73 | 0.84 |
| Utilities | 126 | 1.69 | 0.00 | 0.00 |
| Real estate | 27 | 0.36 | 0.00 | 0.00 |

*Notes:* Compustat classification based on 11 sectors in the Global Industry Classification Standard. Column (1) is the total number of firms in a given industry divided by the total number of firms in the panel (i.e., sample composition). Column (2) is the number of licensing firms in the sector divided by the total number of firms in that sector (i.e., share of licensing firms by sector). Column (3) is the number of firms that license in the sector divided by the total number of licensing firms (i.e., licensing contribution by sector). Column (1) and (3) sum to 100%.

Table 2: Summary statistics at the firm-level

| | Obs. | Mean | SD | Min | P10 | P50 | P90 | Max |
|---|---|---|---|---|---|---|---|---|
| R&D (% of sales) | 5500 | 0.11 | 0.09 | 0.00 | 0.01 | 0.09 | 0.25 | 0.25 |
| Sales (mil.) | 7211 | 2260.17 | 9244.09 | 6.62 | 6.62 | 128.38 | 3540.04 | 119656.00 |
| Assets (mil.) | 7214 | 3806.90 | 17221.99 | 17.21 | 17.21 | 152.25 | 4724.52 | 175889.00 |
| Price cost margin | 7193 | 0.07 | 0.08 | 0.00 | 0.00 | 0.05 | 0.17 | 0.39 |
| Productivity | 7103 | 216.32 | 176.64 | 46.46 | 54.53 | 161.32 | 464.96 | 784.29 |

*Notes:* This table reports summary statistics for the key variables in our empirical tests. We calculate the mean values of the variables at the firm-level. The number of observations varies based on data availability for each measure. Productivity is measured as revenue per employee. Each variable is winsorized.

Table 3: Summary statistics for key variables at the patent-level

| | Obs. | Mean | SD | Min | P10 | P50 | P90 | Max |
|---|---|---|---|---|---|---|---|---|
| Market value (mil.) | 1764463 | 11.45 | 33.22 | 0.00 | 0.07 | 3.61 | 25.36 | 2686.26 |
| Scientific value (cites) | 1764463 | 20.18 | 61.81 | 0.00 | 0.00 | 6.00 | 44.00 | 4974.00 |
| TTR | 1763379 | 0.52 | 0.13 | 0.07 | 0.36 | 0.50 | 0.70 | 0.98 |
| MATTR (50) | 1763379 | 0.65 | 0.09 | 0.11 | 0.54 | 0.65 | 0.77 | 0.98 |
| MATTR (30) | 1763379 | 0.75 | 0.07 | 0.15 | 0.66 | 0.75 | 0.84 | 1.00 |
| MATTR (10) | 1763379 | 0.93 | 0.03 | 0.26 | 0.89 | 0.93 | 0.96 | 1.00 |
| License agreement | 1764463 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |

*Notes:* This table reports summary statistics for the key variables in our empirical tests. We calculate the mean values of the variables at the patent-level. The number of observations varies based on data availability for each measure.

Table 4: Pre- and post-CEM covariate balance

| | **Unmatched Sample** | | | **Matched Sample** | | |
|---|---|---|---|---|---|---|
| | Mean (Licensed) | Mean (Unlicensed) | Diff | Mean (Licensed) | Mean (Unlicensed) | Diff |
| Price cost margin | 0.0545 | 0.0694 | -0.0149*** | 0.0449 | 0.0437 | 0.0012 |
| Sales | 5105.16 | 1821.14 | 3284.01*** | 2904.25 | 2880.77 | 23.48 |
| Assets | 6981.71 | 3190.08 | 3791.63*** | 3581.45 | 3363.81 | 217.64 |
| Sales per employee | 213.09 | 211.48 | 1.60 | 198.57 | 188.16 | 10.40 |
| Capital stock | 207.03 | 75.19 | 131.84*** | 116.11 | 119.83 | -3.72 |
| Operating income | 650.55 | 215.69 | 434.85*** | 285.64 | 245.90 | 39.74 |
| R&D (% of sales) | 0.1679 | 0.0998 | 0.0681*** | 0.1732 | 0.1632 | 0.0101** |
| Observations | 7728 | | | 6051 | | |

Significance levels: * p<0.1; ** p<0.05; *** p<0.01.
*Notes:* We match on firm characteristics between firms that license and firms that do not license. These characteristics include the price cost margin, sales, assets, capital stock, operating income, R&D expenditure, and sector. Mean differences calculated by subtracting the mean value for non-licensing firms from the licensing firms (i.e., licensed mean - non-licensed mean ). The mean difference will be negative if the licensed mean is lower than the non-licensed mean. Robust standard errors used to calculate p-values.

