Essays on Corporate Investment in Scientific Research

by

Lia Sheer

Business Administration
Duke University

Date: ____________________
Approved:

____
Sharon Belenzon, Advisor

____
Ashish Arora

____
Wesley Cohen

____
Alon Brav

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Business Administration
in the Graduate School of Duke University
2021

Abstract

Essays on Corporate Investment in Scientific Research

by

Lia Sheer

Business Administration
Duke University

Date: ____________________
Approved:

____
Sharon Belenzon, Advisor

____
Ashish Arora

____
Wesley Cohen

____
Alon Brav

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Business Administration
in the Graduate School of Duke University
2021

Abstract

In light of a reduction in corporate scientific research in recent decades, my dissertation examines the mechanisms that drive corporate investment in scientific research. More specifically, I explore the relationship between scientific research and its use in invention, how it is organized within the firm, and its aggregated effect on firm-level outcomes, within large firms in the U.S.. To answer my research questions, I construct a novel dataset that traces above 4,000 U.S. publicly traded firms' investment in science and invention for 35 years (1980-2015). The second chapter of the dissertation provides an overview of the dataset and presents its advantages over previous data. The third chapter of the dissertation examines how the production of scientific research by U.S. corporations is related to its use in invention by the focal firm and to spillovers captured by rivals' inventions. The fourth chapter further looks at the heterogeneity in firms' investment in science by examining how the within-firm organization of scientific discovery and invention conditions research output. The findings from chapter three and chapter four suggest that as spillovers of science to rivals increase, and the greater the connectedness between research and invention practices within the firm, the less likely firms are to invest in internal scientific research.

To my beloved family.

Contents

List of Tables

List of Figures

Acknowledgements

I wish to thank my dissertation advisors Ashish Arora, Sharon Belenzon, and Wesley Cohen, for insightful discussions, valuable comments, and continuous support and encouragement. I am forever grateful for the trust you put in me to work on such important research topics early on in my career. I also thank Bronwyn Hall, Adam Jaffe, Manuel Trajtenberg, and Jim Bessen for their pioneered work on patent data that was the building block and inspiration for this dissertation and the DISCERN database. I am also grateful to my Ph.D. colleagues Bernardo Dionisi, Honggi Lee, Dror Shvadron, and JK Suh for their continued help and support in constructing the DISCERN database. Finally, I want to thank the women of the Fuqua STAGE Ph.D. club for their encouragement and support during my time at Fuqua.

1

Introduction

Throughout the 20th century, corporate firms, such as Bell Labs and Xerox, attracted talented researchers who produced Nobel prize-winning research. Corporate research spurred significant inventions, including the transistor, the laser, and the first computer with a graphical user interface, as well as breakthroughs in medicine and pharmacology. The type of research conducted in large firms is different in its nature than that undertaken in universities. Large firms have access to complementary resources and can tackle multidisciplinary problems more easily than universities. While corporate engagement in scientific research is an important activity that has been linked to R&D productivity and economic growth, over the last three decades, corporations have reduced their investment in scientific research (Mowery, 2009; National Science Foundation, National Center for Science and Engineering Statistics, 2019; Arora et al., 2018). Similarly, corporate representation in scientific literature is also shrinking. In an analysis of publications authored by publicly traded U.S. companies, we observe that corporate publication rate fell by about 60% over the sample period (Arora et al., 2021a).

My dissertation examines the mechanisms that drive corporate investment in scientific research. More specifically, I explore the relationship between scientific discovery and its use by inventions, how it is organized within the firm, and its aggregated effect on firm-level outcomes, within large firms in the U.S..

I build on a novel dataset that I constructed that traces above 4,000 U.S. publicly traded firms' investment in science and invention for 35 years (1980-2015). The dataset introduces a major extension and improvement to the historical NBER patent data (Hall et al., 2001; Bessen, 2009). It improves matches and dynamic reassignments by tracking name and ownership changes throughout the sample period. The second chapter of the dissertation, co-authored with Ashish Arora and Sharon Belenzon, provides an overview of the data and compares it to NBER 2006 patent data.

The third chapter of my dissertation, co-authored with Ashish Arora and Sharon Belenzon, revisits the fundamental view that sees science as an input into invention (Bush, 1945; Rosenberg, 1990; Narin et al., 1997). One might suggest that research is becoming less relevant for invention over time. But in fact, we show that patents continue to build on science at ever-increasing rates.

In a firm-level analysis of 4,000 American publicly traded firms between 1980 and 2015, we investigate how the production of scientific research by U.S. corporations is related to its use in invention by the focal firm and spillovers captured by rivals' inventions. We find that over the last 3.5 decades, spillovers have increased relative to internal use. Changes over time in the importance and magnitude of spillovers and internal use are potentially an important cause of the decline in corporate production of scientific research.

We show that the private returns to corporate research depend on the balance between two opposing forces: the benefits from the use of science in own downstream inventions, and the costs of spillovers to rivals. Firms produce more future research when it is used internally, but less research when it is used by rivals.

Put differently, we show that even as firms make greater use of the scientific knowledge produced by rivals, they themselves are less willing to produce such knowledge. This tradeoff between internal and external sourcing has an important implication for managers. Managers must understand how to organize their firm's R&D to capture external opportunities and capitalize on declining internal science.

Following this insight, the fourth chapter of my dissertation examines the organization of R&D within firms. On the one hand, to improve internal use and make research more immediately relevant to the firm's needs, there should be greater connectedness between research and invention practices within the firm (Kline and Rosenberg, 1986; Rosenberg, 1990). However, less specialization can constrain research productivity (Smith, 1776), which is important for long-term significant breakthroughs.

Using data on inventors and authors related to U.S. publicly traded scientific firms for the period 1980-2015, I explore the implications of the internal organization of scientific discovery – either integrated with invention or specialized – on the firm's invention, scientific discovery, and market value outcomes. To the best of my knowledge, my research is the first to directly explore both the benefits and costs of scientific discovery organization at the firm level.

I show that integration of scientific discovery with invention is related to a tradeoff between short-term applied R&D and long-term fundamental R&D. While integration is beneficial for internal invention, it has adverse effects on scientific output, which in turn decreases invention quality in the long run. I find that the negative relationship between integration and publication reduces the direct increase in patents due to integration by approximately 60%. I also show consistent results in terms of market value - the private value of patents increases with integration, while the private value of publications decreases with integration.

Furthermore, I present three main determinants that condition this tradeoff: reliance of invention on science, stage of technology, and external market for technology. I show that firms optimally choose a higher level of integration when their inventions are more fundamental science-based, when their technology is in early stage, and when external technology is more abundant.

Taken together, the results from chapters three and four suggest that part of the reduction in corporate investment in science can be related to firms building more on rival scientific discovery and becoming more integrated over time. This shift, though likely privately profitable, is not without social costs. The declining corporate engagement in research may be contributing to the reported decline in R&D productivity and the associated decline in productivity growth.

2

Matching Patents to Compustat Firms, 1980-2015: Dynamic Reassignment, Name Changes, and Ownership Structures

This chapter is adapted from a joint work with Ashish Arora and Sharon Belenzon that is published in Research Policy Journal (Arora et al., 2021b). For the original article please see: https://doi.org/10.1016/j.respol.2021.104217. This chapter also draws substantially on an unpublished Online Appendix to Arora et al. (2021a). All authors have equal contribution.

2.1 Introduction

An extensive literature uses patent data to answers questions on the determinants and consequences of inventive activity. Many papers use patent data from the NBER, which matches patents granted by the United States Patent Office to publicly traded American firms (henceforth, Compustat firms). There are two versions of the NBER data. The first, introduced by Adam Jaffe, Bronwyn Hall, and Manual Trajtenberg (Hall et al., 2001), pioneered the use of patent data as indicators of inventive activity at scale. We refer to this dataset as NBER '01, which covers the years 1980-1999. The second iteration of this database was developed by Jim Bessen (Bessen, 2009) for the period 1980-2006 to address some shortcomings in the earlier version (henceforth, NBER '06). In particular, Bessen (2009) included an attempt to improve the dynamic reassignment of patents, wherein as firms changed owners, their patents would get reassigned (in the database) to the new owner.

This chapter describes a third iteration of the NBER data (henceforth, ABS), first used in Arora et al. (2021a). We extend the data by a decade to 2015. In addition, we reconstruct the complete historical data covered in the NBER data files. We build on Bessen's work and introduce several improvements focusing on better coverage of name changes and ownership structures.[1] We study the implications of our improved matches on patent value and R&D elasticity of patenting.

We combine data from five main sources: (i) company and accounting information from U.S. Compustat 2018, (ii) patents from PatStat; (iii) subsidiary data from historical snapshots of ORBIS files for 2002-2015; (iv) mergers and acquisition data from SDC Platinum and (v) company name changes from WRDS's "CRSP Monthly Stock". For ownership and subsidiary data, we supplement NBER '06 with a wide range of M&A data, including SDC, historical snapshots of ORBIS files for 2002-2015, and 10-K SEC filings. We further perform extensive manual checks to uncover firms' structure and ownership changes.

There are two main areas of improvement. First, we match more patents. About 20% of patents belonging to Compustat firms were omitted or incorrectly matched in the NBER '06 patent database. Second, we achieve better dynamic reassignment.

[1] Assignee disambiguation is not an important contribution of our data work. We build on existing name harmonization and string matching approaches. Several patent data projects such as the USPTO's PatentsView, UC Berkeley's Patent Database (Li et al., 2014), and Darden's Global Corporate Patent Dataset (GCDP) (Bena et al., 2017) have advanced assignee disambiguation and assignee-matching techniques. Researchers can use our historical standardized name lists to match with their dataset of interest.

Dynamic reassignment means that, for instance, if a sample firm merges with another firm, the stock and flow of patents of the merged firm are linked to the Compustat record from that point onward, but not before. Similarly, once a subsidiary is divested from its parent Compustat firm, its patents are subtracted from the parent firm's stock from that point onward. Lastly, a significant component of the data upgrade is accounting for changes in names. About 30% of the Compustat firms in our sample change their name at least once. Accounting for name changes improves the accuracy and scope of matches to patents (and other assets), ownership structure, and dynamic reassignments of GVKEY codes to companies. Finally, we make our data available to all researchers through a public data repository.[2]

We examine how estimates of the "shadow price" of patents in market value regressions and R&D elasticity in patenting regressions change across NBER datasets and our updated data. In general, we find that our data produces higher estimates of patent value from estimating Tobin's Q specifications, as well as higher R&D elasticity estimates from a patent production function.

Section 2 explains how we construct our sample. Section 3 discusses the main challenges we face when matching patent data to Compustat firms. Section 4 presents the patent matching procedure and the process of dynamic reassignment. Section 5 compares ABS data to NBER '06, including detailed case studies to illustrate improvements to the NBER database. Section 6 analyzes differences in estimates of patent value and R&D productivity between existing NBER datasets and our sample, and Section 7 concludes.

[2] The data can be freely downloaded from 10.5281/zenodo.3594642. The version used for the analysis in this paper is version 5. We are making changes to the data on a regular basis, and they are updated to the Zenodo website. Please check for the latest version available.

2.2 Sample construction

To construct our sample, we start with all North American Compustat records obtained from WRDS in August 2018 and select companies with active records and positive R&D expenses for at least one year from 1980 through 2015 (inclusive). We define an active record as a year with common shares traded (CSHTR_F). We exclude firms that are not headquartered in the United States based on their current headquarter location in the Compustat 2018 file.[3] We further restrict our sample to companies with at least one patent during our sample period.

Following NBER '06, we aggregate the data to the parent company level, which we refer to as an ultimate owner (hereafter, UO). For example, the company GENZYME CORP (GVKEY 12233) is the ultimate owner of the publicly traded companies GENZYME MOLECULAR ONCOLOGY (GVKEY 117298), GENZYME TISSUE REPAIR (GVKEY 118653), GENZYME SURGICAL PRODUCTS (GVKEY 121742), and GENZYME BIOSURGERY (GVKEY 143176). Yet, GENZYME TRANSGENICS CORP (a.k.a. GTC BIOTHERAPEUTICS, GVKEY 028563) is a standalone company, because it has been spun-off by GENZYME CORP. A major contribution of our data is tracing ownership changes over time and identifying the exact years when GENZYME TRANSGENICS CORP falls under GENZYME CORP and the years when it is a standalone firm. An additional important contribution is accounting for private subsidiaries as well as Compustat, publicly-traded, subsidiaries.

A UO firm enters our sample once it is publicly traded and has at least one patent in stock and remains in our data until the end of the sample period unless it is acquired, dissolved, or taken private. All UO firms in our sample have at least 3 consecutive years of active records in Compustat. In total, we match 1.3 million patents to 4,420 U.S. headquartered Compustat firms and their subsidiaries. These

[3] 18.5% of the related Compustat firms are dropped due to the restriction on U.S. headquartered firms.

patents account for about 50% of all utility patent grants of U.S. origin.[4] When a patent has several assignees, we match the patent to multiple firms and assign fractional ownership to each assignee (i.e., 1/number of assignees). In case of an ownership change within the sample, patents are dynamically matched to up to five UO firms (see Section 4.2 for detail). We do not account for patent reassignments that are not part of a corporate ownership change.

2.3 Key measurement challenges

Working with patent and Compustat data presents several challenges. Many of the challenges arise because Compustat uses GVKEY codes to track companies, but GVKEY codes fail to capture changes over time in ownership structure and firm names, and the same company may have multiple codes over time. For example, Ralston Purina is listed under two different GVKEYs: (i) 1980-1993 under "RALSTON PURINA - CONSOLIDATED" (GVKEY 008935) and (ii) 1993-2000 under "RALSTON PURINA CO" (GVKEY 028701)). Compustat does not link related company identifiers, making it difficult to track companies over time only based on GVKEY. For example, "AT&T CORP" (GVKEY 001581) stopped being traded independently in 2005 after it was acquired by "SBC COMMUNICATIONS INC" (GVKEY 009899), which in turn changed its own name to "AT&T INC" Compustat does not provide information on these changes. To overcome this challenge, we develop a specialized database on current and historical ownership structures and name changes for our sample firms.

[4] We match 58% of all U.S. origin patents that are assigned to a U.S. company or corporation during the period 1980-2015 (assignee type classification is based on PatentsView data). When comparing ABS patents with the unmatched patents, we find that the main difference is in the use of in-text and front page NPL citations (Marx and Fuegi, 2020a,b). Unmatched patents cite on average 1.6 in-text and 2.3 front-page citations more than ABS patents. These patterns are consistent with large firms having a much higher propensity to patent. Specifically, they are likely to file a large number of incremental patents (Cohen and Klepper, 1996), which are less likely to cite science. Unmatched patents are also more likely from science-intensive sectors, such as life-sciences.

2.3.1 Company names

A major contribution of this paper is identifying name changes of Compustat firms over the sample years 1980-2015. Company names are important because U.S. patent documents list the name, rather than the CUSIP number or GVKEY of the assignee. Patent records contain the owner's name at the time of their publication, whereas companies appear in the Compustat file under their most current name with no records of previous names. Company names may change over the course of our sample period due to generic name changes (e.g., "MINNESOTA MINING AND MANUFACTURING" changed its name in 2002 to "3M") or due to M&As (e.g., "WESTINGHOUSE ELECTRIC CORP" (GVKEY 011436) acquired "CBS INC" in 1995 and changed its own name to "CBS CORPORATION" in 1997 while maintaining the same GVKEY Compustat firm identifier). Another common reason for name changes is reverse takeovers. For example, in 1993 the private company Dentsply International Inc acquired the public company GENDEX CORPORATION (GVKEY 013700) in a reverse takeover and became publicly traded under the "DENTSPLY INTERNATIONAL INC" name and the original GVKEY. Name changes imply that we cannot simply match the name of Compustat firms to patent assignees because Compustat includes only the most recent name of the focal corporation. For example, when Google reorganized as Alphabet in the summer of 2015, Compustat updated Google's name in all historical records to Alphabet. This means that matching patents to Alphabet in 2015 would exclude Google's patents. This problem may persist even if we have complete ownership information on subsidiaries.

About 30% of the Compustat firms in our sample change their name at least once.[5] To locate historical names, we use the WRDS's "CRSP Monthly Stock" file,

[5] This is comparable to the findings by (Wu, 2010), who finds that during 1925-2000, over 30% of CRSPlisted firms changed their names at some point after going public. The three leading reasons for name changes are: (i) M&As & restructure activity (36%); (ii) change in focus of operation (17%); (iii) brand or subsidiary name adoption (12%).

which includes monthly information on names for each security over time along with its historical CUSIP code and a unique permanent security identification number assigned by CRSP, the PERMNO code, which is kept constant throughout the trading period regardless of changes in name or capital structure. For example, while SPHERIX INC is related to two different GVKEYs (002237 for 1980-2013 and 018738 for 2013-current, it has a unique PERMNO code for the entire period (18148)). Similarly, Google Inc PERMNO code is 90319 and it remains the same after the company reorganized as ALPHABET INC in 2015. We calculate the starting and end years for each name based on the trading dates in the "CRSP Monthly Stock" file.

Using WRDS "CRSP/Compustat Merged Database - Linking Table", we link each PERMNO from CRSP to Compustat GVKEY code. The crosswalk between CRSP and Compustat is not straightforward. As shown above, a PERMNO can have multiple related GVKEYs. In such cases, we apply a dynamic match between a PERMNO and Compustat accounting data. However, CRSP also includes cases where under the same GVKEY there are several PERMNO codes. [6] In such cases, we manually checked using 10K-SEC fillings the years that the name was relevant for each GVKEY.[7]

We perform extensive checks on the name list, including identifying and distinguishing companies with similar names. For instance, RACKABLE SYSTEMS INC (GVKEY 162907) changed its name to SILICON GRAPHICS INTL CORP after it acquired the public company SILICON GRAPHICS INC (GVKEY 012679) in 2009. We ensure that we count SILICON GRAPHICS patent stock and patent flow un-

[6] This is mainly due to significant M&As, including reverse acquisitions that occurred during the years when the firm was not listed.

[7] There is a difference in coverage between CRSP and Compustat for the early sample years. For example, CRSP only includes firms listed in major American exchanges and specifically excludes regional exchanges, while Compustat includes all 10K filer firms in North America. Moreover, CRSP coverage for major exchanges has expanded gradually over the years (e.g., ARCA was only added from 2006). We manually added missing information from Compustat and checked for historical names.

der RACKABLE's GVKEY starting from 2009. Similarly, we need to distinguish between the original BIOGEN INC (GVKEY 002226) and the new BIOGEN INC (GVKEY 024468) that was formed only after BIOGEN's merger with IDEC PHARMACEUTICALS CORP in 2003. Finally, we cleaned and standardized firm names as CRSP tends to abbreviate long words in the company name that it provides.

2.3.2 Name standardization

Prior to matching, we standardize firm names to reconcile company names that may be spelled differently across databases. We compose a standardization code used on both the source and the target names to increase the number of exact matches. Each company name was first standardized by converting all strings to uppercase characters and cleaning all non-alphabetic characters as well as Compustat related indicators (e.g., -OLD, -NEW, -CL A) and other common words (e.g., THE).

Additionally, an important step in standardizing the company names is standardizing abbreviations. We formed a list that includes over 80 abbreviated words matched to their various original words. For example, LABORATORIES, LABORATORY, LABS, LABO, LABORATORIE, LABORATARI, LABORATARIO, LABORATARIA, LABORATORIET, LABORATORYS, and LABORATORIUM were all abbreviated to "LAB". The list was compiled from the most frequently abbreviated words in WOS affiliation field (accordingly, the list is targeted to our sample). This list is presented in Table 2.

Table 2.1: Most frequent abbreviated words

ADV	AEROSP	AGR	AMER	ANAL	ANALYT	ANIM	APPL	APPLICAT
ASSOC	AUTOMAT	BIOL	BIOMED	BIOPHARM	BIOSCI	BIOSURG	BIOSYS	BIOTEC
BIOTHERAPEUT	CHEM	CLIN	COMMUN	COMP	CORP	CTR	DEV	DIAGNOST
DYNAM	EDUC	ELECTR	ENGN	ENVIRONM	FAVORS	GEN	GENET	GRAPH
INSTR	INTERACT	INTL	INVEST	LAB	LTD	MAT	MED	MFG
MICROELECTR	MICROSYS	MOLEC	NATL	NAVIGAT	NEUROSCI	NUTR	ONCOL	ORTHOPAED
PHARM	PHOTON	PHYS	PROD	RES	SCI	SECUR	SEMICOND	SERV
SFTWR	SOLUT	SURG	SYS	TECH	TEL	TELECOM	THERAPEUT	TRANSPORTAT

For each standardized name, we create a cleaner, fully-standardized name by omitting the legal entity endings and other general words (e.g., INC, CORP, LTD, PLC, LAB, PHARMACEUTICAL), where possible, to maximize match rates (e.g., "XEROX CORP" was standardized to "XEROX", "ABBOTT LABORATORIES" to "ABBOTT"). However, in cases where the company name is too short, generic, or can match to other strings within the affiliation field, we preserved the original standardized name to avoid mismatches and extensive manual checks on the match results. For example, omitting the legal entity from "QUANTUM CORP" would result in a potential mismatch between "QUANTUM" and "TEXAS STATE UNIV CTR APPL QUANTUM ELECTR DEPT".

The last step in name standardization is to locate abbreviations that are commonly used by companies instead of their official names. For example, "INTERNATIONAL BUSINESS MACHINES CORP", will also appear under its common abbreviation "IBM" and ''GENERAL ELECTRIC CO" under "GE". We also add the names of prominent R&D laboratories affiliated with companies, such as the T.J. Watson Research Center (IBM) and Bell Labs (initially AT&T and later under Lucent technologies), as authors often omit the name of the company when the address of the laboratory is stated as the publication address.

2.3.3 Ownership structure

Compustat does not link parent companies to their publicly traded subsidiaries, nor does it provide information on private subsidiaries, which can be owned directly by the UO or indirectly via one of its subsidiaries. Because patents can be assigned to any legal entity in the corporation (Arora et al., 2014), we need to develop comprehensive data on corporate ownership structure. Moreover, because ownership can change over time, we need to trace these changes at the UO and subsidiary level so that we can assign patents to their relevant UOs in each year.

We rely on two main data sources of information on ownership structure: (i) annual publications of ORBIS by Bureau Van Dyke, which provide us with annual "snapshots" of private and public subsidiaries and (ii) SDC Platinum, which provides us with detailed information on significant ownership changes, such as mergers, acquisitions, and spin-offs.

There are several challenges in identifying subsidiaries owned by Compustat firms. First, many of the subsidiaries are private, and manual checks are required to verify which of the several similarly named companies is actually owned by a focal UO. Second, subsidiary ownership changes over time. Companies may spin out their subsidiaries, some of which might go public or be sold to other firms. A major contribution of our data is developing comprehensive time-varying data on corporate ownership structure.

Our primary source of information on subsidiaries owned by Compustat firms is ORBIS. We use ORBIS's complete ownership data for each year from 2002 through 2015, because 2002 is the first year ORBIS reported reliable firm coverage for American firms. For earlier years, we rely on NBER files and 10-K SEC filings. Our first step is to match the names of Compustat firms to ORBIS. To do so, we standardize the names of the "Global Ultimate Owner" field in ORBIS, hereafter, GUO, similar to the standardization procedure we used for Compustat firms (see Section 2.3.2). These companies can be UOs themselves or publicly traded subsidiaries of UO firms. Having matched the names of GUOs to all historical Compustat firm names, we retain all the subsidiaries listed in ORBIS of the successfully matched GUOs.

The next step is to match the related subsidiary names to Compustat. We restrict our sample to subsidiaries that are majority-owned by the GUO firm.[8] This yields an

[8] The 10-K SEC filings used to supplement ORBIS data for the pre-2002 period usually list only majority-owned subsidiaries. Therefore, to be consistent throughout the sample period, we use only majority-owned subsidiaries. Since across the sample period, 75% of the subsidiaries in the ORBIS files are majority-owned, we are capturing most of the subsidiaries. Furthermore, in most ORBIS files for our sample period (2002-2012 files), the non-majority-owned subsidiaries related to our sample firms are no more than 20%. Only from 2013, the non-majority-owned subsidiaries

ownership link within Compustat between parent firms and their public subsidiaries. We use this information to aggregate the patent matching to the parent company level. In addition, ORBIS provides us with private subsidiaries for each Compustat firms, which we later use in the patent matching procedure.