Table 5: Linear regression results for license agreement on MATTR

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| MATTR | 0.022*** | 0.023*** | 0.013*** | 0.012*** | 0.012*** |
| | (0.003) | (0.003) | (0.003) | (0.002) | (0.002) |
| Year FEs | No | Yes | No | Yes | Yes |
| Industry FEs | No | No | Yes | Yes | Yes |
| Year × Industry FE | No | No | No | No | Yes |
| R-Sq | 0.00 | 0.00 | 0.01 | 0.02 | 0.02 |
| Observations | 1010943 | 1010943 | 1007930 | 1007930 | 1007928 |

Significance levels: * p<0.1; ** p<0.05; *** p<0.01

*Notes:* This table reports estimation results of likelihood of licensing on MATTR using linear models with different levels of fixed effects and firm controls. The dependent variable is an indicator equal to 1 if the patent was licensed and 0 otherwise. The unit of observation is the patent. The number of observations varies based on data availability. The analysis is performed on the matched sample. Robust standard errors in parentheses clustered by firm.

Table 6: Linear regression results for license agreement on set of patents with over 50% similarity

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| MATTR | 1.317** | 1.039* | 1.000* | 0.963* |
| | (0.538) | (0.564) | (0.527) | (0.554) |
| Patent-Pair FEs | Yes | Yes | Yes | Yes |
| Year FEs | No | Yes | No | Yes |
| Industry FEs | No | No | Yes | Yes |
| R-Sq | 0.02 | 0.31 | 0.14 | 0.39 |
| Observations | 514 | 508 | 510 | 506 |

Significance levels: * p<0.1; ** p<0.05; *** p<0.01

*Notes:* This table reports estimation results of likelihood of licensing on MATTR on patent-pairs that have over 50% similarity based on text analysis. The dependent variable is an indicator equal to 1 if the patent was licensed and 0 otherwise. The sample is not matched. The unit of observation is the patent. Robust standard errors in parentheses clustered by patent-pair and firm.

Table 7: Linear regression results for license agreement on MATTR interacted with relevance proxies

| | (1) | (2) | (3) |
|---|---|---|---|
| MATTR (> med.) | -0.003*** | 0.003*** | 0.017*** |
| | (0.000) | (0.000) | (0.003) |
| R&D intensity | 0.027*** | | |
| | (0.006) | | |
| Price cost margin | | -0.011*** | |
| | | (0.003) | |
| Sales per emp. | | | -0.003*** |
| | | | (0.000) |
| MATTR (> med.) x R&D intensity | 0.039*** | | |
| | (0.005) | | |
| MATTR (> med.) x Price cost margin | | -0.018*** | |
| | | (0.003) | |
| MATTR (> med.) x Sales per emp. | | | -0.003*** |
| | | | (0.000) |
| Year FEs | Yes | Yes | Yes |
| Industry FEs | Yes | Yes | Yes |
| Year × Industry FE | Yes | Yes | Yes |
| R-Sq | 0.02 | 0.02 | 0.02 |
| Observations | 1698567 | 1758306 | 1747347 |

Significance levels: * p<0.1; ** p<0.05; *** p<0.01

*Notes:* This table reports estimation results of likelihood of licensing on MATTR using linear models with different levels of fixed effects. The dependent variable is an indicator equal to 1 if the patent was licensed and 0 otherwise. The unit of observation is the patent. The number of observations varies based on data availability. Robust standard errors in parentheses clustered by firm.

Table 8: Linear regression results for license agreement on MATTR x value proxies

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| MATTR (> med.) | 0.001*** | 0.001*** | -0.001* | -0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| EV (25-75) | 0.000 | 0.000 | | |
| | (0.001) | (0.000) | | |
| EV (>75) | -0.002** | -0.002** | | |
| | (0.001) | (0.001) | | |
| MATTR (> med.) x EV (25-75) | 0.001* | 0.001* | | |
| | (0.001) | (0.001) | | |
| MATTR (> med.) x EV (>75) | -0.001** | -0.002** | | |
| | (0.001) | (0.001) | | |
| SV (25-75) | | | -0.000* | 0.000** |
| | | | (0.000) | (0.000) |
| SV (>75) | | | -0.000 | 0.001 |
| | | | (0.001) | (0.000) |
| MATTR (> med.) x SV (25-75) | | | 0.001*** | 0.001*** |
| | | | (0.000) | (0.000) |
| MATTR (> med.) x SV (>75) | | | 0.006*** | 0.005*** |
| | | | (0.001) | (0.001) |
| Year FEs | Yes | Yes | Yes | Yes |
| Industry FEs | Yes | Yes | Yes | Yes |
| Year × Industry FE | No | Yes | No | Yes |
| R-Sq | 0.02 | 0.02 | 0.02 | 0.02 |
| Observations | 1007930 | 1007928 | 1007930 | 1007928 |

Significance levels: * p<0.1; ** p<0.05; *** p<0.01

*Notes:* This table reports estimation results of likelihood of licensing on MATTR interacted with value proxies using linear models with different levels of fixed effects. The dependent variable is an indicator equal to 1 if the patent was licensed and 0 otherwise. The omitted category is patents below the 25th percentile of economic or scientific value. EV is economic value and SV is scientific value. The unit of observation is the patent. The number of observations varies based on data availability. The analysis is performed on the matched sample. Robust standard errors in parentheses clustered by firm.

Table 9: Linear regression results for license agreement on MATTR by industry

| | All | Biopharma | IT |
|---|---|---|---|
| | (1) | (2) | (3) |
| MATTR | 0.012*** | 0.069*** | 0.003*** |
| | (0.002) | (0.013) | (0.001) |
| Year FEs | Yes | Yes | Yes |
| Industry FEs | Yes | No | No |
| Year × Industry FE | Yes | No | No |
| R-Sq | 0.02 | 0.01 | 0.00 |
| Observations | 1007928 | 103586 | 399409 |

Significance levels: * p<0.1; ** p<0.05; *** p<0.01

*Notes:* This table reports estimation results of likelihood of licensing on MATTR using linear models with different levels of fixed effects overall, in Biopharma, and IT. The dependent variable is an indicator equal to 1 if the patent was licensed and 0 otherwise. The unit of observation is the patent. The number of observations varies based on data availability. The analysis is performed on the matched sample. Robust standard errors in parentheses clustered by firm.

Table 10: Linear regression results for MATTR (> med.) on proximity

| | All | Biopharma | IT |
|---|---|---|---|
| | (1) | (2) | (3) |
| Distance >= 5 km | 0.023 | 0.102 | 0.087** |
| | (0.101) | (0.103) | (0.036) |
| Firm FE | Yes | Yes | Yes |
| R-Sq | 0.46 | 0.40 | 0.47 |
| Observations | 680 | 533 | 75 |

Significance levels: * p<0.1; ** p<0.05; *** p<0.01

*Notes:* This table reports estimation results of MATTR on distance overall, in Biopharma, and IT. We create an indicator for distance between a licensor and licensee that is greater than or equal to 5 kilometers (i.e., firms that are not within the same microgeography). We restrict to companies that have at least two license agreements and compare the relationship between MATTR (> med.) and distance within a firm. The unit of analysis is the license agreement. The number of observations varies based on data availability. Robust standard errors in parentheses clustered by license agreement.