Changes in ownership happen for a diverse set of reasons, including mergers, acquisitions, and spinoffs. We rely on SDC Platinum as an additional source of information for changes in ownership. While ORBIS provides time-series information on ownership structure, its main advantage is mapping subsidiaries to parent firms in specific years. SDC, on the other hand, is a specialized database that focuses only on ownership changes. We ute SDC to track ownership changes both at the UO and subsidiary level.

We downloaded detailed information on the acquirer and target firm names, acquirer and target firm CUSIPs, type of deal, execution dates, and percentage of shares owned after each transaction. We restrict our focus to deals involving a change in ownership that resulted in majority ownership (more than 50% of shares) for the acquirer, and exclude deals involving asset or business unit acquisitions. We standardize target and acquirer names (see Section 2.3.2) and match them to Compustat firms and their related subsidiaries.

For subsidiaries, execution dates are used to define the years a subsidiary begins or ceases to being owned by the GUO (in case of several acquisitions during the sample period). For UOs, we track up to five ownership changes for each firm name after it enters Compustat and one additional reassignment before it became publicly traded if relevant (i.e., if it was a subsidiary of another Compustat firm in our sample prior to its IPO).[9] We assume that if a firm is acquired, all its patents are transferred to

jump to 40%.

[9] For example, Vysis Inc first enters our sample as a subsidiary of Amoco (1991-1997) and is then spun-off and becomes an UO firm in our sample as an independent publicly traded company in 1998 and is eventually acquired and becomes a subsidiary of Abbott in 2001.

the acquirer firm. All subsidiaries are also assumed to move with their parent firm when the parent firm is acquired, unless indicated otherwise. We do not account for reassignments of patents that are not part of the ownership changes that we document.

2.4 Assigning patents to firms

Matching patents to their owner for each year involves two main steps. In the first step, we match names of patent assignees to names of companies that can potentially own the patent. Second, we track changes in the ownership of a patent over time. We explain each step below.

2.4.1 Matching assignees to company names

We match our sample of Compustat firms and their subsidiaries to names of patent assignees from PatStat, which has approximately 5.3 million patents granted between 1980 and 2015. We remove published patent applications (i.e., publication numbers longer than 7 characters), non-utility patents, including Design, Reissue, Plant and T documents, and reexamination certificates. We also remove patents assigned to individuals or government entities (for example, an assignee that includes the string "DECEASED" or "U.S. DEPARTMENT"). This procedure leaves us with roughly 5 million patents and 897 thousand unique standardized assignee names, which we match to sample firms as follows.

We begin by matching firm names to assignees using an exact match procedure. For unmatched patents, we implement several fuzzy matching techniques to account for names that are slightly different, but are in fact the same entity. The final step includes manual checks at the assignee name and patent levels to ensure the correctness of the matches. The matching was carried out twice, both for standardized and for original names.[10] Special care was taken in cases where firm or assignee names are generic, when several different firms share a common portion of a name, or when firm names contain a common given or family name. To resolve ambiguities, we performed web searches and examined the actual patent documents.

For the remaining assignee names not matched during the exact matching process, fuzzy matching was performed to find each of the assignee names from the firm names to identify cases where assignee and firm names do not match exactly but are, in fact, the same firm. Some names are misspelled or include additional letters that prevent an exact match. In other cases, patent assignee names include a specific division title ("ROCKWELL BODY AND CHASSIS SYSTEMS", "ROCKWELL SOFTWARE"), a licensing unit ("MICROSOFT TECHNOLOGY LICENSING LTD", "RCA LICENSING"), or a geographic branch or firm location ("BIOSENSE WEBSTER ISRAEL LTD").

Fuzzy matching was performed using the FuzzyWuzzy library in Python (i.e., Token Set function), and using term frequency-inverse document frequency (TF-IDF). FuzzyWuzzy uses a slightly modified Levenshtein distance to calculate similarities between two strings. More specifically, a vector is created for each assignee name using the words contained in it and then compared to the entire list of firm names (that are also vectorized) to find potential matches. When comparing two vectors, the same elements (words) contained in both vectors are marked as "matched", and the similarity between the remaining different elements are calculated using the Levenshtein distance algorithm after sorting the elements alphabetically. The similarity score between the two strings is higher when the elements that match exactly make up a larger portion of the strings and when the remaining (unmatched) part has a small distance based on the Levenshtein distance. To account for multiple scores

[10] An additional match was performed after dropping legal entities, to account for firms whose names differ only by the legal entity.

that indicate a strong match, the top ten potential matches with the highest scores are examined manually to identify the most appropriate match.

We perform an additional search for the top 300 patenting firm names to find matching assignee names that were not matched through the fuzzy match process. We search for assignee names with at least five related patents that contain any of the fully standardized firm names after the removal of legal entities. Through this process, we include subsidiaries that have the same organic name as the parent UO firm (For example, "EMERSON" firm name matched with "EMERSON CLIMATE TECH", a division within the firm). The search was conducted through a script that receives the list of assignee names and fully standardized firm names and automatically produces all matching pairs. In each search result pair, a firm name is contained within the assignee name string. Following the search, a complete manual check was conducted among all search results to mark the legitimate matches.

As a final check, we employed a team of RAs to verify that assignees with more than 100 patents were correctly matched by the fuzzy matching algorithm. The RAs went through the fuzzy matched names to confirm that they are in fact, the right match. Existing matches were invalidated when they were not the right match, and new matches were added when more appropriate matches were found.

2.4.2 Dynamic reassignment

We build on the methodology developed by Bessen (2009) and perform a dynamic reassignment for our sample of UO Compustat firms. Our matching is done at the firm name level. We assign each firm name a unique identifier labeled as ID_NAME and indicate the first, and last year the name is relevant for a PERMNO_ADJ – our UO identifier. We then perform dynamic matching of names to PERMNO_ADJ based on SDC's M&A data. PERMNO_ADJs are dynamically linked to GVKEYs (to link to Compustat data). The same ID_NAME can be linked to multiple PERMNO_ADJ over time, and each PERMNO_ADJ can be linked to multiple GVKEYs within the same year and over time.

We include up to five ownership reassignments for each firm name that appears in our initial Compustat sample and acquired by another firm in our sample. Our UO and subsidiary historical standardized name lists, including dynamic reassignments, are publicly available.

2.5 Comparison with NBER data, 1980-2006

We compare ABS sample to NBER '06 for the period 1980-2006 (Supplementary Figure 2.1 plots the difference in matched granted patents over time). Table 2.2 presents the comparison results.

About 80% of the patent-GVKEY matches are identical between the NBER and ourselves. We match an additional 18% of the patents mostly due to: (i) improved dynamic linkage of patents to GVKEYs (e.g., Pharmacia), and (ii) linkage of additional patents based on historical name information, wider M&A coverage, and improved matching techniques (e.g., Phillips). In 1% of the cases, we find the same assignment as NBER, but these matches are irrelevant for our sample (e.g., Rhone-Poulenc). Finally, in about 1% of the cases, we are unable to include the NBER matches for a variety of reasons, including possible mistakes on our part. In an unreported analysis, we compare the citation patterns of ABS patents with patents matched only by NBER '06. The difference in yearly weighted forward patent citation per patent (ABS minus NBER '06) is only 0.066 (0.08 for IPC-year weighted forward patent cites), which is less than 6% of the mean value of forward citations per patent.

Table 2.3 examines differences in characteristics of firms that are mismatched by

Table 2.2: Comparison of ABS with NBER for 1980-2006: Patent-GVKEY Assignments, U.S. HQ Firms

Comparison 1980-2006	% Patents	Examples
Agreement	80	
Matched to different GVKEY	4	Improved dynamic matching to Compustat records using historical name. Patents of the merged company included under the GVKEY from acquisition, but not before. Example: PHARMACIA: we matched to PHARMACIA & UPJOHN's GVKEY pre-2000 instead to MONSANTO
Only our Sample	14	Newly matched patents due to (i) availability of historical names; (ii) better M&A data; and (iii) Improved matching. e.g. PHILLIPS PETROLEUM CO: 4000+ patents pre-merger with Conoco Inc in 2002; MONSANTO: 2000+ patents pre-merger with Pharmacia.
Only NBER - we matched but irrelevant gvkey-year	1	(i) NBER match (incorrectly) based on 2006 Compustat name: e.g. 1000 patents of RHONE-POULENC patents matched to RORER's GVKEY pre-merger in 1990; (ii) Improved subsidiary coverage: e.g., 450 patents of HUGHES AIRCRAFT are incorrectly linked to GM's GVKEY pre-1985 acquisition.
Only NBER	1	(i) Withdrawn patents: 600 patents (ii) Misc. could not verify connection, typos, and possible mistakes by us

NBER '06 and firms that are matched correctly. For this exercise, the sample consists of firm-year level observations matched on GVKEY-year pair between NBER '06 and ABS samples for the period 1980-2006. We classify a firm as mismatched if the firm's cumulative patent stock in ABS is at least 10% above or below the number reflected in the NBER '06 sample.[11] Comparing NBER '06 to ABS, 1,664 firms are mismatched

[11] To facilitate comparison between the samples, we consider only patents granted under the current UO.

(approximately 40%). Of the mismatched firms, 129 firms have more patents in NBER '06 than in ABS. That is, for these firms, ABS revise downwards their number of patents (mostly through dynamic reassignment). Comparing mismatched firms with a downward revision of their patents to those with an upward revision of patents, downward revision firms have lower Tobin's Q (3.4 for downward vs. 4 for upward), yet are bigger in terms of sales (2,271 for downward versus 1,858 for upward) and assets (1,345 for downward versus 1,337 for upward). Overall, Table 2.3 shows that mismatched firms are bigger as measured by assets and sales, but have a lower Tobin's Q, and R&D stock and flow. Supplementary Table 2.14 further presents a list of the top 50 firms with the highest number of average annual difference in matched patents. The number of firms with at least one mismatched patent ranges from 288 in 1980 to 658 in 2006. For these firms, the average absolute difference in matched patent per year is 12.2. Mismatches are higher in chemistry and life science with an annual 13 mismatched patents per firm, as compared to 12 and 10.75 mismatched patents in ICT and semiconductors, respectively.

Table 2.3: Difference in Means: Mismatched and Matched Firms (ABS vs. NBER '06, 1980-2006)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Diff. in means	Mismatched			Not-mismatched		
	(3) minus (6)	No. Obs.	Mean	Std. Dev.	No. Obs.	Mean	Std. Dev.
Market value	-157.8	21,146	2,427.8	11,917.3	22,302	2,585.5	20,618.5
Tobin's Q	-0.9**	21,146	4.0	5.7	22,302	4.9	6.3
R&D stock	-108.6**	21,146	248.3	1,054.9	22,302	356.9	2,098.7
R&D expenditure	-19.9**	21,146	58.8	272.8	22,302	78.7	430.7
Assets	307.7**	21,146	1,337.2	6,219.2	22,302	1,029.5	7,606.3
Sales	392.2**	21,146	1,889.6	8,247.3	22,302	1,497.4	8,583.2

Notes: The sample consists of firm-year level observations matched on GVKEY-year pair between NBER '06, and ABS samples. Mismatched firms are those whose cumulative granted patents in ABS is at least 10% above or below the number reflected in the NBER '06 sample. ** $p<0.01$, * $p<0.05$

Table 2.4 presents a regression version of the difference in means analysis. The dependent variable is the absolute number of mismatched annual patents between ABS and NBER '06, normalized by the number of annual patents of the firm. The regression results suggest that larger firms (as measured by assets) are likely to see higher share of mismatched patents. Further, patent-intensive firms (measured as the ratio of patents per R&D expenditures) have a higher share of mismatched patents between ABS and NBER '06.

Table 2.4: OLS Estimates. Absolute Percentage Difference in Annual Patents Assigned (ABS vs. NBER '06)

	(1)	(2)	(3)	(4)	(5)	(6)
DV:Absolute difference in annual patents (ABS minus NBER '06) divided by mean annual patents						
	Unscaled RHS variables			Scaled RHS variables		
VARIABLES	Pooled	Within-firms	Between-firms	Pooled	Within-firms	Between-firms
ln(R&D stock)	0.019*	0.009	0.012			
	(0.009)	(0.019)	(0.009)			
ln(Assets)	0.021**	0.010	0.039**			
	(0.007)	(0.008)	(0.007)			
ln(Tobin's Q)	-0.001	-0.007	0.017	-0.011	-0.009	-0.001
	(0.006)	(0.006)	(0.010)	(0.007)	(0.006)	(0.011)
Patents / R&D exp.	0.058**	0.063**	0.028**	0.048**	0.062**	0.003
	(0.006)	(0.005)	(0.009)	(0.006)	(0.005)	(0.008)
R&D stock / Assets				-0.003	-0.001	-0.007
				(0.003)	(0.004)	(0.004)
Industry fixed effects	Yes	No	Yes	Yes	No	Yes
Firm fixed effects	No	Yes	No	No	Yes	No
Year fixed effects	Yes	Yes	No	Yes	Yes	No
DV sample average	0.385	0.385	0.361	0.385	0.385	0.361
Number of firms	3750	3750	3771	3750	3750	3771
Observations	34,047	34,047	3,771	34,047	34,047	3,771
R-squared	0.08	0.50	0.13	0.07	0.50	0.11

Notes: The sample consists of firm-year level observations matched on GVKEY-year pair between NBER '06, and ABS samples. The absolute difference is divided by mean of ABS and NBER '06 patent for the firm in that year. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by GVKEY. ** p<0.01, * p<0.05

2.5.1 Case studies

We present several case studies to illustrate the outcome of our patent matching procedure.

SEALED POWER and GENERAL SIGNAL

The following example underscores the mismatching consequences of not accounting properly for name and ownership changes. Until 1998, SEALED POWER and GENERAL SIGNAL were two distinct entities. Historical Compustat records include the following information for these companies until 1998:

1. GVKEY 9556, related names:
   (a) SEALED POWER CORP (1962-1988) – original name
   (b) SPX CORP (1988-1997) -name changed retroactively in Compustat
2. GVKEY 5087, related name: GENERAL SIGNAL CORP (1950-1997)

In 1998, SPX Corp acquired General Signal Corp in a reverse merger, and General's GVKEY (5087) became the new security of SPX traded retroactively under the new name "SPX CORP". The original SPX records were renamed retroactively in Compustat as "SPX CORP-OLD". Current Compustat records include the following records for these companies for the complete period they are traded:

1. GVKEY 9556, related name: SPX CORP-OLD
2. GVKEY 5087, related name: SPX CORP

We treat these GVKEYs as two separate companies up to 1997 accounting for all relevant names (SEALED POWER CORP, SPX CORP for GVKEY 9556, and GENERAL SIGNAL CORP for GVKEY 5087) and connect the SPX CORP name to

General's original GVKEY (5087) only from 1998 onward. In the NBER '06 patent dataset (see table 2.5), the two companies are collapsed under the same company (same PDPCO id) and for the purpose of Compustat accounting information General's original GVKEY (5087) is used for the complete period while the original SPX GVKEY (9556) is disregarded. Indeed, in the NBER files SPX patents pre-1998 are matched to General's GVKEY. Moreover, patents related to "GENERAL SIGNAL CORP" (757 patents without considering related subsidiaries) as well as "SEALED POWER CORP" (36 patents without considering related subsidiaries) are located in the 2006 NBER raw patent match but are not assigned to any Compustat record.

Table 2.5: Data for SPX Corp in NBER '06

Current name	gvkey	firstyr	lastyr	pdpco	pdpseq	begyr	endyr
SPX CORP	5087	1950	2006	5087	1	1950	2006
SPX CORP-OLD	9556	1962	1997	5087	-1		

Notes: PDPCO is NBER's Patent Data Project (PDP) unique company id. FIRSTYR is the first year GVKEY company has data. LASTYR is the last year a GVKEY company has data. PDPSEQ is the GVKEY sequence within PDPCO. If PDPSEQ=-1, the related GVKEY is disregarded. BEGYR is the beginning year for GVKEY within PDPCO. ENDYR is the last year for GVKEY within PDPCO. All patents related to SPX CORP will be accounted under GVKEY 5087 from 1950 to 2006, while the original SPX GVKEY (9556) is disregarded.

CONOCO and PHILLIPS PETROLEUM

This example illustrates the importance of historical names and ownership changes. In 1981 Conoco was acquired by Dupont, which later spun it off as a publicly traded company, which was eventually acquired in 2002 by the publicly traded company Phillips Petroleum. The merged entity was renamed ConocoPhillips. Examining current Compustat records, we can only locate the name ConocoPhillips with no record of Philips Petroleum. Compustat does not provide any information on the owner of the record prior to the merger. We use the CRSP monthly stock file to locate all historical names of related securities. Table 2.6 shows the name list for Conoco-Phillips.

Two important changes resulted from this name list. First, we located about 4,000 patents issued to Phillips Petroleum prior to the merger with Conono, which were not previously matched. Second, we dynamically reassign patent stock. For example, U.S. patent Num. "5404954" was granted to Conoco Inc in 1995. At that time, Conoco was a subsidiary of Dupont. In our data, this patent would be included in Dupont's patent flow for 1995. It will also be counted under Dupont's patent-stock for 1996-1997. However, from 1998, when Conoco is spun-off as an independent publicly traded company, this patent would be transferred from Dupont to Conoco's patent stock. Similarly, in 2002 the patent would transfer to ConocoPhillips patent stock. A different patent, issued to Phillips Petroleum in 1999, would be part of the patent flow assigned to Phillips in 1999 and be counted under the patent stock for Phillips Petroleum until 2002, and then would move on to become part of ConocoPhillips patent stock.

Table 2.6: Example of dynamic name list for Conoco-Phillip

ID Name	Name std	fyear 0	nyear 0	Pernmo Adj_0	Name ACQ_0	Fyear 1	Nyear 1	Permno Adj_1	Name ACQ_1	fyear 2	nyear 2	Permno Adj_2	Name ACQ_2
2384	CONOCO INC	1981	1997	11703	DU PONT E I DE NEMOURS & CO	1998	2001	86368	CONOCO INC	2002	2015	13928	PHILLIPS PETR CO
7325	PHILLIPS PETR CO					1998	2002	13928	PHILLIPS PETR CO	2003	2015	13928	CONOCO PHILLIPS
2385	CONOCO PHILLIPS					2002	2015	13928	CONOCO PHILLIPS				
7324	PHILLIPS 66	1980	2011	13928	CONOCO PHILLIPS	2012	2015	13356	PHILLIPS 66				

Notes: This table presents the dynamic reassignment name list related to Conoco-Phillips. ID_NAME is the unique standardized name id. NAME_STD is the standardized firm name. PERMNO_ADJ(0-5) is the UO firm id. A name can be matched dynamically up to 5 times (1-5) and to an additional pre-IPO owner if applicable (0). NAME_ACQ(0-5) is the related UO name. FYEAR(0-5) is the first-year for ID_NAME within PERMNO_ADJ. NYEAR(0-5) is the last-year for ID_NAME within PERMNO_ADJ. For example, a patent granted to Conoco Inc in 1995 would be included in Dupont's patent flow for 1995. It will also be counted under Dupont's patent-stock for 1996-1997. However, from 1998 this patent would be transferred dynamically from Dupont to Conoco's patent stock. Similarly, in 2002the patent would transfer to ConocoPhillips patent stock

TIME-WARNER and AMERICAN ONLINE

This case-study illustrates how properly accounting for name and ownership changes improve the accuracy of patent flow as well as the dynamic reassignment of patents. Warner Communication was an independent and publicly traded company until its merger with Time Inc in 1989 when Time-Warner Inc was formed. In the second half of 2000, Time-Warner was merged with American Online to form AOL Time Warner. In 2003 the company dropped the ”AOL” from its name and was renamed Time-Warner Inc. AOL remained a subsidiary until it was spun-out in 2009. This example illustrates the importance of accounting jointly for name and ownership changes, whereby there are instances where ownership changes without a name change, and other instances where name changes without an ownership change. Accounting for both changes is critical for accurate matching.

A comparison with NBER ’06 reveals the following. First, Warner Communication and its related subsidiary patents are correctly matched to WARNER COMMUNICATIONS INC (GVKEY 11284) up to the merger with Time Inc. However, they are not dynamically assigned after 1988 to Time Warner or any other company. Consequently, the patent stock and patent flow of Time-Warner (and later AOL Time-Warner) from patents related to Warner communication and its subsidiaries (e.g., Warner Bros, WEA Manufacturing), are below the true value after the acquisition in1989.

Second, TIME-WARNER’s related patents from 1991 to 2000 (before the merger with American-Online Inc in late 2000) are matched incorrectly to GVKEY 25056, which during those years was solely AMERICAN-ONLINE INC original Compustat financial records. The current name of GVKEY 25056, TIME WARNER INC, which is likely to have misled NBER to link the Time Warner patents to it, was only adopted retroactively in 2003 when the “AOL” was dropped from the official name.

Moreover, AMERICAN ONLINE INC and AOL related patents (152 patents up to 2006 based on NBER raw patent match) are not linked to any Compustat record. AOL-TIME WARNER related patents, on the other hand, are matched to a "Pro-Forma" Compustat record that is active for only two years 1999-2000: AOL TIME WARNER INC-PRO FORM (GVKEY 142022). All of which implies that AOL Time Warner's flow of patents is smaller than the actual patents it owns.

Having a complete history of names enables us to correctly identify each Compustat record and its origin and dynamically match each firm name in our sample to the correct financial records. In this case (i) AMER ONLINE INC (and later AOL) is matched from 1980 until its spinout in 2009 to GVKEY 25056 and after to AOL INC (GVKEY 183920). (ii) Warner Communication is matched up to the merger with Time Inc to WARNER COMMUNICATIONS INC (GVKEY 11284) and later dynamically transferred ending up in AOL -Time Warner GVKEY (25056) starting 2001. (iii) AOL -Time Warner is matched to AOL -TIME WARNER (GVKEY 25056) starting 2001 after the merger was approved. (iv) Time Inc itself is not included as an UO in our sample as it did not have R&D expenses, but it is included as a subsidiary name under the Time-Warner UO company.

PHARMACIA, UPJOHN and MONSANTO

This example demonstrates that having a complete history of names enables us to correctly identify each Compustat record's historical ownership and dynamically match each firm name in our sample to its relevant financial records in each period. For instance, linking each patent to its correct financial record can be a concern for papers that link patents to market value, specifically those distinguishing different types (e.g., high vs. low cited patents), which rely on the specific patent that was matched and not only the quantity.

In 1995 original Pharmacia merged with Upjohn to form Pharmacia & Upjohn.

In 2000, original Monsanto merged with Pharmacia & Upjohn to form Pharmacia Corporation (New Pharmacia). Between 2000-2002 the new Pharmacia gradually spun off its agricultural operations to a newly created subsidiary, Monsanto Company (New Monsanto). In 2003 the new Pharmacia was acquired by Pfizer and is now a wholly-owned subsidiary of Pfizer. Matching the original assignee name to a current Compusat file can result in misallocating patents of Pharmacia pre-2000 to the New Pharmacia Compustat record, which was originally owned by Monsanto at the time, as well as not accounting for Monsanto's patents granted pre-2000. Table 2.7 summarizes how we deal with this case. On balance, it shows a complex pattern of unallocated patents, as well as misallocated patent stock.

NABISCO

Nabisco's case study illustrates how we account for ownership changes in our data. Table 2.8 shows that during our sample period, Nabisco has changed ownership four times. In 1981 Nabisco merged with the publicly traded company Standard Brands to form Nabisco Brands. In 1985 R.J. Reynolds merged with Nabisco Brands to create RJR Nabisco, which eventually became Nabisco Group holding after the tobacco business was spun out in 1999. In 2000, Nabisco was acquired by Phillip Morris, which combined Nabisco with its Kraft brand. Finally, in 2001 Kraft (together with Nabisco) was spun out as a publicly traded company that later on became Mondelez International Inc. In our dataset all Nabisco related patents and publications are dynamically transferred between Compustat records and UO firms based on its ownership throughout the years as illustrated in Table 2.8.

Examining NBER '06, Table 2.9 shows that all Nabisco related patents are linked to GVKEY 9113 from 1950 to 1999. Though the current name related to GVKEY 9113 is "Nabisco Group Holding Corp", based on the historical name information, we know that up to the merger of R.J. Reynolds with Nabisco it belonged solely to R.J.

Table 2.7: PHARMACIA & UPJOHN and MONSANTO dynamic match

Period	**related GVKEY**	**Relevant Compustat name for period**	**Most recent Compustat name**	**Comments**	**Patent flow per our strategy**	**Patent flow per original NBER match**
1950-1994	11040	UPJOHN CO	PHARMACIA & UPJOHN INC	Original Upjohn before merger with Pharmacia	**2,091** Upjohn related patents	N/A
1995-1999	11040	PHARMACIA & UPJOHN INC	PHARMACIA & UPJOHN INC	**1995**: Upjohn merged with original Pharmacia to form Pharmacia & Upjohn	**479** Pharmacia & Upjohn related patents	N/A
1950-1999	7536	MONSANTO CO	PHARMACIA CORP	Original Monsanto before merger with Pharmacia & Upjohn	**3,228** Monsanto related patents	**2,733** Pharmacia & Upjohn related patents (including patents of Pharmacia before it merged with Upjohn). While Monsanto's **3,228** patents are not linked.
2000-2002	7536	PHARMACIA CORP ("new Pharmacia")	PHARMACIA CORP	**2000**: Original Monsanto merged with Pharmacia & Upjohn to form Pharmacia Corporation (New Pharmacia). All of PHARMACIA, UPJOHN and PHARMACIA & UPJOHN patents are transferred here from 2000. Monsanto's patents are redirected to the new Monsanto spin-off company.	**304** Pharmacia & Upjohn related patents	**304** Pharmacia & Upjohn related patents
2000-2015	140760	MONSANTO CO ("new Monsanto")	MONSANTO CO	**2000-2002**: Pharmacia Corporation (New Pharmacia) gradually spun-off its agriculture operations to a new publicly traded company, Monsanto Co (New Monsanto). All Monsanto related patents are transferred here from 2000.	**553** Monsanto related patents (2000-2006)	**553** Monsanto related patents (2000-2006). NBER links Monsanto's patents to GVKEY 140760 from 1997 while records for 1997-1999 are available on Compustat, they are based on prospective fillings when Monsanto was still traded under GVKEY 140760.
2003-2015	8530	PFIZER INC	PFIZER INC	**2003**: Pharmacia Corporation (New Pharmacia) was acquired by Pfizer and is now a wholly owned subsidiary of Pfizer. All of PHARMACIA, UPJOHN and PHARMACIA & UPJOHN patents are transferred here from 2003.	**472** Pharmacia & Upjohn related patents (up to 2006)	**472** Pharmacia & Upjohn related patents (up to 2006)

Notes:This table presents the comparison between NBER '06 and our data for dynamic patent reassignment for Phamacia-Monsanto related patents at the GVKEY-Period level. Most recent Compustat name is based on Compustat 2018 file. Relevant Compustat name for the period is the historical firm name based on CRSP Monthly Stock file. Patent flow per our strategy is based on NBER raw patent match data for the relevant Compustat name excluding subsidiaries. Patent flow per original NBER match is based on NBER '06 data.

Reynolds. Reynold's patents, on the other hand (Over 419 patents for the period before it spun-out of RJR Nabisco and not including patents of acquired companies such as Heublein Inc), are not assigned by NBER to GVKEY 9113, and they are only being linked to Compustat records after the tobacco business spun-out of RJR Nabisco and became independently traded again under GVKEY 120877 (eventually merging with U.S. operations of British American Tobacco to form Reynolds Ameri-

Table 2.8: Nabisco dynamic match

Years	related GVKEY	Original Compustat owner	Current Compustat name	Comments
1981-1985	7674	STANDARD BRANDS INC	NABISCO BRANDS INCO	1981: Standard Brands company merged with Nabisco Inc to form Nabisco Brands Inc
1986-1999	9113	R J REYNOLDS IND INC	NABISCO GROUP HOLDINGS CORP	1985: R.J. Reynolds Industries merged with Nabisco Brands to form R J R Nabisco Inc
2000	8543	PHILIP MORRIS COS INC	ALTRIA GROUP INC	2000: Nabisco was acquired by Phillip Morris
2001-2015	142953	KRAFT FOOD INC	MONDELEZ INTERNATIONAL INC	2001: Kraft together with Nabisco split from Phillip Morris

Notes: This table presents the dynamic reassignment for Nabisco related patents at the GVKEY-Period level. Current Compustat name is based on Compustat 2018 file. Original Compustat owner for the period is the historical firm name based on CRSP Monthly Stock file.

can Inc). As a result, in 1998, the patent stock in NBER for GVKEY 9113 ("Nabisco Group Holding Corp") is 495 (consisting solely of Nabisco matched patents), whereas it should be 914 if it included R.J. Reynolds related patents. Furthermore, NBER '06 does not dynamically move Nabisco's patent-stock or account for its patent flow after 1999 when it was bought by Philip Morris and eventually became part of Kraft (a total of 529 Nabisco related patents up to 2006).

CHEMTURA CORPORATION

Chemtura Corporation case-study illustrates how having historical names helps account for ownership changes in our data and accurately compute the patent stock.

Table 2.9: Data Entry for Nabisco in NBER '06

Current compustat record name	gvkey	firstyr	lastyr	pdpco	pdpseq	begyr	endyr
NABISCO GROUP HOLDINGS CORP	9113	1950	1999	9113	1	1950	1999
NABISCO INC	7675	1950	1980	9113	-1		
NABISCO BRANDS INC	7674	1950	1984	9113	-1		
NABISCO HLDGS CORP -CL A	31427	1993	1999	9113	-1		

Notes: PDPCO is NBER's Patent Data Project (PDP) unique com y id. FIRSTYR is the first year GVKEY company has data. LASTYR is the last year a GVKEY company has data. PDPSEQ is the GVKEY sequence within PDPCO. If PDPSEQ=-1, the related GVKEY is disregarded. BEGYR is the beginning year for GVKEY within PDPCO. ENDYR is the last year for GVKEY within PDPCO. All patents related to Nabisco will be accounted under GVKEY 9113 from 1950 to 1999, while all other related GVKEYs are disregarded.

Chemtura Corporation traces back to the chemical corporation Crompton & Knowles that was founded in the 19th century. In 1996, Uniroyal Chemical Corporation merged with Crompton & Knowles. In 1999, Crompton & Knowles merged with the publicly traded company Witco to form Crompton Corporation. In 2005, Crompton acquired the publicly traded company Great Lakes Chemical Company, Inc., to form Chemtura Corporation, while Great Lakes Chemical Corporation continued to exist as a subsidiary company of Chemtura.

Based on our strategy, we consider all historical names of the current Chemtura Corporation (PERMNO_ADJ 38420) including:

1. CROMPTON & KNOWLES CORP starting 1980
2. CK WITCO CORP starting 1999
3. CROMPTON CORP starting 2000
4. CHEMTURA CORP starting 2005

Most importantly, because we consider the complete set of historical names, we are able to locate all the relevant M&As throughout the years of the publicly traded firms that exist as an independently traded company in our data prior to an acquisition. Accordingly, we dynamically transfer them post-acquisition to PERMNO_ADJ 38420:

1. Uniroyal Chemical Corporation (acquired 1996)

2. Witco Corp (acquired 1999)

3. Great Lakes Chemical (acquired 2005)

When we examine NBER '06 patent dataset, we find that the only name that was matched to CHEMTURA CORP (GVKEY 3607) is "CHEMTURA CORP" (PDPASS 13245038). As the Chemtura name was adopted in 2005, only one patent was matched for that name. In addition, none of the acquired publicly traded companies were dynamically transferred to CHEMTURA CORP post-acquisition. It is likely that a lack of information on historical names led NBER to rely on post-acquisition name (Chemtura) and thus prevented it from accounting for the M&A activities.

By considering all previous names related to GVKEY 3607: (i) Crompton & Knowles Corp; (ii) CK Witco Corp and (iii) Crompton Corp - based on the NBER raw patent match, we locate 220 additional patents up to 2006 that were not linked to any Compustat record that should be assigned to Crompton & Knowles (77 patents), CK Witco (26 patents)), and Crompton (117 patents).[12] In addition, the acquired Uniroyal Chemical Corp has a patent stock of 379 patents in 2006 (out of which 185 patents are post-acquisition), and the acquired Witco company has a patent stock of 405 in 2006 (out of which 62 patents are from post-acquisition period), and Great Lake Chemicals has a patent stock of 183 in 2006 (out of which three patents are in 2006, the year after the company was acquired). Overall, applying our strategy

[12] This calculation does not include subsidiaries and acquired companies.

to the raw NBER patent match, we find a patent stock of 1,187 patents in 2006 for GVKEY 3607 as opposed to 1 patent in NBER.

2.6 Econometric analysis

We examine how improved matching affects regression estimates of the market value of patent-stock, and of the patenting elasticity of R&D. The standard measurement error model assumes that measurement error is not correlated with either the dependent variable or any independent variable. We generalize the standard model to examine the implications of a failure to dynamically reassign patents, and of omitted patents, for estimates of the market value of patent stock, and R&D elasticity of patents.

2.6.1 Sources of measurement error and likely direction of bias: A simple econometric framework

Using boldface to represent vectors, let $\mathbf{Y}$ represent Tobin's Q (i.e., market value over assets) so that Y_{it} represents the Tobin's Q for firm i and time t. Similarly, let $\mathbf{X}$ represent patent stock over assets. Ignoring control variables for simplicity, the simplest version of the typical regression is of the form

$$Y_{it} = \alpha_0 + \alpha_1 X_{it} + \epsilon_{it} \tag{2.1}$$

Suppose X_{it} is measured with error. The error-laden variable is denoted by $X^*_{it} = X_{it} + m_{it}$, where m_{it} is measurement error. Thus the actual regression specification estimated is

$$Y_{it} = \alpha_0 + \alpha_1 X^*_{it} + \epsilon_{it} \tag{2.2}$$

Let $\mathbf{y} = \mathbf{Y} - \bar{\mathbf{Y}}$, where $\bar{\mathbf{Y}}$ is the sample mean of $\mathbf{Y}$. Define $\mathbf{x}^*$ similarly. We arrive at a simplified regression specification in deviation form, where we suppress the time subscript to avoid clutter.

$$\boldsymbol{y} = \alpha_1^* \boldsymbol{x}^{*+\epsilon}(2.3)$$

Let a_1 be the OLS estimate of α_1 in 2.1, and a_1^* be the corresponding estimate in 2.3. Then

$$\begin{aligned}
a_1^* = \frac{Cov(\mathbf{x}^*, \mathbf{y})}{Var(\mathbf{x}^*)} &= \frac{Cov(\mathbf{x}, \mathbf{y}) + Cov(\mathbf{m}, \mathbf{y})}{Var(\mathbf{x}^*)} \\
&= \frac{Cov(\mathbf{x}, \mathbf{y})}{Var(\mathbf{x})} \frac{Var(\mathbf{x})}{Var(\mathbf{x}^*)} + \frac{Cov(\mathbf{m}, \mathbf{y})}{Var(\mathbf{x}^*)} \qquad (2.4) \\
&= a_1 \frac{Var(\mathbf{x})}{Var(\mathbf{x}^*)} + \frac{Cov(\mathbf{m}, \mathbf{y})}{Var(\mathbf{x}^*)}
\end{aligned}$$

$$E(a_1^*) = \alpha_1 \frac{Var(\mathbf{x})}{Var(\mathbf{x}^*)} + \frac{Cov(\mathbf{m}, \mathbf{y})}{Var(\mathbf{x}^*)} \qquad (2.5)$$

Equation 2.5 highlights the simplest form of expressing the sources of variation in OLS estimates due to measurement error.[13] It is often assumed that measurement error is uncorrelated with the dependent variable, $\mathbf{y}$, which implies that the second term on the right hand side of equation 2.5 is zero. If the measurement error is also uncorrelated with the independent variable,$\mathbf{x}$, so that $Var(\mathbf{x}^*)$=$Var(\mathbf{x})$ + $Var(\mathbf{m})$, then $\frac{Var(\mathbf{x})}{Var(\mathbf{x}^*)} < 1$. Under these assumptions, $a_{1^*} < a_1$, we get the classical result that measurement error leads to attenuation bias.

However, depending on the source, measurement error $\mathbf{m}$ may be correlated with either or both $\mathbf{x}$ and $\mathbf{y}$. The nature of the possible bias differs between the two terms in equation 2.5. The first term deals with the magnitude of the coefficient estimate, and can inflate the magnitude of the coefficient if $Var(\mathbf{x}^*) < \mathbf{Var}(\mathbf{x})$, and deflate it if $Var(\mathbf{x}^*) > \mathbf{Var}(\mathbf{x})$. On the other hand, the second term can either bias up (positive bias), if $Cov(\mathbf{m}, \mathbf{y}) > \mathbf{0}$, or down (negative bias) if $Cov(\mathbf{m}, \mathbf{y}) < \mathbf{0}$

[13] We ignore the complications in replacing sample moments with population moments in moving from 2.4 to 2.5

2.6.2 Matching errors, and failure to dynamically reassign patents

Dynamic reassignment bias

Mergers and acquisitions, and divestitures are an important source of failure of dynamic reassignment. A firm may be acquired by another firm, or a firm may sell a subsidiary to another. From that time onwards, the assets, including the patents of the acquired entity, need to be assigned to the acquiring entity, or else the stock of patents of the acquiring entity to be lower than the true value. We work through a simple case to provide an intuitive understanding of the possible bias when patents are not dynamically reassigned. Suppose a subsidiary is sold by firm s to firm b, but the stock of patents is not adjusted. Thus firm s's stock of patents per dollar of assets (X_s) is higher than the correct number by c, and firm b's is lower. Let S represent the set of sellers, and B represents the set of buying firms. Further, suppose that there are K buying and selling firms out of a sample of N, and let $\gamma = \frac{K}{N}$. Measurement error can be represented as

$$m_i = \begin{cases} c & \text{if } i \in S \\ -c & \text{if } i \in B \\ 0 & \text{otherwise.} \end{cases}$$

Equation 2.5 points to two key quantities: $Var(\mathbf{x}^*)$ and $Cov(\mathbf{m}, \mathbf{y})$. Let $\bar{Y_s}$ represent the mean for $i \in S$, and likewise for $\bar{Y_b}$, $\bar{X_s}$, and $\bar{X_b}$

$$\begin{aligned} Var(\mathbf{x}^*) &= Var(\mathbf{x}) + Var(\mathbf{m}) + 2Cov(\mathbf{x}, \mathbf{m}) \\ Cov(\mathbf{x}, \mathbf{m}) &= \frac{1}{N}\left(c\sum_{i \in S} X_i - c\sum_{i \in B} X_i\right) = \gamma c(\bar{X_s} - \bar{X_b}) \\ Cov(\mathbf{y}, \mathbf{m}) &= \frac{1}{N}\left(c\sum_{i \in S} Y_i - c\sum_{i \in B} Y_i\right) = \gamma c(\bar{Y_s} - \bar{Y_b}) \end{aligned} \tag{2.6}$$

As long as sellers are less patent-intensive than buyers, $Var(\mathbf{x}^*) > Var(\mathbf{x})$. Therefore, from equation 2.5, $|E(a_1^*)| < |\alpha_1|$, implying that the OLS estimate is biased towards zero. However, if sellers have lower Tobin Q than buyers, $Cov(\mathbf{y}, \mathbf{m}) < \mathbf{0}$. If $\alpha_1 > 0$, as one might expect, this would reinforce the downward bias. If sellers are on average more patent-intensive than buyers, then equation 2.6 implies that it is possible that the $Var(\mathbf{x}^*) < Var(\mathbf{x})$. Thus, the first term of 2.5 may be greater than α_1 in magnitude. Notice that the quantitative importance of the bias, if any, grows over time because γ, the proportion of firms involved increases.[14] In summary, if patents are not dynamically reassigned, the estimation bias will depend upon the difference between the average Tobin Q, the difference the average patent intensity of sellers and buyers, the magnitude of the mismatch, c, and on the share of firms, γ involved in transactions

Bias from omitted matches

We match more patents, not fewer patents. That is, most of the errors in NBER are errors of omission, and relatively few are errors of commission (assigning patents to a firm that do not belong to that firm). Suppose C_i patents are incorrectly unmatched to firm $i, i \in M$, where M is the set of firms that suffer from such errors of omission. Let $c_i = \dfrac{C_i}{A_i}$, where A_i be the assets of firm i. For simplicity, let $c_i = c$. The measurement error m can be represented as

$$m_i = \begin{cases} -c & \text{if } i \in M \\ 0 & \text{otherwise.} \end{cases}$$

Following equations 2.6, and abusing the notation to let $\gamma = \frac{M}{N}$ we get

$$Cov(\mathbf{x}, \mathbf{m}) = -\gamma c \sum_{i \in M} (X_i - \bar{X}) \tag{2.7}$$

$$Cov(\mathbf{y}, \mathbf{m}) = -\gamma c \sum_{i \in M} (Y_i - \bar{Y}) \tag{2.8}$$

[14] Recall from section 2.5, the number of mismatched firms grows over time, from 142 in 1980 to 404 in 2006

If mismatched firms have higher Tobin Q than average, then equation 2.8 implies a negative bias. However, Equation 2.7 implies that if firms with missing patents are more patent-intensive than average, then it may be that $Var(\mathbf{x}^*) > Var(\mathbf{x})$. In that case, instead of attenuation bias, we would have amplification bias. If $\alpha_1 > 0$, the negative bias works against the amplification bias. The outcomes under different combinations of assumptions can be worked out similarly.

The foregoing analysis is, of course, simplified.[15] Despite its simplicity, it shows that the direction and magnitude of bias is not straightforward: It will depend on a variety of factors, which themselves depend on the source of measurement error, its magnitude, and the characteristics of the firms subject to the error. Importantly, the different sources of bias reflected in equation 2.5 may point in opposing directions. Simply put, the bias resulting from measurement error due to errors of omission and imperfect dynamic assignment is mostly an empirical matter.

2.6.3 Market value equation

In what follows, we focus the possible biases in the estimate of the coefficient of patent stock in a Tobin Q regression originating from different sources of measurement error. We include only companies that appear in the relevant datasets being compared, so that differences in estimates are not driven by changes in the sample composition (that is, extending the coverage of firms being matched in a given year, rather than the quality of the matches). Table 2.10 presents descriptive statistics for the main variables used in the estimation. We follow Bloom et al. (2013) and Arora et al. (2021a) and estimate the following Tobin's Q specification:

$$\ln\left(\frac{Value_{it}}{Assets_{it}}\right) = \alpha_0 + \alpha_1 \frac{Patent\ stock_{it-2}}{Assets_{it-2}} + \quad \alpha_2 \frac{R\&D\ stock_{it-2}}{Assets_{it-2}} + + \boldsymbol{\eta_i} + \boldsymbol{\tau_t} + \epsilon_{it} \tag{2.9}$$

[15] For instance, we have ignored other control variables.

Tobin's Q is market value over assets. *Patent stock*$_{it-2}$ and *R&D stock*$_{it-2}$ are measured as the stocks of patents and perpetual R&D stock, respectively.[16] We are interested in how the estimated market value of patents, $\hat{\alpha_1}$, differs between the NBER datasets and ours. Tables 2.11 and 2.12 present the estimation results of comparing $\hat{\alpha_1}$ across the different datasets.

Table 2.11 compares $\hat{\alpha_1}$ across NBER '01, NBER '06, and ABS (our sample) for the period 1980-1999 (the sample period used in NBER '01). To isolate the effect of measurement error, only firm-years observations common across databases are included (Panel C in Table 2.10 provides summary statistics for these firms). Columns 1-3 present the estimates for 1980-2001 for NBER '01. Column 1 controls for industry fixed effect in a pooled regression, Column 2 add firms fixed effects, and Column 3 includes only mismatched firms – whose cumulative patents for the entire sample period in ABS is at least 10% above or below that in NBER '06. The estimated $\hat{\alpha_1}$ ranges from 0.006 to 0.043 but is not statistically significantly different from zero. Columns 4-6 repeat the same specifications using the NBER '06 dataset. The estimates range from 0.028 to 0.072, and $\hat{\alpha_1}$ is statistically significant except for mismatched firms. Comparing Columns 3 to Column 6 indicates that when focusing only on mismatched firms, $\hat{\alpha_1}$ increases from 0.010 to 0.028, but remains statistically indistinguishable from zero. Columns 7-9 replicate the same specifications using ABS data. The estimated $\hat{\alpha_1}$ ranges from 0.090 to 0.043, and all estimates are statistically significantly different from zero. In particular, the estimate for mistmatched firms is 0.043, close to the estimate for the sample as a whole, 0.049. In other words, the mismatched firms appear to be representative of the ABS dataset, but not so for the NBER datasets.

Table 2.12 compares NBER '06 to our dataset by including only firm-year obser-

[16] to facilitate comparison between the samples, the patent stock includes only patents granted under the current UO

vations that appear in both for the period 1980-2006.[17] The estimate of $\hat{\alpha_1}$ obtained from NBER '06 is lower, especially for the sample of mismatched firms (0.016 versus 0.042 from Columns 3 and 6). Column 7 shows that $\hat{\alpha_1}$ remains identical to its 1980-2006 value when using our expanded sample of 1980-2015. Overall, improvements in measurement result in higher estimated private value of patent stock, particularly for firms whose patents are measured with error.

[17] Supplementary Table 2.15 presents results by main industries

Table 2.10: Summary Statistics for Main Variables

	(1)	(2)	(3)	(4)	(5)	(6)
				Distribution		
VARIABLES	Obs	Mean	Std. Dev.	10th	50th	90th
Panel A: ABS (1980-2015)						
R&D stock ($mm)	57,837	425	2,339	1.04	34.64	540
R&D expenditure ($mm)	57,837	95	498	0.41	7.99	126
Tobin's Q	57,837	5	6	0.36	1.80	20
Assets ($mm)	57,837	1,612	9,687	1.99	56.34	2,239
Sales ($mm)	57,837	2,143	11,054	2.62	116.23	3,465
Patent stock (ABS)	57,837	256	1,621	1.00	11.00	300
Patent flow (ABS)	57,837	23	135	0.00	1.00	33
Panel B: NBER '06 & ABS Match (1980-2006)						
R&D stock ($mm)	43,448	304	1,675	0.84	26.84	402
R&D expenditure ($mm)	43,448	69	365	0.35	6.19	94
Tobin's Q	43,448	4	6	0.33	1.70	17
Assets ($mm)	43,448	1,179	6,970	2.06	47.73	1,709
Sales ($mm)	43,448	1,688	8,424	2.95	100.94	2,818
Patent stock (ABS)	43,448	170	962	1.00	9.00	211
Patent flow (ABS)	43,448	18	89	0.00	1.00	28
Patent stock (NBER '06)	43,448	147	918	0.00	5.00	154
Patent flow (NBER '06)	43,448	15	85	0.00	1.00	20
Panel C: NBER '01 & NBER '06 & ABS Match (1980-1999)						
R&D stock ($mm)	29,075	242	1,335	0.65	19.94	323
R&D expenditure ($mm)	29,075	56	293	0.31	4.79	77
Tobin's Q	29,075	4	6	0.27	1.39	13
Assets ($mm)	29,075	991	4,997	2.26	44.74	1,621
Sales ($mm)	29,075	1,522	6,844	3.51	102.98	2,857
Patent stock (ABS)	29,075	131	649	1.00	7.00	180
Patent flow (ABS)	29,075	16	70	0.00	1.00	26
Patent stock (NBER '06)	29,075	112	616	0.00	4.00	132
Patent flow (NBER '06)	29,075	13	67	0.00	1.00	19
Patent stock (NBER '01)	29,075	84	528	0.00	0.00	91
Patent flow (NBER '01)	29,075	10	59	0.00	0.00	13

Notes: This table presents summary statistics of the main variables examined in the paper.