# Appendix for:

# Transferability MATTRs: Towards Understanding Antecedents of Strategic Licensing

Dafna F. Bearson and Maria P. Roche

# Tables

Table A1: Variable Descriptions

| Variable | Description | Source |
|---|---|---|
| **Invention** | | |
| *Patents* | The patents owned by public firms in the U.S. from 1980-2015. | Kogan et al. (2017). |
| *Licensed patents* | An indicator equal to one if the patent is licensed by the firm based on company 10-K or 8-K filings to the U.S. SEC from 1995 to 2015. | RoyaltyStat. |
| **Relevance** | | |
| *Value creation measures* | | |
| *Scientific value* | The total number of forward citations that a patent receives. | Kogan et al. (2017). |
| *Market value* | The movements in stock prices in a three day "annoucement window" following a patent being granted to a firm. | Kogan et al. (2017). |
| *R&D expenditure as a % of sales* | All costs incurred during the year that relate to the development of new products or services divided by gross sales. | Compustat. |
| *Value capture measures* | | |
| *Price cost margin* | The operating profits net of depreciation, provisions less financial cost of capital divided by gross sales. The cost of capital is assumed to be 0.085 for all firms and time periods, following Aghion et al. (2005). | Compustat. |
| **Transferability** | | |
| *Type token ratio (TTR)* | The number of unique words divided by the total number of words in a patent abstract. | PatStat. |
| *Moving average type token ratio (MATTR)* | The number of unique words divided by the total number of words in a patent abstract, adjusted for length of the patent over smooth moving ranges of 10, 30, and 50. | PatStat. |
| *Geographic proximity* | An indicator for if the licensor-licensee pairs are within 5 kilometers of one another. | Compustat; Google API; manual search |

Table A2: Summary statistics for key variables at the patent-level

| | Obs. | Mean | SD | Min | P10 | P50 | P90 | Max |
|---|---|---|---|---|---|---|---|---|
| **Non-licensed patent** | | | | | | | | |
| Market value (mil.) | 1761009 | 11.45 | 33.22 | 0.00 | 0.06 | 3.61 | 25.37 | 2686.26 |
| Scientific value (cites) | 1761009 | 20.05 | 61.13 | 0.00 | 0.00 | 6.00 | 44.00 | 3222.00 |
| TTR | 1759930 | 0.52 | 0.13 | 0.07 | 0.36 | 0.50 | 0.70 | 0.98 |
| MATTR (50) | 1759930 | 0.65 | 0.09 | 0.11 | 0.54 | 0.65 | 0.77 | 0.98 |
| MATTR (30) | 1759930 | 0.75 | 0.07 | 0.15 | 0.66 | 0.75 | 0.84 | 1.00 |
| MATTR (10) | 1759930 | 0.93 | 0.03 | 0.26 | 0.89 | 0.93 | 0.96 | 1.00 |
| **Licensed patent** | | | | | | | | |
| Market value (mil.) | 3454 | 10.01 | 31.36 | 0.01 | 0.33 | 2.02 | 21.84 | 569.48 |
| Scientific value (cites) | 3454 | 85.16 | 204.63 | 0.00 | 2.00 | 28.00 | 201.00 | 4974.00 |
| TTR | 3449 | 0.62 | 0.14 | 0.13 | 0.43 | 0.61 | 0.82 | 0.97 |
| MATTR (50) | 3449 | 0.71 | 0.09 | 0.37 | 0.60 | 0.71 | 0.83 | 0.97 |
| MATTR (30) | 3449 | 0.80 | 0.07 | 0.43 | 0.71 | 0.80 | 0.88 | 0.99 |
| MATTR (10) | 3449 | 0.94 | 0.03 | 0.60 | 0.90 | 0.94 | 0.98 | 1.00 |
| **Total** | | | | | | | | |
| Market value (mil.) | 1764463 | 11.45 | 33.22 | 0.00 | 0.07 | 3.61 | 25.36 | 2686.26 |
| Scientific value (cites) | 1764463 | 20.18 | 61.81 | 0.00 | 0.00 | 6.00 | 44.00 | 4974.00 |
| TTR | 1763379 | 0.52 | 0.13 | 0.07 | 0.36 | 0.50 | 0.70 | 0.98 |
| MATTR (50) | 1763379 | 0.65 | 0.09 | 0.11 | 0.54 | 0.65 | 0.77 | 0.98 |
| MATTR (30) | 1763379 | 0.75 | 0.07 | 0.15 | 0.66 | 0.75 | 0.84 | 1.00 |
| MATTR (10) | 1763379 | 0.93 | 0.03 | 0.26 | 0.89 | 0.93 | 0.96 | 1.00 |

*Notes:* This table reports summary statistics for the key variables in our empirical tests. We calculate the mean values of the variables at the patent-level. The number of observations varies based on data availability for each measure.