Table 2.11: Tobin's Q and Patent Stock: NBER '01, NBER '06 and ABS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable:	ln(Tobin's Q)								
Dataset(sample years):	NBER '01			NBER '06			ABS		
	Pooled	All	Mismatched	Pooled	All	Mismatched	Pooled	All	Mismatched
Patent stock$_{t-2}/Assets_t$	0.006	0.043	0.010	0.072**	0.044**	0.028	0.090**	0.049**	0.043*
	(0.013)	(0.029)	(0.046)	(0.007)	(0.015)	(0.034)	(0.006)	(0.013)	(0.018)
R&D stock$_{t-2}/Assets_t$	0.162**	0.124**	0.123**	0.143**	0.117**	0.121**	0.122**	0.104**	0.100**
	(0.004)	(0.011)	(0.015)	(0.005)	(0.012)	(0.015)	(0.005)	(0.012)	(0.017)
Industry FE	Yes	No	No	Yes	No	No	Yes	No	No
Firm FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Variable Avg.	3.365	3.365	3.205	3.365	3.365	3.205	3.365	3.365	3.205
Firms	2764	2764	1162	2764	2764	1162	2764	2764	1162
Observations	19,897	19,897	9,463	19,897	19,897	9,463	19,897	19,897	9,463
R-squared	0.41	0.73	0.72	0.42	0.73	0.72	0.42	0.73	0.72

Notes: The sample consists of firm-year level observations matched on cusip-year pair across NBER '01, NBER '06, and ABS samples. The sample for mismatched-firm analysis includes firms whose cumulative patents in ABS is at least 10% above or below the number in NBER '06. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by cusip. ** p<0.01, * p<0.05

Table 2.12: Tobin's Q and Patent Stock: NBER '06 and ABS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	ln(Tobin's Q)							
Dataset(sample years):	NBER '06 ('80-'06)			ABS ('80-'06)			ABS('80-'15)	ABS('07-'15)
	Pooled	All	Mismatched	Pooled	All	Mismatched	All	All
Patent stock$_{t-2}/Assets_t$	0.060**	0.031**	0.016	0.080**	0.042**	0.042**	0.043**	0.056**
	(0.004)	(0.009)	(0.019)	(0.004)	(0.008)	(0.012)	(0.007)	(0.014)
R&D stock$_{t-2}/Assets_t$	0.137**	0.119**	0.123**	0.117**	0.107**	0.103**	0.110**	0.133**
	(0.003)	(0.007)	(0.010)	(0.003)	(0.008)	(0.011)	(0.006)	(0.015)
Industry FE	Yes	No	No	Yes	No	No	No	No
Firm FE	No	Yes	Yes	No	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable Avg	4.125	4.125	3.757	4.125	4.125	3.757	4.348	5.001
Firms	3644	3644	1567	3644	3644	1567	4190	1861
Observations	31,439	31,439	15,810	31,439	31,439	15,810	43,470	9,111
R-squared	0.43	0.71	0.69	0.43	0.71	0.69	0.69	0.84

Notes: The sample consists of firm-year level observations matched on GVKEY-year pair between NBER '06 and ABS samples. The sample for mismatched-firm analysis includes firms whose cumulative patents in ABS is at least 10% above or below the number in NBER '06. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by GVKEY. ** p<0.01, * p<0.05

2.6.4 Patent production function

If measurement error in patents is systematically correlated with R&D stock, measurement error will also bias the estimates of elasticity of patenting with respect to R&D in patent production function estimates. We estimate the following patent production function to assess how the elasticity of patents with respect to R&D expenditures changes across the different samples. As before, to understand the implications of measurement error, we use firm-year observations common across all datasets, so that changes in the composition of the sample are not at work.

$$\ln(Patents)_{it} = \beta_0 + \beta_1 \ln(R\&D\ stock_{it-2}) + \beta_2 \ln(Assets)_{it-2} + \boldsymbol{\eta_i} + \boldsymbol{\tau_t} + \epsilon_{it} \tag{2.10}$$

Our interest is the coefficient $\hat{\beta}_1$. As discussed, a primary source of measurement error is in patents themselves. Unlike the Tobin's Q estimates, here, measurement error is in the dependent variable, which results in larger standard errors of the estimate, or a bias that depends directly on the correlation between the measurement error and some independent variable, or both. For patents flows, the most likely source of error is mismatching (e.g., due to name changes or other issues in matching). Incomplete dynamic reassignment mostly affects lagged patent stocks rather than the flow of patents.

Table 2.13 presents the estimation results.[18] Columns 1-6 compare NBER '01, NBER '06, and ABS for the period 1980-99. Comparing Columns 1, 3, and 5, we see that the estimated elasticity of patents with respect to R&D stock is similar across the datasets for the entire sample. However, considering only the mismatched firms, we see a marked increase in the measured estimated elasticity, consistent with a negative correlation between the error and R&D: Firms for which NBER failed to

[18] Supplementary Table 2.16 presents pooled results.

match patents are also likely to have higher R&D stock. This is consistent with the results in Table 2.4.[19]

2.7 Summary and conclusions

This paper reports on updates and improvements to the well-known NBER patent database. We extend the database from 2006 to 2015. We also improve the accuracy of the matches by accounting for changes in company names, and changes to corporate boundaries through mergers and acquisitions of firms or their subsidiaries. Doing so enables us to improve the accuracy of the match between firms and the patents they own. This results in an increase in the number of patents we match to sample firms. It also enables us to dynamically assign patent stocks to firms as the patent assignees change owners. We find that approximately 40% of the sample firms are mismatched firms – whose cumulative patents for the entire sample period in ABS is at least 10% above or below that in NBER '06.

We explore the implications of the measurement error for two important relationships that the literature has investigated. The first is the market value of patents, and the second is the patent production function, which relates patent flow to R&D investment. We provide a simplified framework to guide intuition about possible bias. We find that measurement error results in modest under-estimates of patent value. We also find that estimates of the elasticity of patenting with respect to R&D are also under-estimated, especially for firms where measurement error is significant.

[19] Table 2.4 showed that absolute difference between ABS patents and NBER '06 patents is positively related to R&D stock. Because matching errors are mostly errors of omission rather than commission, the absolute difference is also the difference itself. If ABS is the true patent count, then Table 2.4 is the measurement error, implying a positive correlation between measurement error and R&D stock. If y_{it} is the true patent flow for firm i in year t and the error-laden patent flow is $y*_{it} = y_{it} + m_{it}$, then the estimated patent production function is $y*_{it} = \boldsymbol{x}_{it}\alpha + \epsilon_{it}$. This can be rewritten as $y_{it} = \boldsymbol{x}_{it}\alpha + (\epsilon_{it} - m_{it})$. The regression error, therefore, contains the negative of the measurement error. Therefore, a positive correlation between the measurement error and x implies a negative correlation between the regression error and x, and hence a negative bias in the estimate of α.

Overall, these results are good news in that existing estimates reported in the literature are not biased in any significant way. However, additional research would be required to evaluate the implications for more restricted samples (where the fraction of firms affected by error may be larger). In addition, analyses of trends over time, based on the older data, may need to be re-examined, insofar as the incidence of measurement error may have increased over the sample period.

Table 2.13: Patent Production Function Estimates Across Datasets

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
DV:	ln(1+Number of patents)											
	'80-'99						'80-'06				'80-'15	'07-'15
	NBER '01		NBER '06		ABS		NBER '06		ABS		ABS	
Sample:	All	Mismatch	All	Mismatch	All	Mismatch	All	Mismatch	All	Mismatch	All	All
ln(R&D stock)	0.164**	0.092**	0.250**	0.140**	0.301**	0.231**	0.271**	0.139**	0.339**	0.255**	0.345**	0.283**
	(0.025)	(0.026)	(0.031)	(0.034)	(0.030)	(0.035)	(0.028)	(0.028)	(0.025)	(0.029)	(0.024)	(0.050)
ln(Assets)	0.068**	0.042**	0.102**	0.064**	0.127**	0.122**	0.095**	0.063**	0.119**	0.118**	0.121**	0.064**
	(0.012)	(0.010)	(0.013)	(0.014)	(0.014)	(0.018)	(0.010)	(0.011)	(0.010)	(0.014)	(0.009)	(0.012)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DV Avg.	10.088	6.314	13.597	8.652	15.947	13.597	15.059	9.722	17.933	15.264	23.285	39.548
Firms	3188	1295	3188	1295	3188	1295	3935	1660	3935	1660	4347	2220
Obs.	28,785	13,146	28,785	13,146	28,785	13,146	42,949	20,805	42,949	20,805	57,082	14,133
R-squared	0.90	0.89	0.82	0.79	0.85	0.84	0.79	0.75	0.84	0.83	0.83	0.93

Notes: The sample used for 1980-1999 consists of firm-year level observations matched on cusip-year pair across NBER '01, NBER '06, and ABS. The sample used for years 1980-2006 consists of firm-year level observations matched on GVKEY-year pair between NBER '06 and ABS. The sample used for 1980-2015 and 2007-2015 consists of firm-year observations from ABS. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firm. ** $p<0.01$, * $p<0.05$

2.8 Supplementary Results

FIGURE 2.1: Patents assigned to U.S. HQ public corporations and their related subsidiaries

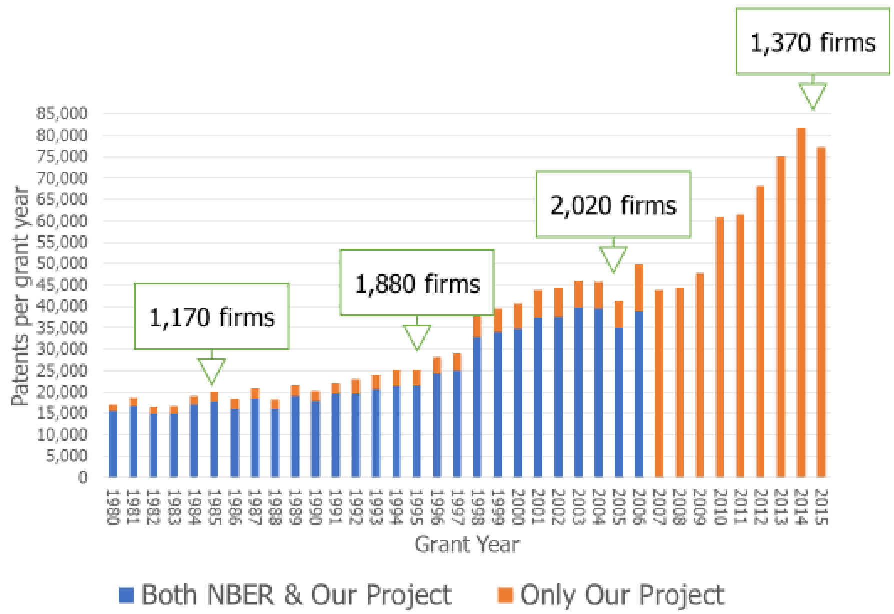


Table 2.14: Top 50 Mismatched Firms: ABS minus NBER '06 (1980-2006)

GVKEY	Company	Average annual difference in matched patents	Total difference in matched patents	Avg Tobin's Q
11436	CBS CORP -OLD	331	6,626	1.3
7475	MOBIL CORP	262	4,973	0.7
29356	VERITAS SOFTWARE CORP	218	1,525	10.5
5693	HONEYWELL INC	183	3,474	1.0
9947	SPERRY CORP	154	921	0.2
8549	CONOCOPHILLIPS	148	3,997	0.8
7435	3M CO	139	3,762	3.0
11636	XEROX CORP	130	3,498	1.0
9653	SHELL OIL CO	119	712	0.8
31166	GUIDANT CORP	118	1,416	10.3
2285	BOEING CO	109	2,935	0.9
14380	PIONEER COS INC -CL A	101	1,216	0.9
157858	FREESCALE SEMICONDUCTOR INC	95	286	2.1
2991	CHEVRON CORP	91	2,454	1.0
7536	PHARMACIA CORP	90	2,073	1.8
142953	MONDELEZ INTERNATIONAL INC	88	526	1.3
25279	BOSTON SCIENTIFIC CORP	85	1,282	7.5
4060	DOWDUPONT INC	83	2,244	1.0
5047	GENERAL ELECTRIC CO	83	2,228	2.0
11040	PHARMACIA & UPJOHN INC	81	1,624	2.1
6266	JOHNSON & JOHNSON	79	2,142	4.3
116526	CONEXANT SYSTEMS INC	72	575	3.1
8543	ALTRIA GROUP INC	69	1,865	1.6
3532	CORNING INC	68	1,834	2.0
4510	FMC CORP	68	1,833	0.7
157415	TRW AUTOMOTIVE HOLDINGS CORP	68	203	0.3
5860	ITT INC	67	1,819	0.9
1608	AMP INC	61	1,165	2.8
4503	EXXON MOBIL CORP	55	1,493	1.3
6774	LOCKHEED MARTIN CORP	54	1,466	0.8
9203	ROCKWELL AUTOMATION	54	1,451	1.1
15708	ALLERGAN INC	53	961	7.8
4087	DU PONT (E I) DE NEMOURS	53	1,424	1.3
1478	WYETH	50	1,353	4.0
4462	NEWMARKET CORP	49	1,335	0.9
15855	SYMANTEC CORP	48	669	8.2
29849	WESTERN ATLAS INC	45	180	1.1
7679	NALCO CHEMICAL CO	44	836	2.8
166414	ALEXZA PHARMACTCLS INC	43	43	16.7
60979	WABTEC CORP	40	474	2.0
10301	TRW INC	38	844	0.8
65399	MERITOR INC	38	380	0.4
11038	UNOCAL CORP	38	941	1.1
9372	ST JUDE MEDICAL INC	37	963	7.6
28742	BORGWARNER INC	37	518	0.8
134932	ON SEMICONDUCTOR CORP	37	256	2.3
3810	DAY INTERNATIONAL INC	36	250	0.1
8488	APPLIED BIOSYSTEMS INC	35	943	4.1
24205	INTL SPECIALTY PRODUCTS INC	35	383	1.0
10983	UNITED TECHNOLOGIES CORP	33	890	0.8

Table 2.15: Tobin's Q and Patent Stock by Main Industries

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	ln(Tobin's Q)						
Years:	1980-1999			1980-2006		1980-2015	1980-2017
Patent Data	NBER '01	NBER '06	ABS	NBER '06	ABS	ABS	ABS
Panel A: Chemicals and Life Sciences							
Patent stock$_{t-2}/Assets_t$	-0.017	0.017	0.031	0.016	0.027*	0.037**	0.032*
	(0.035)	(0.018)	(0.017)	(0.011)	(0.011)	(0.008)	(0.016)
R&D stock$_{t-2}/Assets_t$	0.105**	0.097**	0.087**	0.105**	0.097**	0.113**	0.172**
	(0.017)	(0.017)	(0.018)	(0.011)	(0.011)	(0.009)	(0.022)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable sample average:	5.496	5.496	5.496	6.792	6.792	7.442	9.132
Number of firms	532	532	532	727	727	893	480
Observations	3,802	3,802	3,802	6,486	6,486	9,479	2,285
R-squared	0.83	0.83	0.83	0.79	0.79	0.77	0.85
Panel B: ICT							
Patent stock$_{t-2}/Assets_t$	0.471**	0.168**	0.130**	0.071**	0.071**	0.070**	0.117*
	(0.138)	(0.061)	(0.050)	(0.024)	(0.021)	(0.018)	(0.053)
R&D stock$_{t-2}/Assets_t$	0.141**	0.130**	0.113**	0.135**	0.125**	0.117**	0.065
	(0.025)	(0.026)	(0.029)	(0.014)	(0.015)	(0.013)	(0.034)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable sample average	4.923	4.923	4.923	5.449	5.449	5.225	4.732
Number of firms	334	334	334	592	592	731	336
Observations	2,018	2,018	2,018	3,952	3,952	5,855	1,423
R-squared	0.72	0.72	0.72	0.65	0.65	0.63	0.81
Panel C: Semiconductors							
Patent stock$_{t-2}/Assets_t$	0.052	0.053	0.054	0.036	0.046*	0.068**	0.096**
	(0.070)	(0.046)	(0.037)	(0.027)	(0.023)	(0.019)	(0.033)
R&D stock$_{t-2}/Assets_t$	0.124**	0.115**	0.100**	0.114**	0.101**	0.097**	0.068
	(0.033)	(0.033)	(0.038)	(0.021)	(0.024)	(0.017)	(0.042)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable sample average	2.899	2.899	2.899	3.447	3.447	3.453	3.504
Number of firms	485	485	485	618	618	699	307
Observations	3,290	3,290	3,290	5,146	5,146	7,117	1,500
R-squared	0.67	0.67	0.67	0.64	0.64	0.62	0.79

Notes: The sample used for 1980-1999 consists of firm-year level observations matched on cusip-year pair across NBER '01, NBER '06, and ABS. The sample used for years 1980-2006 consists of firm-year level observations matched on GVKEY-year pair between NBER '06 and ABS. The sample used for 1980-2015 and 2007-2015 consists of firm-year observations from ABS. Panel A consists of firms (at the firm-year level) that operate in Chemicals and Life Sciences, Panel B consists of firms (at the firm-year level) that operate in IT & Telecommunications, and Panel C consists of firms (at the firm-year level) that operate in Semiconductors. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firm. ** p<0.01, * p<0.05

Table 2.16: Patent Production Function Across Datasets - Pooled

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	ln(1+Number of Patents)				
	1980-1999			1980-2006	
	NBER '01	NBER '06	ABS	NBER '06	ABS
ln(R&D stock)	0.160**	0.278**	0.322**	0.288**	0.331**
	(0.015)	(0.014)	(0.013)	(0.011)	(0.011)
ln(Assets)	0.131**	0.099**	0.124**	0.092**	0.119**
	(0.010)	(0.010)	(0.009)	(0.008)	(0.008)
Industry fixed-effects	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes
Dependent variable sample average	5.453	8.087	9.636	8.792	10.710
Number of firms	3188	3188	3188	3935	3935
Observations	3,188	3,188	3,188	3,935	3,935
R-squared	0.39	0.50	0.63	0.51	0.62

Notes: The sample used for 1980-1999 consists of firm-year level observations matched on cusip-year pair across NBER '01, NBER '06, and ABS. The sample used for years 1980-2006 consists of firm-year level observations matched on GVKEY-year pair between NBER '06 and ABS. Standard errors (in brackets) are robust. ** p<0.01, * p<0.05

3

Knowledge Spillovers and Corporate Investment in Scientific Research

This chapter is adapted from a joint work with Ashish Arora and Sharon Belenzon that is published in American Economic Review (Arora et al., 2021a). All authors have equal contribution. For the original article please see:
https://doi.org/10.1257/aer.20171742

3.1 Introduction

Although economists often speak of R&D as a single construct, it is useful to distinguish between research ("R") and development ("D"). Research is an input into invention: scientific discoveries may lead to new products and processes, but even when inventions do not directly arise from a scientific discovery (Kline and Rosenberg, 1986; David et al., 1992), research enhances the efficiency of inventive activity (Nelson, 1959).[1] Research is typically thought of as being performed by universities and funded by the government, but many significant scientific breakthroughs have

[1] In Vannevar Bush's words, "New products and new processes do not appear full-grown. They are founded on new principles and new conceptions, which in turn are painstakingly developed by research in the purest realms of science." (Bush, 1945, p.241)

come from scientists working not in universities but in the labs of companies such as Du Pont, ICI, Merck, Xerox, IBM, and AT&T. In 2018, the business sector in the United States invested about $337 billion in internal R&D. Of this, $67.5 billion, or a fifth, was spent on research ("R"), with development ("D") accounting for the rest. This $67.5 billion represents nearly a third of all research conducted in the United States that year, which amounted to $211.5 billion.[2]

The two components of R&D differ in another respect as well. Whereas new products and processes can be protected from potential imitators by patents, copyrights, and trade secrecy, research is typically disclosed in scientific publications, even when it is conducted by corporate scientists. Research, therefore, is more likely than development to generate knowledge spillovers (e.g., Dasgupta and David (1994a); Arrow (1962); Nelson (1959)). When these spillovers accrue to rivals, they are not merely externalities; they may actually *reduce* private returns from research. Indeed, Rosenberg (1990) posed the question of why firms invest in research in the first instance. He suggested that, in addition to sometimes producing commercially valuable findings, research investments enabled firms to benefit from academic science. The literature has offered several other possible mechanisms for private returns from research. Common features of these explanations are that they do not require the firm to use the research for innovation, and that spillovers to other firms do not reduce returns from research.[3]

In this paper, we focus on how private returns to corporate research depend on

[2] R&D performed by the business sector in 2018, including that financed by the government and other sources, amounted to $422 billion, with research accounting for $91.8 billion. *Source*: National Science Foundation, National Center for Science and Engineering Statistics 2019, Tables 2, 3, 4. Available at https://ncses.nsf.gov/pubs/nsf20307, last accessed July 20th, 2020.

[3] These explanations include absorptive capacity (Cohen and Levinthal, 1989; Cockburn and Henderson, 1998; Griffith et al., 2004, 2006; Aghion and Jaravel, 2015), incentives for high-skilled scientist-inventors (Stern, 2004; Henderson and Cockburn, 1994; Audretsch, 1996; Cockburn and Henderson, 1998), and signaling to investors, prospective customers or regulators (Azoulay, 2002; Hicks, 1995).

the balance between two opposing forces: the benefits from the use of science in own inventions, and the profit-reduction from knowledge outflows to rivals.[4] Because spillovers, as used in the literature, include knowledge inflows from other firms, to avoid confusion we henceforth use the term “spillouts” to refer to knowledge outflows to rivals.

Our analysis includes all publicly traded American manufacturing firms with at least one year of positive R&D expenditures and at least one patent over the period 1980-2015. The final sample consists of an unbalanced panel of 3,807 firms and 53,110 firm-year observations. We measure the use of internal research in invention by citations made by the firm’s patents to its own scientific publications. We measure spillouts by citations from the patents of rivals to the focal firm’s publications.

With these newly constructed data, we present two main findings. First, we show that there is a positive relationship between the market value of a firm and its stock of scientific output. This relationship is stronger when the firm’s patents use the science that the firm’s scientists produce. Conversely, a firm’s stock of scientific output is less valuable to the firm when rivals use its science. Second, and consistent with this, we find that a firm produces more scientific publications if it is more likely to use the science in its patents, but produces fewer publications if the science is more likely to be used by rivals’ inventions.

Although we do not claim that our estimates of these relationships are causal, the patterns of association are consistent with the notion that firms obtain greater value from scientific research when they are able to use it for their inventions, but the value is reduced when knowledge is used by rivals. The relationships endure

[4] Xerox’s Palo Alto Research Center (PARC) illustrates the tradeoff. Xerox’s failure to commercialize PARC discoveries, which were ultimately used by companies such as Apple, Microsoft, and 3Com, is frequently cited as a reason for PARC’s ultimate demise. Yet, the benefits Xerox obtained from PARC’s research in areas that were closer to Xerox’s core business, such as the laser printer, were substantial. These inventions allowed the firm to recoup its investment in PARC despite the spillovers, at least for a time.

even after controlling for firm fixed effects, as well as a variety of time-varying firm characteristics. We also present estimates where we instrument for citations by rivals using tax credits as instruments for patenting by rivals, following Bloom et al. (2013) and Lucking et al. (2018).

Our work connects to two streams of research in the economics of innovation literature. One stream of prior work, (e.g., Mansfield (1980) and Griliches (1986)) has relied on confidential data to distinguish between research and development. Using a sample of approximately 1,000 large manufacturing firms from 1957 through 1977, Griliches (1986) found that firms that spent a larger share of R&D on basic research were substantially more productive. In a more recent paper, Akcigit et al. (2013) use confidential data on French firms to distinguish between basic and applied research. They argue that spillovers from basic research are broader than those from applied research.

Instead of using confidential data on R&D inputs, we use publicly available data on outputs, namely publications and patents. This enables us to trace knowledge flows from research to innovation using patent citations to corporate publications. Therefore, we can explore the tradeoff between internal use and use by rivals in a firm's decision to invest in research. Despite these differences, our empirical results are consistent with the findings of this literature. Consistent with Griliches (1986), we find that research is privately valuable, in part because research enhances the productivity of innovation. Further, consistent with Akcigit et al. (2017), we find that knowledge spillovers are more likely to be associated with publications rather than patents.

The second literature has focused on knowledge spillovers. Building on Jaffe (1986), who measures spillovers using external R&D, Bloom et al. (2013) (hereafter, BSV) distinguish between the R&D expenditures of product-market rivals and technology neighbors. The latter potentially represent beneficial knowledge-inflows that improve the focal firm's innovation outcomes. However, the R&D expenditures of product-market rivals, though they may also have benefits, have a rent-stealing effect as well: an increase in the knowledge base of competitors hurts the focal firm in the product market.

We build on BSV, with three main differences. First, we consider research, which is upstream of invention, as an input into invention. Research is often disclosed in scientific publications, and is a more potent source of spillovers than inventions, which are protected by patents. Second, we focus on knowledge outflows or spillouts, as opposed to knowledge inflows. That is, we examine how the use of its knowledge by outsiders affects the focal firm, rather than how a focal firm is affected by knowledge produced by other firms.[5] Unlike knowledge spillovers in general, which benefit other firms but do not directly affect the firm producing the knowledge, spillouts to rivals directly reduce the rents from innovation. Spillouts are, therefore, an indirect cost of research. Third, we introduce a direct measure of spillouts. While previous work typically measured potential knowledge flows using R&D performed by other firms, we measure knowledge flows directly as patent citations to science produced by a focal firm.

Our paper contributes to the ongoing policy discussions on the apparent decline in inventiveness (Bloom et al., 2017) and the associated slowdown in productivity growth. If inventions build on science, particularly corporate science, then a decline in corporate science may be implicated in the declining novelty of inventions. Figure 3.1 shows that the share of basic and applied research in the total domestic R&D funded and performed by the business sector has declined from over 31 percent in

[5] Our paper is closer to Belenzon (2012), which examines the relationship between the private value of a firm's patent and the citations it receives from the firm's own patents as well as the patents of its rivals. In related work, Ceccagnoli (2005) investigates a model where some firms invest in R&D that can spill out to rivals, who may not invest in R&D. However, Ceccagnoli (2005) does not trace spillouts, and does not distinguish between research and development.

1986 to about 20 percent in 2008, and has remained at that level thereafter.[6] The share of research in total R&D performed by business, including that financed by the government and other sources, shows a similar decline, from a peak of around 30 percent in 1991 to about 20 percent in 2008, albeit rising modestly thereafter. That is, the composition of business R&D appears to be shifting away from research.

Many leading American firms began to withdraw from research in the 1980s (Mowery, 2009). Their research labs were shut down or were oriented towards more applied activities. Bell Labs was separated from its parent company AT&T and placed under Lucent in 1996, Xerox PARC was spun off into a separate company in 2002, and Du Pont closed its Central Research and Development Laboratory in 2016. These accounts are consistent with Figure 3.2, which presents trends in the annual number of publications ("R") and patents ("D") divided by sales, for our sample firms. The corporate publication rate fell by about 60% over the sample period, whereas the patenting rate does not show any clear trend. This suggests that the composition of corporate R&D is changing over time, with less "R" and more "D" (Arora et al., 2018). Changes in the balance between internal use and spillouts may be related to the declining share of research in corporate R&D. If firms become more sensitive to their research leaking out to rivals, then they focus on research projects with the highest likelihood of internal use, cutting back on more broad-ranging research initiatives.

Our paper also contributes by developing new data and measures. We create and validate a new measure of use of science in invention. Although some recent work has also used patent citations to science, ours is the first large scale study measuring the flow of corporate science to corporate invention, within and across

[6] Based on Tables 2-4, National Science Foundation, National Center for Science and Engineering Statistics 2019. National Patterns of R&D Resources: 2017–18 Data Update. NSF 20-307. Alexandria, VA. Available at https://ncses.nsf.gov/pubs/nsf20307.

firm boundaries, for a period of over a third of a century.[7] We match publication records from Web of Science to front-page non-patent literature (NPL) references and link both to Compustat firms. In so doing, we also improve and extend the NBER patent database, adjusting for changes in corporate names and ownership. The outcome is a more accurate and comprehensive match between firms and their stock of patents and publications, which accounts for changes in names, and for mergers, acquisitions, and divestitures. This process is described in section 3.3 below.

The chapter proceeds as follows. Section 3.2 presents the analytical framework that guides our empirical investigation. Section 3.3 discusses the data, section 3.4 outlines the econometric specifications, and section 3.5 summarizes the results. Section 3.6 concludes with a discussion of how trends in spillouts and internal use may be related to the changes in the composition of corporate R&D.

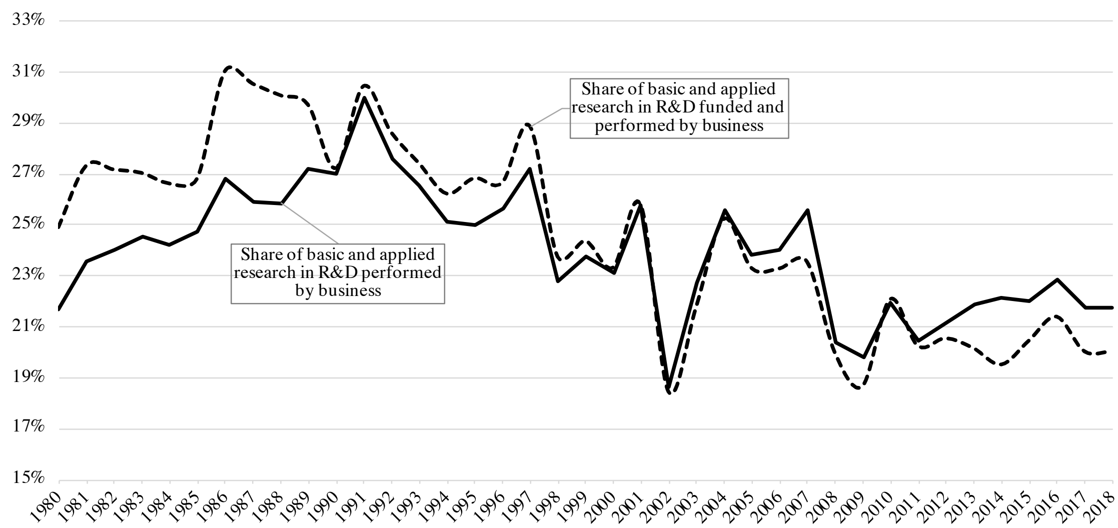
3.2 Analytical framework

We outline a simple framework to motivate the empirical analysis. We follow the framework in BSV but differ in three important respects. First, we distinguish between research and innovation. Research reduces the cost of innovations, and innovations increase profits. Second, we focus on knowledge "spillouts" from research; inventions are protected by patents and are assumed not to generate "spillouts". Finally, whereas BSV study how spillovers are beneficial externalities that enhance the efficiency of R&D, our focus is on how spillouts to rivals reduce private returns from research, and therefore also the incentives to invest in research.

Consider two firms, indexed by 0 and 1. Both invest in innovation, d_0 and d_1,

[7] Patent citations are imperfect measures of knowledge flow, but Roach and Cohen (2013) judge patent to publication citations to be better sources of tracking flow of scientific knowledge than patent to patent citations, which have been more extensively used in the literature (e.g., Jaffe et al. (1993)). In recent work, Azoulay et al. (2019) use patent citations to track the flow of National Institutes of Health (NIH) funded research to biomedical inventions, and Bryan et al. (2020) compare front-page citations to in-text citations to a fixed set of journals. We also validate our measure using the Carnegie Mellon Survey of R&D performing firms.

Figure 3.1: Research in Business R&D, 1980-2018

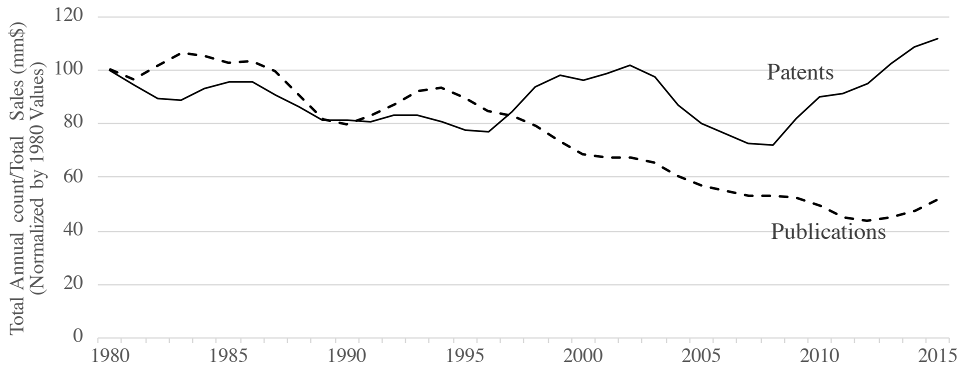


Data for this plot are generated from Tables 2-4, National Science Foundation, National Center for Science and Engineering Statistics 2019. National Patterns of R&D Resources: 2017–18 Data Update. NSF 20-307. Alexandria, VA. Available at https://ncses.nsf.gov/pubs/nsf20307.

respectively, to compete in the product market. For simplicity, only firm 0 invests in research. Research by firm 0 reduces its own cost of innovation, but also spills out to the rival firm, reducing the rival's cost of innovation. There are three stages. In stage 3, the firms compete in the product market. Each firm's product market profit depends upon its own innovation output and that of the rival. The reduced form profit functions are $\Pi_0(d_0, d_1)$ and $\Pi_1(d_0, d_1)$. In stage 2, firms choose their innovation output. The cost of innovation for firm 0 is $\phi(r_0; \lambda)d_0$, where r_0 is the investment in research by firm 0, and λ represents internal absorptive capacity, or the ability to learn from internal research. The innovation cost for firm 1 is $s(r_0; \theta)d_1$, where θ reflects the ability of the rival to learn from "spillouts" from firm 0's research. In stage 1, firm 0 chooses its research investment, r_0, to maximize $v_0 = \Pi_0(d_0, d_1) - \phi(r_0; \lambda)d_0 - \gamma(r_0)$, where $\gamma(r_0)$ is the direct cost of research.[8]

[8] If firm 1 also invests in research, then ϕ and s also depend on r_1. In this case, the results that follow should be interpreted as holding r_1 constant (i.e., characterizing firm $0's$ reaction function from stage 1).

Figure 3.2: Trends in Corporate Scientific Publications and Patents, 1980-2015



The figure presents the trend in publications and patents over time. The figures are measured by total annual count divided by the sum of annual sales (mm$) across our complete sample firms from 1980 through 2015. Both lines start at a normalized index value of 100 for 1980 and future years are measured relative to 1980. The sample for publications is conditional on firms with at least 1 publication stock.

Table 4.1 summarizes the effects of spillouts on value, research, and innovation, if Π_0 is additively separable and concave. The first set of results relate to value. Table 3.5 empirically examines whether research is more valuable if it is used by the firm, but less valuable if it is used by rivals. The same tradeoff also applies to investment in research, but with one important difference: investment in research depends on the marginal returns to research. The marginal return to research, $\frac{\partial v_0}{\partial r_0} = -d_0 \frac{\partial \phi}{\partial r_0} + \frac{\partial \Pi_0}{\partial d_1} \frac{\partial d_1}{\partial r_0} - \frac{\partial \gamma}{\partial r_0}$, has three terms. The first term represents how research lowers the cost of innovation. This term is positive and represents the benefit from research. The second term represents the impact of spillouts to rivals. We expect the rival's innovation to increase because it benefits from knowledge spillouts, so that spillouts to rivals are an indirect cost of doing research. The final term represents a direct marginal cost of research.

In general, internal use increases the marginal return to research, whereas spillouts decrease it. However, strategic complementarity in stage 2 between internal and rival innovation can potentially offset these direct effects. For instance, internal use increases internal innovation, thereby increasing marginal returns to research. However, if rival innovation also increases because of strategic complementarity, this indirect cost of research reduces marginal return. That is, the same underlying mechanisms may result in different relationships when we compare how value and research respond to spillouts and internal use. We explore these relationships empirically in Table 3.6.

Strategic interactions are particularly relevant for innovation outcomes. In their absence, internal use increases innovation and spillouts decrease innovation. However, under strategic complementarity, spillouts may even increase innovation. In Table 3.7, we explore empirically the extent to which the use of scientific knowledge makes innovative activity more productive, and how spillouts affect innovation.

Table 3.1: Predicted Relationships

VARIABLE	VALUE	RESEARCH	INNOVATION
Spillout to rivals	Decrease	Decrease	No effect
Internal use	Increase	Increase	Increase

3.3 Data and Main Variables

We combine data from six sources: (i) company and accounting information from S&P North America Compustat (Standard & Poor's, 2018b); (ii) scientific publications from Web of Science (WoS) by Clarivate (Clarivate Analytics, 2016); (iii) patent and non-patent literature (NPL) citations from PATSTAT (European Patent Office, 2016); (iv) subsidiary data from ORBIS (Bureau van Dijk, 2018); (v) acquisition data from SDC Platinum (Securities Data Company Platinum, 2018); and (vi) company name changes from WRDS's "CRSP Monthly Stock" (Center for Research in Security Prices, 2018b,a).

We re-construct the NBER data (Hall et al. (2001), Bessen (2009)) from 1980 and extend it to 2015 while introducing several improvements to accommodate changes in corporate names and ownership structures. We use scientific publications as our measure of the production of scientific knowledge and patents as our measure of inventive activity. We treat a citation by a patent to a corporate publication as an indicator that the patented invention used the knowledge in the publication. For this purpose, we also develop new data on corporate publications matched to NPL citations, which we discuss below.

3.3.1 Accounting panel data

We start with all North American Compustat records and select companies with positive R&D expenses for at least one year during our sample period of 1980-2015. We exclude firms that are not headquartered in the United States and firms without patents. As in Bloom et al. (2013), we further restrict the sample to manufacturing firms. Our final sample consists of an unbalanced panel of 3,807 firms and 53,110 firm-year observations.

Approximately 30% of the Compustat firms in our sample changed their name at least once, making it challenging to match publication and patent data. Accounting for name changes is difficult because there is no single source that tracks different names of the same firm, and to the best of our knowledge, this has not been done previously on a large scale. We identify name changes in two ways: (i) we link Compustat records to WRDS's "CRSP Monthly Stock" file, which records historical names for each month a security is traded, and (ii) perform extensive manual checks using Securities and Exchange Commission (SEC) filings to verify all related names for our sample period.

The second major challenge comes from ownership changes. A parent company and a majority-owned subsidiary may have different identification numbers and records in Compustat. Moreover, a single company may correspond to multiple firm identifiers due to changes in ownership (such as mergers, acquisitions, and spinoffs). We identify ownership structures and ownership changes in three ways. First, we match our sample firms to ORBIS ownership files for the years 2002-2015 for annual subsidiary information (using each publication year as a separate "snapshot" of ownership structure).[9] Second, for firms that exit Compustat before 2002, we manually collect subsidiary names based on SEC filings and rely on the NBER patent database for pre-2002 ownership data. Third, we match our firms to M&A data from SDC Platinum to supplement information on ownership changes.

3.3.2 Main variables

Corporate publications

We match our sample firms to the Web of Science database. We include articles from journals covered in the "Science Citation Index" and "Conference Proceedings Citation Index - Science", excluding social sciences, arts, and humanities articles. Using the affiliation field and all historical company names, we identify approximately 800 thousand articles published between 1980 and 2015 that have at least one author employed by our sample of Compustat firms or their majority-owned subsidiaries at the time.

Corporate patents

We match patents to our sample of Compustat firms and their subsidiaries. We account for firm name changes as well as M&A reassignment of patents based on SDC and ORBIS data. As with publications, when the ownership of the patenting entity changes, the patents associated with the entity are reallocated to the new owner. We match approximately 1.3 million patents to our sample firms and their

[9] The year 2002 is the first year with reliable coverage of ownership information in ORBIS.

subsidiaries.

Patent citations to corporate publications

We match non-patent literature (NPL) citations to publications as our measure of the use of corporate science in invention. Using all patents granted in the period 1980-2015, we perform a many-to-many match between NPL citations and WoS publications (approximately 10 million citations matched to 800 thousand corporate publications), allowing for more than one publication to be matched to each citation. For each possible match, we construct a score that captures the degree of textual overlap between the free-text NPL format and the structured WoS record, which includes the following fields: article title, journal, and authors. To exclude mismatches, we use a more detailed secondary matching algorithm that is based on different WoS fields: standardized authors' names, number of authors, article title, journal name, and year of publication. The matching algorithm accounts for misspelling, unstructured text, incomplete references, and other issues that may cause mismatches. We manually verify the accuracy of the matches. We then focus on citations made by our sample of corporate patents. This process resulted in 70 thousand unique corporate publications cited by 140 thousand unique corporate patents.[10]

3.3.3 Descriptive statistics

Our main sample and variables are at the firm-year level. Table 3.2 presents descriptive statistics for our main variables over the sample period. Our sample includes a wide distribution of firm sizes, with market value ranging from 6 million dollars (10th percentile) to 4 billion dollars (90th percentile) and sales ranging from 3 million

[10] Papers and patents are matched "dynamically." For instance, if a sample firm merges with another firm, then the patents of the merged firm are included in the stock of patents linked to the Compustat record from that point onward, but not before. Most importantly, we can identify more accurately an internal or external citation based on the owner of the citing patent and the affiliation of the author for the cited paper at the time the paper was published.

dollars (10th percentile) to 3.7 billion dollars (90th percentile). About 70% of firms have at least one publication during the sample period (all sample firms have at least one patent). These firms produce, on average, 19 publications per year. The distribution of publications is skewed, with the median firm producing one publication per year. We observe a similar pattern for patents, with an average of 24 patents per firm-year and a median of 2 patents.

Table 3.2: Summary Statistics for Main Variables

					Distribution		
VARIABLE	# Obs.	# Firms	Mean	Std. Dev.	10th	50th	90th
Scientific publications count	41,664	2,781	19	101	0	1	22
Scientific publications stock	41,664	2,781	92	537	0	4	95
Patents count	53,110	3,807	24	138	0	2	36
Patents stock	53,110	3,807	111	644	1	7	158
R&D expenditures ($mm)	53,110	3,807	96	501	0.41	8	124
R&D stock ($mm)	53,110	3,807	391	2,353	0.5	21	432
Market value ($mm)	53,110	3,807	3,381	21,351	6	130	3,962
Tobin's Q	53,110	3,807	4	6	0	2	17
Sales ($mm)	53,110	3,807	2,253	11,470	3	119	3,640
Assets ($mm)	53,110	3,807	1,684	10,065	2	58	2,315

Notes: This table provides summary statistics for the main variables used in the econometric analysis. The sample is at the firm-year level and includes an unbalanced panel of 3,807 U.S. headquartered publicly traded manufacturing ultimate owner parent companies (of which 2,781 are publishing companies) over the sample period of 1980-2015.

Table 3.3 provides an additional descriptive analysis of citation patterns. Among the 2,781 publishing firms, 734 cite their own publications in their patents at least once, and 984 produce publications that are cited by other firms. Cited publications receive substantially more external citations than internal citations (10.3 vs. 4.6). Yet, the number of external citations drops sharply when accounting for the product market proximity between the citing and cited firms (from 10.3 to 3.7), indicating that a substantial portion of spillouts are unlikely to be harmful to the focal firm.

In an additional analysis (not reported in the table), we find that publications that are cited internally are almost ten times more likely to receive an external citation. Publications that are cited by the firm's own patents receive 1.1 external citations, compared to only 0.1 external citations for publications that are not internally cited. Furthermore, we find that firms with an above-mean ratio of internal citations to total citations received have more productive R&D programs (measured by the number of publications and patents per dollar of R&D) and are more R&D intensive (measured by R&D expenditures over sales). These patterns are consistent with our main premise that, although research may spill out to rivals, as long as the benefit of internal use offsets the private cost of spillouts, firms might have sufficient incentives to invest in scientific research.

Table 3.3: Summary Statistics for Patent Citations to Corporate Publications

	(1)	(2)	(3)	(4)
VARIABLE	Number of firms with positive values	Citations per firm-year	Number of citing patents per firm-year	Number of cited publications per firm-year
Patent citations, all	1,123	14.98	9.82	7.67
Internal patent citations	734	4.64	2.33	2.78
External patent citations, corporate	984	10.33	7.49	5.27
External patent citations, RIVAL	975	3.7	2.58	2.08

Notes: This table provides summary statistics for patent citations to scientific publications by our sample firms. The sample is at the firm-year level and includes only firms with at least one publication that is cited by a patent (either internally or by a corporate patent). RIVAL measures citations from product market rivals.

3.3.4 Validating patent citations to scientific articles as a measure of use of science in innovation

We use patent citations to scientific publications to measure the use of knowledge. Although patent citations are widely used, they are also widely criticized as imperfect measures of knowledge flows (Jaffe and Trajtenberg, 2002; Duguet and MacGarvie, 2005; Roach and Cohen, 2013). Roach and Cohen (2013) point out, however, that patent to publication citations, though imperfect, are much better than patent to patent citations at tracing knowledge flows, especially from public research to firms. Our interest is in knowledge flows from corporate research to other firms. To validate our measure of use of science—NPL citations to scientific articles—we use the Carnegie Mellon Survey (CMS) data on industrial R&D (Cohen, Wesley M., Richard R. Nelson, and John P. Walsh, 2000). As part of the survey, lab directors in R&D performing firms were asked about the extent to which their R&D projects used scientific knowledge from various sources. Of the firms in our sample, 772 are also covered in the CMS, with a total of 28,318 patents granted between 1991 and 1999.

Table 3.4 confirms that firms whose patents cite scientific publications also reported that science contributed to their R&D projects, even after controlling for firm size, number of backward patent citations to other patents, and four-digit industry SIC code dummies. Furthermore, the fields of science that contribute the most to a firm are also those whose publications the firm's patents cite, and firms that draw on public science also tend to cite public science in their patents. Column 4 is especially important. It documents a strong relationship between our measure of patent citations to corporate science and the reported value of other firms' research as an input into own innovation.[11]

[11] We thank Michael Roach and Wesley Cohen for providing the Carnegie Mellon Survey data to us.

Table 3.4: Citations to Science as Measures of Use: Supporting Evidence from the Carnegie Mellon Survey

	Dependent variable: CMS questions				
	(1)	(2)	(3)	(4)	(5)
Response to CMS questions:	Importance of public research findings (Q.18)		Importance of main research field's findings (Q.22)	Importance of other firm's research findings (Q.16)	Basic research share (Q.45)
Citations to top 200 *universities articles*	0.337				
	(0.146)				
Citations to public science articles		0.246			1.821
		(0.120)			(0.697)
Citations to articles in main research field			0.148		
			(0.065)		
Citations to corporate articles				0.453	
				(0.161)	
Citations to patents	0.001	0.001	-0.002	-0.003	-0.043
	(0.006)	(0.006)	(0.005)	(0.007)	(0.037)
$ln(Sales)$	0.078	0.074	0.040	-0.016	0.023
	(0.032)	(0.034)	(0.020)	(0.027)	(0.174)
Industry dummies	Yes	Yes	Yes	Yes	Yes
Observations	555	555	495	555	557
R-squared	0.39	0.39	0.46	0.41	0.39

Notes: This table presents OLS estimation results for the relationship between average patent citations to publications per patent and the 1994 Carnegie Mellon survey (CMS) response (Cohen, Wesley M., Richard R. Nelson, and John P. Walsh, 2000) related to the importance of research findings as an input to the firm's R&D projects. Columns 1 & 2 are based on CMS question 18: "During the last three years, what percentage of your R&D unit's projects made use of the following research outputs produced by universities or government research institutes and labs?". Column 3 is based on CMS question 22: "Referring to the fields listed above, indicate the field whose research findings in general (not just university and government research) contributed the most to your R&D activities during the last three years". Column 4 is based on CMS question 16: "Below are some sources of activities or information on the R&D activities or innovations of other firms in your industry. Please score each of following in terms of the importance of that information's contribution to a recently completed major project". Column 5 is based on CMS question 45: "Approximately what percentage of your R&D effort is: a. Basic Research; b. Applied Research; c. Design and/or Development; d. Technical Service.". The sample includes only patenting firms. In Column 3, the sample is restricted to firms that indicated their main research field in question 22 (excluding 'Others' category). For Citations to articles in main research field, publications were classified to research fields based on the Web of Science journal subject category. Citations to corporate articles include citations to publications by our main sample of Compustat firms. Citations to patents include backward citations to patents. The unit of observation is a firm. Robust standard errors in parentheses.

3.4 Econometric framework

The analytical framework in section 3.2 provides two sets of predictions. First, that research would increase the value of the firm to the extent it is used internally, but that it would be less valuable to the extent that it spills out to rivals. Second, and consistent with this, the firm would produce less research if it is more likely to spill out to rivals, and more research if the firm is more likely to use it internally. We turn next to the empirical investigation of these predictions.