Table A3: Linear regression results for license agreement on MATTR with controls

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| MATTR | 0.025*** | 0.022*** | 0.012*** | 0.009*** | 0.010*** |
| | (0.003) | (0.003) | (0.002) | (0.002) | (0.002) |
| R&D (% of sales) | 0.078*** | 0.081*** | 0.074*** | 0.077*** | 0.071*** |
| | (0.011) | (0.012) | (0.012) | (0.012) | (0.012) |
| Price cost margin | -0.024*** | -0.023*** | -0.037*** | -0.036*** | -0.037*** |
| | (0.005) | (0.004) | (0.007) | (0.007) | (0.007) |
| Year FEs | No | Yes | No | Yes | Yes |
| Industry FEs | No | No | Yes | Yes | Yes |
| Year × Industry FE | No | No | No | No | Yes |
| R-Sq | 0.01 | 0.01 | 0.02 | 0.02 | 0.03 |
| Observations | 983211 | 983211 | 980565 | 980565 | 980562 |

Significance levels: * p<0.1; ** p<0.05; *** p<0.01

*Notes:* This table reports estimation results of likelihood of licensing on MATTR using linear models with different levels of fixed effects and firm controls. The dependent variable is an indicator equal to 1 if the patent was licensed and 0 otherwise. The unit of observation is the patent. The number of observations varies based on data availability. The analysis is performed on the matched sample. Robust standard errors in parentheses clustered by firm.

Table A4: Linear regression of MATTR (> med.) on patent value

| | (1) | (2) | (3) |
|---|---|---|---|
| EV (log) | 0.023*** | | 0.022*** |
| | (0.008) | | (0.008) |
| SV (log) | | 0.011*** | 0.009*** |
| | | (0.002) | (0.002) |
| Year FEs | Yes | Yes | Yes |
| Industry FEs | Yes | Yes | Yes |
| R-Sq | 0.09 | 0.09 | 0.09 |
| Observations | 1007930 | 1007930 | 1007930 |

Significance levels: * p<0.1; ** p<0.05; *** p<0.01

*Notes:* This table reports estimation results of MATTR (> med.) on economic value and scientific value. Economic value and scientific value are denoted EV and SV respectively. The dependent variable is an indicator equal to 1 if the MATTR is above median and 0 otherwise. The unit of observation is the patent. The number of observations varies based on data availability for each measure. The analysis is performed on the matched sample. Robust standard errors in parentheses clustered by firm.

Table A5: Linear regression results for license agreement

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| EV (log) | -0.001*** | 0.001 | | | 0.000 |
| | (0.000) | (0.001) | | | (0.001) |
| $EV^2$ | | -0.001*** | | | -0.000*** |
| | | (0.000) | | | (0.000) |
| SV (log) | | | 0.001*** | -0.002*** | -0.002*** |
| | | | (0.000) | (0.001) | (0.001) |
| $SV^2$ | | | | 0.001*** | 0.001*** |
| | | | | (0.000) | (0.000) |
| Year FEs | Yes | Yes | Yes | Yes | Yes |
| Industry FEs | Yes | Yes | Yes | Yes | Yes |
| R-Sq | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| Observations | 1008708 | 1008708 | 1008708 | 1008708 | 1008708 |

Significance levels: * p<0.1; ** p<0.05; *** p<0.01

*Notes:* This table reports estimation results of likelihood of licensing on economic and scientific value using linear models with different levels of fixed effects. Economic value and scientific value are denoted EV and SV respectively. The dependent variable is an indicator equal to 1 if the patent was licensed and 0 otherwise. The unit of observation is the patent. The number of observations varies based on data availability for each measure. The analysis is performed on the matched sample. Robust standard errors in parentheses clustered by firm.