3.4.1 Market value equation

We follow Bloom et al. (2013) and their predecessors (Griliches (1986) and Hall et al. (2005)) and estimate the following Tobin's Q specification (bold indicates vector representation):

$$\begin{aligned} \ln \frac{Value_{it}}{Assets_{it}} =& \alpha_0 \frac{G_{it-1}}{Assets_{it-1}} + \alpha_1 \ln \left(\textit{Cumulative internal use}_{it-1} \right) \\ &+ \alpha_2 \ln \left(\textit{Cumulative SPILLOUT}_{it-1} \right) \\ &+ \boldsymbol{Z}'_{it-1} \boldsymbol{\gamma} + \boldsymbol{\eta_i} + \boldsymbol{\tau_t} + \epsilon_{it} \end{aligned} \tag{3.1}$$

Tobin's Q is market value over assets. G is knowledge assets, measured as the perpetual stocks of R&D, publications, and patents. The use of internal science in innovation is measured by the cumulative number of citations made by the focal firm's patents to its own publications. More precisely, internal use counts all citations made by the focal firm's patents to its own research published up to and including year $t-1$ (including citations received in years greater than t to those publications). Similarly, spillout is measured as the cumulative number of *rivalry-weighted* external citations to the focal firm's publications. To account for rivalry, we follow BSV and measure product-market rivalry as the Mahalanobis similarity of vectors representing the shares of industry segment sales (Standard & Poor's, 2018a) for each pair of firms, labeled as SEG. Citations received by firm i from firm j are weighted by SEG_{ij}, the "distance" of citing firm j from the cited focal firm i in the product market.[12] If internal use is valuable (see Table 4.1 in Section 3.2), we expect $\hat{\alpha_1} > 0$, and if spillouts reduce the value of research, we expect $\hat{\alpha_2} < 0$.[13]

While we focus on *spill-outs*, the earlier literature (e.g., Jaffe (1986); Bloom et al. (2013)) has stressed *spill-ins* or incoming knowledge flows. To facilitate comparison with that literature, we also present specifications that control for potential incoming knowledge flows. Accordingly, $\boldsymbol{Z}$ is a vector of controls, including the sum of stocks of R&D, patents, and publications by other firms weighted according to the proximity of these firms to the focal firm in the product and technology spaces. As in Bloom et al. (2013), $SPILLSIC_{it}$ is the sum of weighted R&D by product market rivals and is computed as $\sum_j SEG_{ij} \times GRD_{jt}$. GRD_{jt} is the perpetual R&D stock of a potential rival firm j. Similarly, $SPILLTECH_{it}$ is the sum of outsiders' R&D stock weighted by the technology distance, computed as $\sum_j TEC_{ij} \times GRD_{jt}$. Technology proximity, TEC_{ij}, is measured analogously as the Mahalanobis similarity of vectors

[12] *Cumulative* $SPILLOUT_{it-1} = \sum_{j=1}^{N} n_{ij} \times SEG_{ij}$, where n_{ij} is the number of citations from patents of firm j to publications by firm i published up to year t-1 (inclusive), and SEG_{ij} is the product market proximity between the two firms. To get SEG_{ij} we follow BSV's procedure and weight the share of firm $i's$ sales in industry segment s (defined by 4-digit SIC codes) by the market share of firm i in industry segment s (industry segments are from Compustat's operating segments database). Define W_i as the vector, whose individual component w_{is} is the share of segment s in firm $i's$ total sales, multiplied by the share of firm i in the total sales in segment s. For example, if firm i and firm j have similar sales shares across operating segments, the proximity score of the firms would be high. The Mahalonobis distance allows industry relatedness to be firm-specific by accounting for how dominant each firm is in an industry, with higher weights assigned for more dominant competitors. For instance, if firm i and firm j's sales both account for a large share of total industry sales in industry segment A, then segment A would be given high weight in determining the proximity score between i and j. The proximity between firm i and firm j is the vector cosine similarity: $SEG_{ij} = \dfrac{W_i' \cdot W_j'}{\|W_i'\| \|W_j'\|}$.

[13] If a firm invests in research to signal quality to regulators and customers or to attract talented researchers, citations of its publications by others would validate its claims to quality and reinforce the signal. That is, external citations, rather than representing profit-reducing spillovers, would increase profits. Similarly, higher-quality research, which is more likely to garner citations, would be positively related to profits. Thus, if citations by rivals are negatively related to value, this strongly suggests that spillouts of knowledge to rivals reduce profits from research.

representing the shares of patents across 4-digit international patent classes (IPC) for each pair of firms, i and j. That is, we use the same formulation as for SEG_{ij}, but instead of share of sales by industry segments we use share of patents by IPC. Finally, η_i and τ_t are firm and year fixed effects, while ϵ_{it} is the $i.i.d$ error term.

3.4.2 Publication equation

The relationship between scientific research, internal use, and spillout is specified as follows:

$$\begin{aligned}\ln(Publications_{it}) =& \beta_0 + \beta_1 \ln(Internal\ use_{it-1}) + \beta_2 \ln(SPILLOUT_{it-1}) \\ &+ \boldsymbol{Z'_{it-1}\gamma} + \boldsymbol{\eta_i} + \boldsymbol{\tau_t} + \epsilon_{it}\end{aligned} \tag{3.2}$$

Our coefficients of interest are β_1 for internal use and β_2 for spillouts to rivals. From Section 3.2 we expect $\hat{\beta}_1 > 0$ and $\hat{\beta}_2 < 0$, respectively. Here, internal use is constructed as the lagged number of patent citations made by firm i's patents granted at year $t-1$ to its own scientific articles published up to year $t-1$ (inclusive).[14] Similarly, spillout is the equivalent measure with SEG-weighted external citations. The key difference in how we measure internal use and spillout in the publication equation as compared to the market value equation is the exclusion of future citations in the publication equation (that is, citations to publications published up to year $t-1$ by patents granted in years greater than t). The reason for excluding future citations is to mitigate the concern that common technological shocks may drive the relationship between publications, internal use and spillout. Current and future shocks to research opportunity can affect both focal publications and the number of citing patents by the focal firm and its rivals. Focusing on *pre-determined* citations mitigates this concern.

[14] Our results are not sensitive to including higher-order citation lags, or to including $future$ patent citations to existing publications (citations to articles published up to year $t-1$ by patents that are granted at years greater than t).

Another issue is a potential bias due to scale. Firms with more publications and patents will tend to have more internal citations, leading to an upward bias in $\hat{\beta}_1$. To mitigate this concern, all specifications include firm fixed effects as well as controls for firm scale, such as patent and R&D stocks. We also present results when citations by rivals are instrumented using variation in state R&D taxes (see Section 3.4.4 below), and control for the R&D by product-market rivals and by technology neighbors (as in the Tobin's Q specification).

3.4.3 Patent equation

We estimate the following patent equation to test our premise that internal and external research are related to innovation.

$$
\begin{aligned}
\ln(Patents)_{it} =& \omega_0 + \omega_1 Publications\ stock_{it-1} + \omega_2 \ln(Internal\ use_{it-1}) + \\
& \omega_3 \ln(SPILLOUT_{it-1}) + \omega_4 \ln(Citations\ to\ rivals_{it-1}) \\
& + \omega_5 \ln(R\&D\ stock_{it-1}) + \boldsymbol{\eta_i} + \boldsymbol{\tau_t} + \epsilon_{it}
\end{aligned}
\tag{3.3}
$$

In Equation 4.2, $Patents_{it}$ is the flow of patents for firm i in year t. The framework presented in Section 3.2 assumes that the use of science, be it internal or external, would reduce the cost of innovation. To measure the use of external science by the focal firm, *Citations to rivals* represents the average number of citations to rivals' publications per patent of the focal firm granted at year $t-1$.[15] Internal use is measured as the number of citations to internal publications, per patent of the focal firm granted in year $t-1$. If the use of science leads to more innovation, we expect $\hat{\omega}_2 > 0$ and $\hat{\omega}_4 > 0$. However, internal research can also enhance innovation indirectly, such as by redirecting the firm's innovative activities to more promising avenues (Nelson, 1982), or by attracting talented inventors (Stern, 2004). This indirect effect

[15] Section 3.2 had ignored research by rivals. Empirically, we explore how the use of rivals' research conditions innovation by the focal firm by allowing its innovation to also depend on the use of external science.

is represented by ω_1, the coefficient of the stock of publications.

3.4.4 Instrumental variable strategy

Investment in research and citations made by external patents may be driven by common unobserved time-varying effects, leading to an upward bias in the OLS estimate of the spillout coefficients $\hat{\alpha_2}$ and $\hat{\beta_2}$ in the Tobin's Q and publication equations, respectively. For instance, if a particular line of scientific inquiry becomes economically promising, research in that field may receive more citations as other firms ramp up inventive activity in that field. Similarly, an expansion in demand may increase research by the focal firm as well as its use by others. Formally, a firm and its rivals may have common shocks to the marginal benefits of R&D. These common shocks would result in a positive correlation between the research conducted by the focal firm and the patenting output of its rivals, and hence, between the research conducted by the focal firm and the citations the research receives from patents filed by rivals.

We follow Bloom et al. (2013) and Lucking et al. (2018) and use state-level variation in tax credits as an instrument for rivals' patenting.[16] In effect, our IV strategy is to use the variation in the cost of R&D as a source of exogenous variation in inventive activity (i.e., patents). R&D tax credits affect the marginal cost of R&D, but not the benefit. Therefore, they offer a source of variation in R&D that is independent of the confounding variation. For each sample firm, we calculate its cost of R&D and regress the number of patents against this cost. The predicted number of patents from this regression is used as our input into calculating a focal firm-specific aggregate number of predicted patents by its rivals, where the aggregation is based on the weighting procedures discussed in Section 3.4.1. The aggregate rival patents

[16] The coefficient estimate of internal use remains vulnerable to an upward bias from such common shocks.

are used as our instrument for $SPILLOUT$ in the different equations.

We implement the IV approach by first projecting our patent count variable on both the state and the federal tax credit components of R&D user costs. Next, we calculate the predicted value of logged patent-count using the regression estimates, $\hat{\pi}_{it}$. For each firm i, we compute $Riva\hat{l}PAT_{it} = \sum_i SEG_{ij}\hat{\pi}_{jt}$, where SEG_{ij} is the distance in product space between firm j and focal firm i, using the Mahalanobis distance described earlier. Finally, we use $Riva\hat{l}PAT_{it-2}$ as an instrument for *Rival citations*$_{it-1}$. This procedure follows Bloom et al. (2013) and is used because $SPILLOUT_{it-1}$ is itself constructed by weighting citations from firms by the Mahalonobis distance.

3.5 Estimation results

3.5.1 Market value equation

Table 3.5 presents the estimation results for market value. Column 1 shows that the coefficient of R&D stock is 0.11. Column 2 adds publication and patent stocks, and shows a stronger relationship of value with patents than with publications. Column 3 adds internal use and spillout. As expected, $\hat{\alpha_1} > 0$ and $\hat{\alpha_2} < 0$. Both estimates are statistically different from zero (p-value< 0.05).

Column 4 presents the second stage, where we instrument for spillout.[17] The estimates indicate that the value of an additional internal citation is offset by 2.2 citations made by rivals.[18] The coefficient estimate of spillout is significantly larger

[17] Column 1 in Supplementary Table 3.9 presents the first stage results of regressing spillout against $Riva\hat{l}PAT$. More patenting by rivals leads to more citations by these rivals to the focal firm's publications. Supplementary Table 3.8 presents the estimation results of regressing rival patents, which generates our instrument $Riva\hat{l}PAT$, on rival cost of R&D (Column 1). There is a strong negative relationship between rival R&D cost and rival patenting, and with the R&D expenditures of rivals (Column 2).

[18] Average values for internal use and spillout are 29 and 32, respectively. The marginal effect of an additional rival citation, evaluated at the sample mean, is $-0.006 = (-0.175 \times 4/33) \times 0.3$ (mean SEG value is 0.3, mean Tobin's Q is 4, and 33 is one plus average value of spillout). The same calculation for internal use is $0.013 = (0.096 \times 4/30)$.

in magnitude than the coefficient in the previous specifications. One possible explanation for the higher IV estimate is that rival citations may reflect the quality of the firm's science as well as spillouts. We expect that the former is positively associated with value, whereas the latter is negatively associated, with the resulting OLS estimate being negative but smaller in magnitude. Another possible source of bias is common shocks. For example, shocks to demand or technical opportunities would result in a positive correlation between patent citations received from rivals and the firm's own value. Overall, we find that private returns to research are positively related to its internal use in invention, but negatively related to its external use in invention by rivals.

The literature on knowledge spillovers has used aggregate R&D by firms overlapping in technology (Jaffe, 1986), and by firms competing in the product market (BSV) to proxy for potential incoming knowledge flows, or spill-ins. Accordingly, we include these measures in Columns 5-9. Following BSV, these are labeled $SPILLTECH$ and $SPILLSIC$, respectively. By so doing, we also account for the knowledge use that patent citations to publications may miss; others may benefit from a firm's research without necessarily citing it in their patents. The results in Columns 5-7 show that the use of a firm's scientific knowledge by rivals, as measured by citations, continues to be negatively related to value, even after controlling for external R&D by rivals and other related firms. Although the coefficient of citations by rival patents drops in magnitude (citations by rival patents is positively related to R&D by rivals), it remains comparable to the coefficient of citations by internal patents, and statistically different from zero. However, R&D by rivals and by firms operating in similar technology fields is negatively related to value, indicating that knowledge spill-ins might be offset by other rent reducing effects, such as more intense product market competition, or preemption in the technology space.

To study these opposing effects of external R&D, we replace aggregate external

R&D with disaggregated measures of research and innovation by other firms, using their publications and patents, in Columns 8 and 9. If publications are a source of external knowledge that firms can use freely in their own inventions, then we would expect external publications (whether authors by product market rivals or technology neighbors) to be positively related to value. Patents also disclose knowledge. However, patent holders have claimed property rights over some or all of this knowledge. External patents potentially preempt the focal firm from inventing in the related technical space, or presage forthcoming innovations that might reduce the focal firm's profits.[19] Therefore, external patents may be negatively related to value.

The results in Columns 8 and 9 are consistent with these conjectures. In Column 8, the coefficient of publications by rivals, SPILLSIC-PUB, is positive, whereas the coefficient of rival patents, SPILLSIC-PAT is negative. Similarly, in column 9, the coefficient of publications by firms operating in neighboring technology areas, SPILLTECH-PUB, is positive (but statistically not different from zero), whereas the coefficient of patents, SPILLTECH-PAT is negative. Overall, these results reinforce the basic premise of this paper that the different constituents of R&D, namely research and development, have very different economic properties. Research, which is upstream, is the more potent source of spillovers. Innovation by rivals, on the other hand, is more likely to be covered by means of appropriation, such as intellectual property rights or secrecy, and thus, also more likely to have a market-stealing component.

[19] For instance, the patenting firm may expand into the focal firm's market or license to entrants into the focal firm's market, or the focal firm may have acquired licenses to some of those patents.

Table 3.5: Stock Market Value and Patent Citations to Corporate Science

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent variable:					ln(Tobin's Q)				
	R&D	"R" vs "D"	Internal use and spill-outs	2nd Stage IV, Rival	Spill-ins, BSV	Spill-ins, Jaffe	Spill-ins, Jaffe and BSV	SPILLSIC, "R" vs "D"	SPILLTECH, "R" vs "D"
$ln(Cumulative\ internal\ use)_{t-1}$			0.043	0.096	0.047	0.054	0.052	0.051	0.048
			(0.021)	(0.030)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
$ln(Cumulative\ SPILLOUT)_{t-1}$			-0.083	-0.175	-0.059	-0.065	-0.057	-0.053	-0.057
			(0.021)	(0.041)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
$Publication\ stock_{t-1}/Assets$		0.018	0.020	0.023	0.019	0.020	0.020	0.015	0.015
		(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
$Patent\ stock_{t-1}/Assets$		0.061	0.060	0.056	0.057	0.057	0.057	0.057	0.058
		(0.007)	(0.007)	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)
$R\&D\ stock_{t-1}/Assets$	0.105	0.067	0.069	0.084	0.075	0.075	0.076	0.080	0.078
	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
$ln(SPILLSIC,\ GRD)_{t-1}$					-0.345		-0.195		-0.335
					(0.060)		(0.079)		(0.076)
$ln(SPILLTECH,\ GRD)_{t-1}$						-0.563	-0.369	-0.616	
						(0.089)	(0.119)	(0.103)	
$ln(SPILLSIC,\ PUB)_{t-1}$								0.235	
								(0.072)	
$ln(SPILLSIC,\ PAT)_{t-1}$								-0.356	
								(0.064)	
$ln(SPILLTECH,\ PUB)_{t-1}$									0.130
									(0.109)
$ln(SPILLTECH,\ PAT)_{t-1}$									-0.313
									(0.085)
Weak identification				F=698.05					
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Tobin's Q sample average	4	4	4	4	4	4	4	4	4
Number of firms	3,653	3,653	3,653	3,383	3,653	3,653	3,653	3,653	3,653
Observations	43,432	43,432	43,432	39,861	43,432	43,432	43,432	43,432	43,432
R-squared	0.68	0.68	0.68	0.69	0.68	0.68	0.68	0.69	0.68

Notes: This table presents estimation results for the relationship between Tobin's Q with internal use and spillouts. SPILLSIC is the product market distance weighted sum of all other firms' R&D/Publications/Patents stocks (as appropriate). SPILLTECH is the technology-distance weighted sum of all other firms' R&D, Publication, or Patent stocks (as appropriate). All specifications include a dummy variable that receives the value of one for observations where lagged publications stock is equal to zero; and a dummy variable that receives the value of one for observations where lagged R&D stock is equal to zero. One is added to logged control variables. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

3.5.2 Publications equation

The tradeoff between internal use and spillouts will be reflected in the decisions to invest in research as well. However, the production of research depends not on the average return, but rather the marginal returns to research. As discussed in Section 3.2, an increase in internal use has a direct effect of increasing the marginal return, and an indirect effect that depends on how the rival responds. If there is strategic substitutability, the indirect effect reinforces the direct effect through an increase in the firm's investment in innovation. With strategic complementarity, the indirect effect is in the opposite direction. Similarly, an increase in spillouts has a direct effect of reducing the marginal return to research, but strategic complementarity results in an offsetting indirect effect. Put differently, the relationship of spillouts and internal use with research can be empirically different from the relationship with value, though both sets of relationships reflect the same economic forces.

Table 3.6 presents the estimation results for publications. As expected, $\hat{\beta}_1$ is positive and statistically significant (Column 1), indicating that firms produce more publications if their past publications were used internally. Not all citations to science are equally relevant for investing in research. We expect internal citations to be more relevant to a firm's decision to invest in research when the cited publication (i) is more recent (the lag between the grant year of the patent and publication year of the article is shorter) and (ii) is cited by the firm's more valuable patents. These predictions are confirmed in Columns 2 and 3. Column 2 distinguishes between citations to old and new science. Internal citations to new science consist of citations to articles published no earlier than five years from the grant year of the citing patent, and citations to all earlier articles are treated as citations to old science. Only the coefficient estimate of citations to recent science is positive and statistically significant (the estimates on new and old science are statistically different from each other with a p-value < 0.05). Column 3 distinguishes between citations made by high and low quality patents.[20] The coefficient estimate for high-quality patents is positive and statistically significant, while the estimate for low-quality patents is statistically zero (the estimates are statistically different from each other with a p-value < 0.05).

In summary, Columns 1-3 are consistent with the view that patent citations to own science that matter for the production of future science are citations that come from high-quality patents of the sponsoring firm to recent publications. These relationships bolster the view that scientific output is an input into downstream inventive activity, and that to justify investment in research, managers need to demonstrate that the knowledge produced is useful for the downstream inventive activity of the sponsoring firm.

Column 4 adds spillout. The coefficient estimate $\hat{\beta}_2$ is negative and statistically significant. A firm whose research spills over to rivals is likely to reduce its investment. The direct effect of spillout is to reduce the payoff from research by increasing innovation by rivals. However, if innovation strategies are strategic complements, an increase in innovation by rivals would induce the focal firm to increase innovation as well, which increases the marginal payoff to research. Empirically, it appears that the direct effect outweighs the potentially offsetting indirect effect.

Columns 5-6 alleviate concerns that our results are driven by firm-year observations with zero publications. Column 5 presents estimates from a Negative Binomial publications count specification with pre-sample fixed effects (5-year pre-sample average number of publications) following Blundell et al. (1999). Column 6 presents results for inverse hyperbolic sine transformation. Our main results remain robust.

Column 7 presents estimates from instrumenting for spillout with $Riva\hat{l}PAT$.[21]

[20] Patent quality is based on the number of citations a patent receives divided by the average number of citations received by all patents granted in the same year as the focal patent. Patents are classified into high and low quality using the median value from the corporate patents sample.

[21] Column 2 of Supplementary Table 3.9 presents the first stage results of regressing spillout against

The IV estimate of spillout is larger in magnitude, indicating a larger negative effect of rival citations on focal publications. This is consistent with the results from Table 3.5. Based on the IV estimates, the positive contribution of an additional internal citation is offset by 8 external citations.[22]

Columns 8-10 add controls for potential spill-ins. Similar to our findings from Table 3.5, the coefficient of publications by rivals, $SPILLSIC-PUB$, is positive, whereas the coefficient of rival patents, $SPILLSIC-PAT$ is negative (but statistically not different from zero). In column 10, the coefficient of publications by firms operating in similar technology areas, $SPILLTECH-PUB$, and the coefficient of patents, $SPILLTECH-PAT$ are both positive. This is a point of difference from the market value estimates, which show a negative estimate for the coefficient of $SPILLTECH-PAT$. A possible explanation is that, although more patenting by firms operating in similar technology fields hurts the profits of the focal firm, strategic complementarity in the innovation stage results in more patents and, consequently, more research by the focal firm.[23] That is, the effect of $SPILLTECH-PAT$ on the average return to research is negative, while its effect on the marginal value of research is positive.

$Riva\hat{l}PAT$, with the expected sign.

[22] The mean for internal use is 14, and for spillout is 75. An additional rival citation lowers publications by $0.014 = (-0.229 \times 15/76) \times 0.3$ (0.3 is average SEG value–the contribution of an additional citation by a rival). An additional internal citation increases publications by $0.117 = (0.125 \times 15/16)$.

[23] In unreported specification, we confirm that $SPILLTECH-PAT$ is positively related to research.

Table 3.6: Publications and Patent Citations to Corporate Science and Publication Output

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable:	ln(1+Number of publications)				Number of publications	Inverse hyperbolic sine	ln(1+number of publications)			
	Internal citations	New vs. Old science	High vs. Low Quality Patents	Rival and internal citations	Negative Binomial, Pre-sample controls	OLS	2nd Stage IV, Rival	Spill-ins, Jaffe and BSV	SPILLSIC, "R" vs "D"	SPILLTECH, "R" vs "D"
$ln(Internal\ use)_{t-1}$	0.056			0.087	0.150	0.113	0.125	0.085	0.084	0.084
	(0.017)			(0.017)	(0.027)	(0.021)	(0.022)	(0.017)	(0.017)	(0.017)
$NEW\ publications$		0.132								
		(0.018)								
$OLD\ publications$		-0.028								
		(0.018)								
$High\ quality\ citing\ patents$			0.073							
			(0.019)							
$Low\ quality\ citing\ patents$			-0.019							
			(0.019)							
$ln(SPILLOUT)_{t-1}$				-0.075	-0.075	-0.071	-0.229	-0.082	-0.085	-0.083
				(0.025)	(0.027)	(0.031)	(0.049)	(0.026)	(0.026)	(0.025)
$ln(SPILLTECH,\ GRD)_{t-1}$								0.120	0.052	
								(0.083)	(0.071)	
$ln(SPILLSIC,\ GRD)_{t-1}$								0.010		-0.045
								(0.053)		(0.053)
$ln(SPILLTECH,\ PUB)_{t-1}$										0.278
										(0.078)
$ln(SPILLTECH,\ PAT)_{t-1}$										0.091
										(0.070)
$ln(SPILLSIC,\ PUB)_{t-1}$									0.143	
									(0.048)	
$ln(SPILLSIC,\ PAT)_{t-1}$									-0.075	
									(0.046)	
$ln(R\&D\ stock)_{t-1}$	0.139	0.139	0.139	0.139	0.328	0.172	0.149	0.136	0.134	0.134
	(0.014)	(0.014)	(0.014)	(0.014)	(0.028)	(0.017)	(0.016)	(0.014)	(0.014)	(0.014)
$ln(Patent\ stock)_{t-1}$	0.084	0.084	0.085	0.089	0.170	0.100	0.114	0.089	0.093	0.088
	(0.014)	(0.014)	(0.014)	(0.014)	(0.023)	(0.017)	(0.015)	(0.014)	(0.014)	(0.014)
Pre-sample FE					0.433					
					(0.024)					
Weak identification							F=364.78			
Firm fixed-effects	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Publications sample average	15	15	15	15	14	15	16	15	15	15
Number of firms	3807	3807	3807	3807	3030	3807	3521	3807	3807	3807
Observations	49,303	49,303	49,303	49,303	34,889	49,303	45,210	49,303	49,303	49,303
R-squared	0.88	0.88	0.88	0.88		0.88	0.89	0.88	0.88	0.88

Notes: This table presents estimation results for the relationship between publications with internal use and spillout. All specifications include a dummy variable that receives the value of one for firms that never published up to the focal year; a dummy variable that receives the value of one for firms without yearly granted patents; and a dummy variable that receives the value of one for firms without annual patent citations to own publications. One is added to logged control variables. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

3.5.3 Patent equation

The framework developed in Section 3.2 assumes that scientific knowledge lowers the cost of innovation (or equivalently, increases the efficiency of investments in innovation). In Table 3.7, we directly explore whether the use of science enhances innovation. As is customary in the literature, we use patents to measure the flow of innovations produced by a firm, controlling for R&D investment. Our interest is in how innovation is related to the focal firm's scientific knowledge, to the use by the focal firm of its own and outside knowledge, and to spillouts of its knowledge to rivals.

Column 1 includes publications stock without controlling for use. As expected, publications are positively related to patenting. Column 2 adds internal use. Interestingly, the coefficient estimate of publications stock does not change by much, indicating that scientific research, in addition to directly producing commercially valuable discoveries (as captured by patent citations to publications), may also enhance innovation in other ways. For instance, investing in research may guide innovation activities into more productive avenues and away from less productive ones (Nelson, 1982; Fleming and Sorenson, 2004), and attract talented inventors (Stern, 2004). Column 3 shows that citations by the focal firm to publications by rivals (spill-ins) are positively related to the patenting output of the focal firm. This result supports the assumption that a firm's innovation cost falls when it can use externally produced knowledge. Column 4 adds spillouts to rivals. Spillouts are positively related to patents, but statistically not different from zero. Column 5 instruments for rival citations using the same instrument used in Table 6. There is a positive effect of rival citations on patenting, larger in magnitude, albeit still not statistically different from zero (Column 3 of Supplementary Table 3.9 presents the first-stage estimation results). Column 6 confirms that the results are similar when controlling for $SPILLSIC$ and $SPILLTECH$.

These results are consistent with the simple story that scientific knowledge enhances innovation. This knowledge may be produced internally but may also be produced by rivals. Knowledge that spills out to rivals reduces value, but how the firm's research and innovation responds to spillouts is more nuanced and also depends upon the nature of strategic interactions. In particular, strategic complementarity in innovation may result in innovation increasing with spillouts while research nonetheless falls.

Table 3.7: Patenting and citations to corporate science

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	ln(1+Number of patents)					
	Publications stock	Internal use per patent	Spill-ins from rivals	Spill-outs	2nd Stage IV, Rival	2nd Stage IV, Rival
$ln(Publications\ stock)_{t-1}$	0.245	0.213	0.207	0.206	0.225	0.232
	(0.024)	(0.023)	(0.023)	(0.023)	(0.025)	(0.025)
$ln(Internal\ use)_{t-1}$		0.208	0.181	0.177	0.151	0.149
		(0.022)	(0.023)	(0.021)	(0.023)	(0.023)
$ln(Citation\ to\ rival\ pub.)_{t-1}$			0.239	0.237	0.223	0.230
			(0.035)	(0.034)	(0.035)	(0.035)
$ln(SPILLOUT)_{t-1}$				0.011	0.082	0.118
				(0.030)	(0.059)	(0.060)
$ln(SPILLSIC,\ GRD)_{t-1}$						0.072
						(0.088)
$ln(SPILLTECH,\ GRD)_{t-1}$						-0.539
						(0.122)
$ln(R\&D\ stock)_{t-1}$	0.298	0.289	0.285	0.285	0.303	0.314
	(0.018)	(0.018)	(0.018)	(0.018)	(0.020)	(0.020)
Weak identification					F=369.3	F=379.0
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Patents sample average	25.43	25.43	25.43	25.43	26.93	26.93
Number of firms	3807	3807	3807	3807	3521	3521
Observations	49,303	49,303	49,303	49,303	45,210	45,210
R-squared	0.85	0.86	0.86	0.86	-	-

Notes: This table presents estimation results of a patent equation that examines the relationship patents with use of science (own and by rivals) and spillouts. All specifications include dummy variables that receive the value of one for firms without citations at the focal year; a dummy variable that receives the value of one for observation where lagged publications stock is equal to zero; and a dummy variable that receives the value of one for observation where lagged R&D stock is equal to zero. One is added to logged control variables. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

3.6 Conclusion and discussion

American firms are investing a smaller share of their R&D budget in research. The decline in corporate participation in science, even as inventions themselves become more dependent on science, is piquant. Even as firms make greater use of scientific knowledge, they themselves are less willing to produce such knowledge, preferring to shift attention and resources from upstream research to downstream development. This shift, though likely privately profitable, is not without social costs. The declining corporate engagement in research may be contributing to the reported decline in R&D productivity and the associated decline in productivity growth (e.g., Bloom et al. (2017)). As we show in this paper, firms invest in research to boost innovation. However, when the research spills out to rivals, leading to an increase in inventions by rivals, the private returns to research fall. The changing tradeoff between the benefits of internal use against the cost of spillouts to rivals may be one reason for the shift in corporate R&D in America, with less "R" and more "D". It is plausible that firms have become more vigilant in ensuring that the knowledge they produce creates value for them. If so, they would also have become more sensitive to rivals using their research. The heightened sensitivity would lead firms to invest in fewer research projects that are more carefully targeted to internal needs, and less likely to spill out to others.

We conjecture that this is part of a broader change in the U.S. innovation ecosystem since the 1980s. In addition to the changes in industrial structure, there have been significant changes in the intellectual property regime, universities have become more active in commercializing research, and startups have become more important sources of innovation (Arora et al., 2020a). The growing division of innovative labor has seen firms increasingly acquiring knowledge and inventions from others through licensing, alliances, and outright acquisitions. Not surprisingly, there is greater attention on knowledge as an asset, and more efforts to monetize it in a variety of ways.[24]

Changes in intellectual property protection may also be at work (Galasso and Schankerman, 2014; Moser, 2005). Here there are contradictory impulses. On the one hand, it is widely acknowledged that intellectual property protection in the United States was strengthened in the 1980s. Consistent with this, there are significant differences across industries in the extent to which firms are withdrawing from research. Life sciences have seen the smallest decline compared to sectors such as materials, chemicals, and information technology, perhaps because profit-reducing spillouts are less likely because intellectual property protection is the strongest in life-sciences (Williams, 2013). On the other hand, the last decade has seen a push back, with several court cases weakening patent protection. Moreover, if scientific findings have become more broadly applicable, then even without changes in the patent regime, the patents filed by a firm may cover a smaller fraction of the applications of its scientific discoveries, more so if firms themselves are more narrowly focused on fewer products.

Indeed, patent citations to papers suggest that the likelihood of a research finding spilling out has increased. Figure 3.3 shows that the propensity of patents to cite corporate science (measured as the ratio between citations to corporate science per patent and total number of available corporate publications in a given year) has been rising over time for both internal and rival citations. But, while the propensities to cite internally and by rivals are identical in 1990, by 2015 rival citations become twice as likely than internal citations. Moreover, such spillouts may also potentially reduce profits if rivalry is more intense (Aghion et al., 2005), although the hypoth-

[24] American corporations reported \$66 billion of income from licensing industrial technology in 2002, and IRS data imply an annual growth of 11 percent between 1994 to 2004, well above the average GDP growth of 3.42 percent over the same period (Robbins, 2009). For 2011, income to American corporations from licensing technology stood at nearly \$82 billion. Patent assignments also increased between 1987 and 2014. See Arora et al. (2020a) for more details.

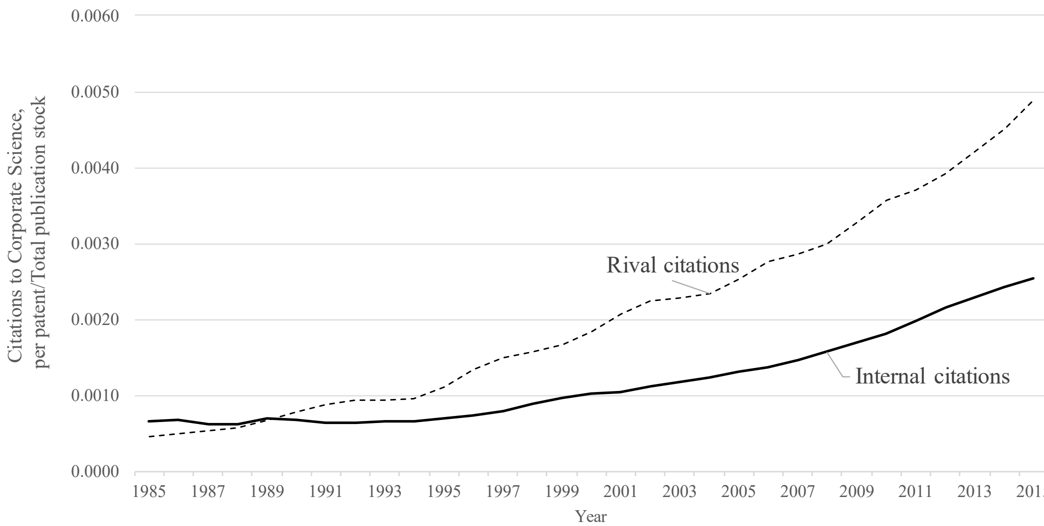
esis of greater competition has to contend with other research pointing to growing market concentration and the rise of superstar firms (Autor et al., 2020; Gutiérrez and Philippon, 2017). [25]

If firms are indeed more concerned with ensuring a return on their investments in research, then they are likely to focus on fewer research projects, more likely to serve clearly identified internal needs, and perhaps narrower in scope and less likely to spillout to others. As a result, even if the threat of spillouts increases, it could well be that spillovers in the aggregate may fall, with deleterious consequences for overall productivity growth. This suggests that understanding how the economic institutions, such as intellectual property and anti-trust laws, affect knowledge-flows across firms, and how the nature of the knowledge itself conditions these relationships, remain important subjects for future research.

[25] The greater rent-destroying potential is pithily illustrated in the following quote from a former Bell Labs researcher (Odlyzko, 1995, p. 4, emphases added):

> Xerography was invented by Carlson in 1937, but it was only commercialized by Xerox in 1950. Furthermore, there was so little interest in this technology that during the few years surrounding commercialization, Xerox was able to invent and patent a whole range of related techniques, while there was hardly any activity by other institutions. [... By contrast] when Bednorz and Mueller announced their discovery of high-temperature superconductivity at the IBM Zurich lab in 1987, it took only a few weeks for groups at University of Houston, University of Alabama, Bell Labs, and other places to make important further discoveries. *Thus, even if high-temperature superconductivity had developed into a commercially significant field, IBM would have had to share the financial benefits with others who held patents that would have been crucial to developments of products.*

Figure 3.3: Trends in Use of Science by Corporate Patents, 1985-2015



This figure presents trends in the propensity of corporate patents to cite corporate science, measured as the ratio between citations to science per corporate patent divided by total number of corporate publications in each year (y-axis values are multiplied by 1,000). The sample is conditional on firm-year observations with at least one granted patent. Rival citations weigh external citations by the product market proximity between the citing and cited firms.

3.7 Supplementary Results

Table 3.8: Predicting Patents and R&D using Federal and State R&D Tax Credit

	(1)	(2)
Dependent variable:	ln(1+Number of patents)	ln(R&D)
ln(Federa Tax Credits)	-2.202	-4.557
	(0.450)	(0.335)
ln(State Tax Credits)	-0.474	-0.389
	(0.128)	(0.101)
Firm fixed-effects	Yes	Yes
Year dummies	Yes	Yes
Joint F-test of the tax credits	F=19.10	F=101.16
Dependent variable sample average	30.20	109.57
Number of firms	3,451	3,451
Observations	42,642	42,6421
R-squared	0.83	0.92

Notes: Data on Federal and State R&D tax credit is based on Lucking, Bloom, Van Reenen (2018). Restricted to firm-years with available data. Standard errors (in brackets) are robust to arbitrary heteroscedasticity.

Table 3.9: Instrumental Variable Estimation (First Stage): Federal and State R&D Tax Credit

	(1)	(2)	(3)
Specification:	Tobin's Q	Publications	Patents
Dependent variable:	$ln(SPILLOUT_{t-1})$	$ln(SPILLOUT_{t-1})$	$ln(SPILLOUT_{t-1})$
	First Stage	First Stage	First Stage
$Predicted\ RIVAL\ Patents)_{t-1}$	0.187	0.104	0.104
	(0.007)	(0.005)	(0.005)
$ln(Internal\ use)_{t-1}$	0.352	0.230	0.224
	(0.022)	(0.016)	(0.016)
$Publication\ stock_{t-1}/Assets$	0.005		
	(0.004)		
$Patent\ stock_{t-1}/Assets$	-0.005		
	(0.003)		
$R\&D\ stock_{t-1}/Assets$	0.007		
	(0.003)		
$ln(R\&D\ stock)_{t-1}$		0.004	0.011
		(0.004)	(0.004)
$ln(Patent\ stock)_{t-1}$		0.015	
		(0.006)	
$ln(Publication\ stock)_{t-1}$			0.007
			(0.006)
$ln(Citation\ to\ rival\ publications)_{t-1}$			0.060
			(0.019)
Firm fixed-effects	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes
Number of firms	3,383	3,521	3,521
Observations	39,861	45,210	45,210

Notes: This table presents first stages of instrumental variable estimations for the effect of SPILLOUT on Tobin's Q, publications and patents. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

4

Sitting on the Fence: integrating the two worlds of scientific discovery and invention within the firm

4.1 Introduction

Throughout the U.S. corporate history, there has been a constant debate on the "appropriate" organization of scientific discovery[1] within firms. Managers have mainly focused on whether integration between research and development practices is more effective than the specialization of research activity (Wise, 1985; Hounshell and Smith, 1988). This debate goes as far back as the early 20th century when large firms established central corporate research labs (e.g., DuPont, GE, Xerox-PARC, and AT&T-Bell Laboratories). For example, Dupont had invested in both the "Eastern Lab", where researchers were working side by side engineers, mainly on applied research related to the firm's immediate product and process improvements needs, as well as in the "Experimental Station" and its Central Research Department (CRD), where scientist initially focused on long-term-fundamental research, separate from

[1] Henceforth, I shall use scientific discovery interchangeably with scientific research or simply research.

the manufacturing and product lines (Hounshell and Smith, 1988).[2]

The decline in central corporate research laboratories, starting in the 1980s (Mowery, 2009; Pisano, 2010; Arora et al., 2018, 2020a) has only made the topic of the organization of scientific discovery within firms more relevant and important. While past research mainly highlighted the benefits of integration of research and development practices to the invention output, in the current paper, I examine both the benefits and the costs of integration, as well as its determinants, in a within-firm analysis.

The literature presents two main views on the organization of scientific discovery.[3] On the one hand, a specialized organization (Smith, 1776), which supports a clear division of research and invention practices to increase productivity.[4] On the other hand, an integrated organization[5], which supports the connectedness and the continuous interaction between research and invention practices to increase productivity (Kline and Rosenberg, 1986; Rosenberg, 1990). As Rosenberg (1990) stated, *"When basic research in industry is isolated from the rest of the firm, whether organizationally or geographically, it is likely to become sterile and unproductive"*. These two views are not mutually exclusive; a firm can decide to implement a combination of

[2] Dupont's CRD later changed its focus towards more applied research based on the needs of its business units, and accordingly, it was renamed the Central Research and Development Department. It continued to operate until 2016, when DuPont merged with Dow Chemical and decided to downsize and reorganize it as the Science & Innovation group, which was more closely aligned with the firm's business units.

[3] While originally the concept of specialization and the division of innovative labor (Smith, 1776; Arora and Gambardella, 1994), was discussed in terms of different firms and organizations specializing in the stages of the innovation process where they have a comparative advantage, this paper takes a within-firm perspective of the topic.

[4] Smith (1776) suggests that productivity is enhanced by increasing dexterity and time saved by avoiding task changing: *"Subdivision of employment in philosophy, as well as in every other business, improves dexterity, and saves time. Each individual becomes more expert in his own peculiar branch, more work is done upon the whole, and the quantity of science is considerably increased by it."* (Smith (1776):12-13)

[5] Integration in this research should not be confused with vertical integration, and with the concept of "Technology integration" (Iansiti, 1997), which relates to the capability of choosing among technological options (i.e., research outputs), and effectively integrating them into an application.

both as well as to change its scientific discovery organization over time. For example, in 2014, after many years of clear separation between its research and product units, Microsoft reassigned half of its research unit to a new group, MSR NEXT. Scientists in this new group work alongside engineers on applied projects with immediate impact on Microsoft's products (e.g., the Skype translator) rather than focusing on basic research initiatives (Bloomberg, 2016)[6]. Similarly, Google has its hybrid AI teams (Spector et al., 2012), where researchers work closely with its product groups and focus on more short-term applied initiatives related to Google's current products, as well as Google Depmind - a separate specialized AI research unit that focuses on basic research initiatives.

Building on Ronald Coase and Oliver Williamson's idea that *"All feasible forms of organization are flawed"*[7], in this research, I examine the tradeoffs and the determinants that are associated with the organization of scientific discovery - either specialized or integrated - within firms.

I demonstrate the short-term benefits of integration in the form of higher invention productivity and the long-term costs of integration in terms of lower research productivity. I find that a one standard deviation increase in integration is related to an increase of 9 patents and to a decrease of 1.3 scientific papers per firm-year. These relationships are stronger for science-based technologies. Building on the view that science is an input to invention (Bush, 1945; Rosenberg, 1990; Narin et al., 1997; Arora et al., 2021a), the cumulative decrease in scientific discovery quality will, in turn, dilute the firm's invention quality and breakthroughs in the long-run - resulting in approximately a 60% reduction in the direct increase in patents due to integration. Finally, the organization of scientific discovery also has value impli-

[6] https://www.bloomberg.com/features/2016-microsoft-research

[7] Interview with Oliver E. Williamson following the announcement of his 2009 Nobel Prize: https://www.nobelprize.org/prizes/economic-sciences/2009/williamson/26025-interview-with-oliver-e-williamson/

cations for the firm. I find that the positive relationship between the market value of a firm and its scientific knowledge stock is weaker with integration, especially for science-based firms and for weak markets for technology. Conversely, the positive relationship between market value and scientific invention stock is stronger with integration, especially for science-based firms and for early-stage technology.

The paper contributes to the organizing for innovation literature that examines the relationship between internal organization and innovative output (Kay, 1988; Argyres and Silverman, 2004; Arora et al., 2014; Argyres et al., 2019; Aggarwal et al., 2020). R&D organizational structure, and more specifically, the organization of scientific discovery, is a strategic choice, and the current research documents and quantifies the tradeoffs associated with this choice. The results documented in this paper imply that managers must understand how to organize scientific discovery while balancing short-term and long-term R&D initiatives (Laverty, 1996), as well as internal and external technology sources. Furthermore, the paper contributes to the recent line of research on corporate science (Mowery, 2009; Pisano, 2010; Simeth and Cincera, 2016; Arora et al., 2018, 2020a, 2021a), by offering an explanation of why conducting corporate basic research remains important and useful for corporate invention.

In terms of data, the paper offers a new and more concrete measure of the organization of scientific research at the firm level. I build on firm-level data, covering 35 years of publications and patents, and their link to each other (DISCERN, 2020) to measure the connectedness of research and invention practices within the firm. I compute a firm-level measure of integration: *the share of integrated-authors (of all authors)*. Integrated authors are scientists who perform both research and invention (i.e., both authors and inventors) and collaborate with specialized inventors (i.e., who only patent). The data enable me to perform cross-industry within-firm analyses and answer relevant questions of how integration relates to firm-level outcomes, which were limited in previous author-inventor co-authorship studies that were mostly at the patent level and for specific industries (Gittelman and Kogut, 2003; Bonaccorsi and Thoma, 2007; Breschi and Catalini, 2010).

The paper proceeds as follows. Section 2 presents the theory and literature review, Section 3 presents the framework that motivates the empirical analysis. Section 4 discusses the data and main measures, Section 5 presents the descriptive statistics and non-parametric evidence, and Section 6 summarizes the econometric results. Section 7 concludes.

4.2 Theory and literature review

4.2.1 The relationship between scientific discovery and invention

Scientific discoveries and inventions are two distinct worlds. The former focuses on the general principles and methods, and the latter on commercial application. For the purpose of this paper, we can think of *Scientific Discovery* as research efforts that yield a scientific publication, and of *Invention* as downstream development of an artifact that results in a patent.

The simplest view of the relationship between scientific discovery and invention was the so-called "linear model" associated with Bush (1945), who asserted that technical progress rests upon scientific advance. Over the years, this view was modified to a more complex relationship that includes a more interactive connection between scientific discovery and invention. More specifically, there is a need to account for both "demand-pull" and "discovery-push" in driving the technological innovation process (Marquis and Allen, 1966; Kline and Rosenberg, 1986; Rosenberg and Nathan, 1994).

The organization of scientific discovery is closely related to the relationship between science and invention. Applying the terminology of Stokes (1997)'s quadrant model: at one extreme specialization, which imposes a clear division between scientific discovery and invention practices, results in "discovery-push" research - a high fundamental understanding of the underlying principals with no direct technological application (i.e., Bohr's pure basic research quadrant). On the other hand, integration of scientific discovery and invention results in "demand-pull" research with direct technological application: either use-inspired basic research (Pasteur's Quadrant) or pure applied research (Edison's Quadrant). Put differently, the organization of scientific discovery has direct implications on the nature of research the firm undertakes and the firm's invention output.

While integration connects research to the immediate technical needs of the firm and facilitates the transfer of science to invention, it is unlikely to result in fundamental research, which is important for significant long-term breakthroughs (Nelson, 1959; Arrow, 1962). As Google-Deepmind's founder puts it,

"A lot of research in industry is product led," Hassabis says. "The problem with that is that you can only get incremental research. [That's] not conducive to doing ambitious, risky research, which, of course, is what you need if you want to make big breakthroughs." (Wired, 2019)[8]

When organizing scientific discovery, managers thus need to understand how to balance between the short-term benefit from integration and its long-term costs; maximizing the impact of corporate science on invention, and at the same time, protecting its long-term properties.

4.2.2 The short-term benefits from integration

The concept of integration between research and invention practices relates to early work by Thomas Allen and Michael Tushman (Allen, 1966; Allen and Cohen, 1966; Allen, 1969; Tushman, 1977; Allen et al., 1979; Tushman and Katz, 1980). Integration nurtures the development of scientists as boundary spanners as well as gatekeepers - occupying a central position in diffusing knowledge between the firm's functions

[8] https://www.wired.co.uk/article/deepmind-protein-folding

and between external scientific sources and the firm. Put differently, integration bridges between scientific discovery and invention, facilitating the transfer of science to invention as well as directing research based on the firm's development needs.

An extensive line of research examines the innovative performance of corporate scientists, who both invent and do research (Henderson and Cockburn, 1994; Stern, 2004; Gittelman and Kogut, 2003; Bonaccorsi and Thoma, 2007; Sauermann and Roach, 2012; Motohashi, 2020).[9] Overall, empirical studies document that firms with integrated author-inventor scientists have higher quality patents. Gittelman and Kogut (2003), for instance, show for a sample of biotechnology firms, that the availability of scientists who both publish and patent, has more impact on the quality of a patent (examined at the individual patent level using forward patent citations to the focal patent) than the firm's stock of scientific publications. Similarly, examining research programs of major pharmaceutical firms, Henderson and Cockburn (1994) find that scientists, who are promoted on the basis of their publications and reputation in the broader scientific community, are related to an increase in invention productivity. Focusing on the nanotechnology industry, Bonaccorsi and Thoma (2007) further show that corporate patents that include at least one inventor who is also an author, are of higher quality than patents with inventors only - including receiving more forward citations and having a wider patent scope.

In related work, Clark et al. (1987) and Holbrook et al. (2000) further suggest that firms that apply cross-functional coordination mechanisms outperform those that do not. For example, Fairchild's early success and its significant breakthroughs in the planar process and integrated circuits are attributed mainly to the cross-functional

[9] past work on author-inventor has mainly focused on the individual scientist and was examined at the patent-level as opposed to the firm-level analysis of a firm's choice of integration and firm-level outcomes that I pursue in the current paper. The firm-level analysis provides a more holistic examination that advances our understanding of the link between the organization of scientific discovery and firm-level invention and scientific outcomes

integration between research and production (Holbrook et al., 2000).[10]

In summary, this literature concludes that for a given level of investment in internal research, a higher level of integration between science and invention practices increases the firm's absorptive capacity capabilities (Cohen and Levinthal, 1990; Rosenberg, 1990). In other words, integration gives the firm a competitive advantage to build on internal and external science, which is openly accessible to everyone, in its inventions (Cockburn and Henderson, 1998; Singh, 2005). Building on the view that science acts as a guide to invention, it follows that firms that are better able to use research should generate more or better-quality patents (Fleming and Sorenson, 2004; Arora et al., 2021a). In addition, integration connects research to the immediate technical needs of the firm, and thus, as asserted by Rosenberg (1990), it increases R&D productivity. Consistent with this, I find that higher integration between scientific discovery and invention practices increases downstream invention quantity and quality.

If integration between research and invention is, in fact, fruitful for invention – why don't all firms choose to organize their scientific discovery in such a way? What the aforementioned papers do not examine is the relationship between integration and long-term scientific discovery, invention, and value outcomes. In this paper, I address this gap in the literature by examining both the benefits as well as the costs of integration.

4.2.3 The long-term costs of integration

Previous research sheds some light on the direct costs of integration in terms of scientific research output. From an opportunity cost perspective, integration draws

[10] Moreover, Fairchild's decline in the mid-1960s is mainly contributed to the move towards more specialized research: "When the firm grew large and geographically dispersed and when R&D had become highly centralized, and an end in itself, this type of coordination broke down, contributing to Fairchild's decline" (Holbrook et al. (2000): p. 1030)

scientists away from research, which results in lower scientific output. In terms of coordination costs, integration involves researchers teaming up with inventors, two very different groups in their nature (Allen, 1984; Vincenti, 1990). Researchers and inventors use different technical language and coding schemes (Allen, 1984). Furthermore, they have different incentives and rewards system. Researchers are more concerned with non-pecuniary motives such as independence, autonomy in their research agenda, the publication of their work, and their reputation in the broader scientific community (Stern, 2004; Sauermann and Cohen, 2010). Inventors' goals, on the other hand, are more tied to the implementation of their work and their achievements within the firm (Ritti, 1971; Allen, 1984). Inventors' clearly defined research objectives based on development needs might contradict the researcher's basic motives. Integration can, therefore, also decrease the recruitment of researchers, who prefer independence and have a taste for basic research, which will eventually lead to further deterioration in basic science (Stern, 2004).[11]

Integration also has an indirect cost in terms of long term invention output. Building on the fundamental view that sees scientific research as an input into invention (Bush, 1945; Rosenberg, 1990; Narin et al., 1997; Arora et al., 2021a), the decline in scientific research due to integration will decrease invention in the long run (Gambardella, 1992; Li et al., 2017; Poege et al., 2019).[12]

In the current paper, I find that integrating scientific discovery with invention practices lowers the quantity and quality of publications as well as their private value. I also show that Scientific discovery complements the invention process. Taken

[11] While scientists' preferences are not directly considered in this paper, scientists select into organizations that fit their preference in terms of taste for either science or commercialization (Stern, 2004; Sauermann and Roach, 2012). Insofar as more productive scientists are more likely to work in firms with clearer separation between discovery and invention, integration will reduce the scientific capability of the firm.

[12] for example, Gambardella (1992), shows a positive relationship between the quantity of basic research a pharmaceutical firm performs and the patents it produces

together, I document the cost of integration in terms of long term invention output.

The current paper is closely related to a stream of literature that explores R&D organizational structure and innovation outcomes (Kay, 1988; Argyres and Silverman, 2004; Arora et al., 2014; Argyres et al., 2019), which suggests that R&D structure - centralized or decentralized[13] - conditions the nature of research the firm undertakes. Firms with decentralized R&D, managed at the business unit level with close interactions between research and production (similar to the concept of integration in the current paper) tend to produce more short-term applied research specific to their products and services. In contrast, firms with a centralized research lab are likely to conduct more general basic scientific research that benefits the firm as a whole. Another related stream of research is the innovation networks literature (Breschi and Catalini, 2010; Argyres et al., 2019). Examining networks of author-inventors, Breschi and Catalini (2010) investigate whether one function comes at the expense of the other for the scientist. They find that maintaining a central position in the technological network may come at the expense of filling a similarly central position in a scientific network. While Breschi and Catalini (2010) do not consider the institutional affiliations of authors-inventors, they suggest that some tradeoff may exist between the two functions. In more recent work, Argyres et al. (2019) bridge between the R&D organization literature and the network literature to examine the relationship between formal R&D structure, internal inventor networks, and innovative outcomes. They find that co-invention networks mediate the relationship between structure and innovative outcomes. That is, they show evidence that suggests, as I argue in the current paper, that co-invention practices within a firm are shaped by choice of R&D structure. Furthermore, the concept of integration is related to a line of research that examines the diversity of inventors' knowledge on invention teams and firms' recom-

[13] measured by resource allocation decision (Argyres and Silverman, 2004; Argyres et al., 2019), and patents assignment between the headquarter and affiliates (Arora et al., 2014)

binant capabilities (Henderson and Clark, 1990; Fleming, 2001; Katila and Ahuja, 2002; Singh and Fleming, 2010; Carnabuci and Operti, 2013; Aggarwal et al., 2020; Nagle and Teodoridis, 2020) and their relationship to invention output. For example, Singh and Fleming (2010) show that diversity within-teams can trim poor outcomes, which is another mechanism through which integration can contribute to invention quality. Aggarwal et al. (2020) further distinguish between within-team versus across-team knowledge diversity using a firm-level analysis. When considering both within-team coordination costs and cross-team knowledge flows, they find that concentrated knowledge structures, with low within-team diversity and high across-team diversity, are associated with higher firm-level innovation quality. This latter finding emphasizes the importance of examining organizational measures and their outcomes at the firm-level, as I pursue in the current paper. Lastly, this paper is closely related to the literature on the interplay between exploration and exploitation and the concept of ambidexterity (March, 1991; Tushman and O'Reilly III, 1996; Lavie et al., 2010; O'Reilly III and Tushman, 2013; Stettner and Lavie, 2014).

4.2.4 The determinants of integration

The coupling between discovery and invention practices can vary within-industries and within-firms over time. To better understand firms' integration choice, I present three main factors that condition the benefits and costs associated with integration: market for technology, reliance on science, and stage of technology.

A reduction in scientific discovery due to integration that will result in lower invention quality, suggests that firms would need to rely on external sourcing for technology in the long run. Firms can obtain external technology through (i) markets for technology (MFT) (Arora et al., 2001; Serrano, 2010), (ii) markets for firms (MFF), and (iii) Mobility of people (Singh and Agrawal, 2011; Tzabbar et al., 2015). As the market for external technology rises, the firm can increase integration, even at the cost of lowering internal scientific discovery, knowing that external technology sourcing options are available. Alternatively, as outside opportunities decrease, firms will need to rely on internal development for breakthroughs in the long run - integration should thus be minimized to maximize the payoff from internal scientific discovery.[14] This idea also fits with previous research findings that decentralized firms that have a stronger connection between science and invention at the unit level, tend to rely more on external acquisitions, while centralized firms draw more value from internal R&D (Arora et al., 2014). Consistent with this reasoning, we should expect integration to increase and be less harmful for scientific discovery value with the availability of external technology sourcing.

In terms of the nature of scientific research in the field, Cohen et al. (2020), find that in more applied-engineering fields, where research and invention practices are more connected, the opportunity cost for academic scientists being author-inventors is quite low relative to more fundamental fields (e.g., physical sciences). In the current paper, I extend their argument to the level of the firm. I argue that in fundamental science-based fields, where research results are further away from the commercial end, although integration is meaningful (i.e., connecting two distinct practices), the opportunity costs for research are high - integration requires scientists to depart further away from the traditional research in their field.[15] Consistent with this reasoning, we should expect integration to be less valuable for scientific discovery value, yet more beneficial for invention value, in more fundamental science-based

[14] one can argue that as the market for technology increases, firms would increase their specialization in scientific discovery in order to become a seller of technology in the market. As my sample consists of large firms, it seems less likely that they are innovating to sell their inventions (Figueroa and Serrano, 2019). In fact, I find that only 13% of the firms in my sample sell more than 30% of their patents granted throughout the complete sample period.

[15] In fact, in applied fields, where the practice of science is inseparable from the development (such as in the clinical trial phase in the pharmaceutical and biotech industry), we can expect corporate scientists to be both researchers and inventors in the same mind, such that there is no real need for integration.

fields.

Similarly, the benefits from integration are more significant in the early stages of a technology field (Lieberman, 1978; Faulkner, 1994), when invention is more knowledge led, and is less so in later stages, where inventions are incremental and are more likely to result from trial and error. Therefore, for firms with frequent introductions of new technologies, we should expect higher benefits from integration in terms of invention value.

To better understand the tradeoffs and the determinants related to integration, I briefly outline a simple framework, which I then build on in the empirical analysis.

4.3 Conceptual framework

I present a framework in which a firm's payoff depends on internal inventions, d_1 and on external technology sourcing. Scientific discovery, r, benefits internal invention by directing the firm's search focus to a defined set of possible avenues to explore (Nelson, 1982; Fleming and Sorenson, 2004). This reduces the cost of invention and increases its quality. Firms also choose their long-term scientific discovery organization strategy – either specialized or integrated with invention. Both scientific discovery, r, and integration, t, are inputs to invention, such that, for a given level of investment in scientific discovery, a higher level of integration increases the firm's invention quality. That is, $d_1 = d_1(r, t, Z)$, where Z is various technology and industry characteristics, where $\dfrac{\partial d_1}{\partial r} > 0$ and $\dfrac{\partial d_1}{\partial t} > 0$. Investment in science has a marginal cost of c. Both investment in science, s, and integration, t, determine the firm's scientific discovery output. While integration increases internal invention, it reduces scientific discovery. That is, $r = r(s, t, Z)$, where $\dfrac{\partial r}{\partial s} > 0$ and $\dfrac{\partial r}{\partial t} < 0$. Similarly, for external technology, I assume that the net value of the firm's external sourcing is $P(t)d_2$, where $\dfrac{\partial P}{\partial t} > 0$. That is, external sourcing of inventions increases with the thickness of the market for technology d_2 and with integration, such that higher integration makes it easier for the firm to discern relevant external inventions and observe them inside the firm post-acquisition. The firm's decision problem is to choose its research investment, s, and integration, t, to maximize its profit: $V = \Pi(d_1) + P(t)d_2 - cs$, where $\dfrac{\partial \Pi}{\partial d_1} > 0$. Since integration is negatively related to scientific discovery and positively related to invention, this should be reflected not only in the level of publications and patents, but also in the firm's value. That is, I examine the cross partial of V with respect to integration and scientific discovery (or invention). I find that $\frac{\partial^2 V}{\partial t \partial r} < 0$, and $\frac{\partial^2 V}{\partial t \partial d_1} > 0$. The optimal coupling between scientific discovery and invention practices varies by main determinants of integration. In this model, I examine the availability of external technology sourcing, d_2, as well as the technology and industry characteristics, Z, such as reliance of invention on science and stage of technology in the field as main determinants. In terms of the market for technology, when external sourcing opportunities are low, firms need to rely on internal development for breakthroughs in the long run. Therefore, we should expect lower benefits from integration in terms of scientific discovery value when the supply of external technology is low. In terms of the nature of technology, in science-based fundamental fields, research and invention are two distinct practices. While integration should increase invention value, by connecting two distinct practices, the opportunity cost for research is higher - it requires scientists to depart further away from the traditional research in their field. We should, therefore, expect integration to be negatively related to scientific discovery value in fundamental science-based fields. Similarly, in the early stages of a technology field, where an invention is more knowledge led, we should expect higher benefits from integration in terms of invention value. Yet, the opportunity cost of integration in terms of scientific discovery value should still be high. Table 4.1 summarizes how integration conditions main outcomes: scientific discovery, invention, and patent and publication value, and how those relationships vary by main determinants. Table 4.2 presents the predicted relationship between integration and the main determinants. Section 4.6, explores these relationships empirically. Table 4.7 explores the relationship between integration and these main determinants. Table 4.8 examines whether research quantity and quality decrease with integration. Similarly, Table 4.9 explores the positive relationship between invention and both integration and scientific discovery. Taken together, they show that the decline in scientific research due to integration may decrease invention quality in the long run. In Table 4.8, I also pursue an Instrumental Variable estimation strategy for the relationship between integration and research output that is motivated by external sourcing as a potential determinant of integration. Table 4.10 presents the relationship between integration and patent and publication value. Table 11 further shows how the above determinants of integration condition the relationship between integration and patent and publication value, $\frac{\partial^3 V}{\partial Z \partial t \partial r}$, $\frac{\partial^3 V}{\partial Z \partial t \partial d_1}$, $\frac{\partial^3 V}{\partial d_2 \partial t \partial r}$, and $\frac{\partial^3 V}{\partial d_2 \partial t \partial d_1}$.

Table 4.1: Framework:Outcomes

VARIABLE	SCIENTIFIC DISCOVERY	INVENTION	PUBLICATION VALUE	PATENT VALUE
Integration	Decrease	Increase	Decrease	Increase
Integration ×				
Reliance on science			Decrease	Increase
Early stage of technology			Decrease	Increase
Thick MFT			Increase	-

Notes: The full model contains proofs and extensions.

Table 4.2: Framework: Determinants

VARIABLE	RELIANCE ON SCIENCE	EARLY STAGE TECHNOLOGY	THICK MFT
Integration	Increase	Increase	Increase

Notes: The full model contains proofs and extensions.

4.4 Data and main variables

I combine data from 4 main sources: (i) firm-level data on publications and patents from DISCERN dataset[16] (Arora et al., 2021a), (ii) company and accounting information from S&P North American Compustat (Standard & Poor's, 2018b), (iii) scientific publications and author information from Web of Science (WoS) (Clarivate Analytics, 2016), and (iv) patent and inventor information from PatStat (European Patent Office, 2016).

4.4.1 Accounting panel data

DISCERN dataset covers U.S.-headquartered publicly listed firms and their subsidiaries over the period 1980-2015. For the purpose of this paper, which focuses on the organization of R&D, the sample is restricted to publishing manufacturing firms[17] with at least ten publications during the sample period.[18] The final sample includes an unbalanced panel of 1,506 ultimate owner parent companies and 24,510 firm-year observations over the sample period 1980-2015.

[16] The data can be freely downloaded from 10.5281/zenodo.3594642. The version used for the analysis in this paper is version 5.

[17] As in Bloom et al. (2013), manufacturing firms are based on SIC codes in Compustat segment file.

[18] out of 2,653 firms in the sample with at least one granted publication throughout the sample period, 1506 firms have at least 10 publications. The restriction on publications is to assure that there is a sufficient amount of publications, such that the measure of integration is meaningful, and the firm is, in fact, likely to face a choice regarding its scientific discovery organization. In a robustness check, I further show that results hold both for small and large R&D firms in my sample.

4.4.2 Main variables

Integration

Integration in this research is defined as the connectedness between scientific discovery and invention practices within the firm. While scientific discovery organization structure is difficult to observe, I compile a measure that is correlated with it: *share integrated authors of all authors.* Integrated authors are scientists who perform both research and invention (i.e., both authors and inventors) and collaborate with specialized inventors (i.e., who only patent). That is, integration involves collaborations between individuals who work on research and those who specialize in invention.[19] These collaborations could be in the form of cross-functional interactions between separate research and development units[20], as well as within-unit interactions. Furthermore, these collaborations can be imposed by the firm's research organization strategy (e.g., Microsoft's MSR NEXT group), by the firm's organization and reporting structure, by R&D budget allocation decisions and workers' incentive plan (e.g., IBM's incentive plan change in 1989 that emphasizes patenting over scientific publication), as a result of physical co-location of authors and inventors, as well as a consequence of the technology focus of the firm on more applied initiatives.

I pursue several robustness checks to get a better understanding of what my measure of integration captures. First, I find that my measure of integration is different from self-use of science in invention (Arora et al., 2021a). Specifically, I find that only 20% of the patents by scientists classified as "integrated" self-cite inventors' science, which supports the idea that integration is an outcome of direct interaction between science and invention practices rather than merely an outcome of researchers

[19] conditioning on collaborations with specialized inventors is essential for capturing real interactions between science and invention practices and not simply a lone author or a group of authors who patent their discovery.

[20] Integration, in this sense, is closely related to Lawrence and Lorsch (1967)'s concept of "unity of effort among the various subsystems" and to Clark et al. (1987)'s cross-functional coordination.

choosing to work on applied problems that are subsequently cited by their patents. In Supplementary Table 4.13 Columns 5-8, I further present robustness checks for integration using a measure of co-location of inventors and authors.[21] For both the publication and patent equation, I receive consistent and significant results with the co-location dummy at the industry level, but insignificant results at the firm level. This reassures that my measure of integration does not simply capture interactions of research and inventions due to co-location of research and invention activities.[22] Lastly, in section 4.5.3, I validate my measure of integration using the 1994 Carnegie Mellon Survey (CMS) data (Cohen et al., 2000). I find that my measure is correlated with cross-functional communication within firms.

To compute the integration measure, I start by identifying all authors listed on the scientific publications and inventors listed on the patent documents related to the corporate firm sample. My goal is to match authors and inventors related to the same ultimate owner (UO) firm during a 5-year cohort period in order to identify individuals who both patent and publish.[23] Any individual who had at least one publication during a 5-year cohort period is considered an author; any individual who had at least one patent during the same period is considered an inventor.

One challenge I face using WoS data is that not all authors are linked to an institutional affiliation address.[24] This can cause a problem for collaborative publi-

[21] I look at the top city of authors and the top city of inventors for each 5-yr cohort and compute a dummy variable that equals one if they are located in the same city.

[22] Supplementary Table 4.13 presents additional robustness checks for the integration measure. In Columns 1 and 2 integration is computed excluding scientific publications from new journals post-1990. Results for both patent and publication equations continue to hold, which reassures that my results are not affected from new journals (and perhaps more applied journals) that were added in the latter part of the sample. Similarly, in Columns 3 and 4 integration is computed excluding applied scientific publications with below-median JIF, and results continue to hold.

[23] The measure is per 5-year period as the main assumption is that integration is a long-term organizational feature.

[24] for example, in some cases, WoS only documents the link between the reprint author and her institution address, while in the original publication, all authors are linked to an address

cations that consist of authors from several institutions. To overcome this challenge, I compile three lists of author names- (i) the complete list of authors listed on each corporate publication, (ii) a list of all authors that were linked in WoS to their UO firm, and (iii) a list of all authors related to non-collaborate scientific publications (i.e., where all institutions listed on the publication are related to a unique UO firm). In addition, for the purpose of compiling the inventor name list, I exclude from the sample patents with multi assignees, which are far less prominent than collaborative scientific publications (less than 2% of patents in the sample). I then use this inventor list to match the above author list (i). In other words, limiting the sample of inventors provides me the certainty that all matched author-inventor individuals are related to the focal firm. Finally, after matching, I use author lists (ii) and (iii) combined with the matched author-inventor results to compute the total number of unique authors for each firm-cohort.

One other challenge is that patent data report inventors' first names and last names, whereas WoS data list last names and initials of authors. To resolve the differences, I first standardized inventor and author names in a similar way – last names and initials. Since the match is done within firm-cohort, I am less exposed to mismatches that identify different individuals as the same person, which could lead to an overestimation of integrated scientists. Nonetheless, I conduct extensive manual checks to verify the matches, especially of common and short names.[25]

I divide the sample period into seven cohorts of 5 consecutive calendar years. I fuzzy match the standardized list of inventor names with the list of standardized author names to identify for each UO firm-cohort individuals responsible for both a patented invention and a scientific publication. Next, I merge the matched results back to the patent level data to identify integrated scientists who perform both

[25] I also run the match with different levels of restrictions for false-positive matches. I confirm that results hold even when I use a very conservative definition for matches

research and invention (i.e., both authors and inventors) and are part of an inventor team, which includes at least one specialized inventor (i.e., only inventor). By doing so, I essentially exclude from my measure of integration cases of paper-patent pairs as well as cases of lone inventors, where there is no real interaction between author and inventor teams.

Lastly, I aggregate the data to the firm-cohort level by counting the number of unique integrated scientists in each UO firm-cohort and dividing by the total number of unique authors in each UO firm-cohort, to compute my main measure for the analyses- *share integrated authors of all authors*.

Supplementary Table 4.12 illustrates the effectiveness of my measure in capturing changes in the organization of scientific discovery using case studies. Column 1 explores the major change in International Business Machines's (IBM) R&D organization strategy practices in the late 1980s (Gomory, 1989; Bhaskarabhatla and Hegde, 2014). Up to 1989, IBM's research under the academically oriented leadership of Ralph Gomory (1970-1986) and his successor John Armstrong (1986-1989), was mainly focused on basic research separated from development. In 1989, following a change in U.S. patent law in the early 1980s, and with the appointment of James McGroddy as director of IBM Research, the company adopted pro-patent IP management practices. Simultaneously, it also shifted its focus towards more applied research initiatives, including joint research and development programs. Column 1 presents the trend in the integration measure for IBM. The sharp increase in integration from an average share of 0.24 pre-1990s to 0.55 post-1990s is consistent with the company's documented shift in scientific discovery organization. Furthermore, (Bhaskarabhatla and Hegde, 2014) find that in the decade following this shift, IBM increased their patent applications and decreased its publications, which is consistent with the results I document in this paper for integrated organization structure. Table 4.12 Column 2 further explores the effect of Bell-Labs' separation from AT&T

CORP in 1996 on integration. It shows that following Bell's acquisition by Lucent, AT&T became more integrated - reflecting the firm's loss of Bell's specialization in basic research.

To the best of my knowledge, this is the first paper to compile data on inventor-author integration at the firm level for a wide variety of industries, across 3.5 decades. Previous work mostly examined co-authorship data at the patent level (e.g., Bonaccorsi and Thoma (2007)), within networks of inventors (e.g., Breschi and Catalini (2010)), and for specific industries and limited years (e.g., Gittelman and Kogut (2003)).

Market for technology (MFT)

I measure market for technology based on patent trading activity in invention classes relevant to the focal firm's patent portfolio (Serrano, 2010; Hochberg et al., 2018; Figueroa and Serrano, 2019; Arora et al., 2020b). Patent transactions are from the USPTO Patent Assignment dataset. [26] For each sample firm's patent portfolio, I compute the probability (averaged across related IPC classes) that a patent related to the focal firm's portfolio will be sold.[27]

Reliance on science in invention

I measure reliance on science by citations to external WoS scientific publications located on the front page of a patent.[28] Patent citations are from PatStat and their matched publications from WoS.

[26] The clean patent transaction data is based on Arora et al. (2020b). Following Figueroa and Serrano (2019), the transactions exclude patents that are reassigned due to pure M&As as well as deals with more than 25 patents transferred. That is, they may include acquisitions of small startups.

[27] Specifically, for each firm i, year t related IPC codes in its patents granted between [t,t-5], I compute the share of external patents sold up to year t out of all patents granted between [t,t-8]. The share is then averaged across IPC classes.

[28] I rely only on citations to external science to make sure that the measure is not directly related to the main dependent variable, annual publications.

Stage of technologies

Early-stage technologies are defined as patents granted no more than 10 years from the related IPC inception year. IPC inception date is from PatStat.

4.5 Descriptive statistics and non-parametric evidence

The main sample and variables are at the parent company-year level. Table 4.3 presents descriptive statistics for the main variables over the sample period, 1980-2015. The sample includes a wide distribution of firms in terms of size and R&D investment: market value ranging from 24 million dollars (10th percentile) to 10 billion dollars (90th percentile) and R&D expenditures ranging from 2.74 million dollars (10th percentile) to 319 million dollars (90th percentile). The sample also varies in terms of R&D employees: authors range from 1 author (10th percentile) to 244 authors (90th percentile) and inventors from 2 inventors (10th percentile) to 442 inventors (90th percentile). The firms produce, on average, 31 publications and 49 patents per year.

Integration varies across the sample ranging from zero (10th percentile) to 50% (90th percentile), with a mean of 22%, and it tends to be higher in science-based industries (Electronics and Semiconductors- 26%, Pharma 24%, whereas IT and Software 18%). Supplementary Figures 4.1 and 4.2 present trends in integration for main industry groups: (i) IT & Software, (ii) Electronics & Semiconductors, (iii) Telecommunication, (iv) Chemicals, (v) Energy, and (iv) Pharma & Biotech. There is substantial heterogeneity in integration over time by industry. Figure 4.1 shows an increase in integration in the first three groups (IT & Software, Electronics & Semiconductors, and Telecommunication). For life-science related industries (Figure 4.2), integration trend is less clear, with very similar rate at the end of the sample period as at the beginning of the sample period.

4.5.1 Integration and the tradeoff between scientific discovery and invention

Table 4.4 presents mean comparison tests for differences in scientific discovery and invention between firms with high and low integration. It shows that firms with above-median integration share also have more and better-quality inventions (an average of 0.48 citation-weighted patents to R&D ($mm) per firm-year for low integration share versus an average of 0.76 citation-weighted patents to R&D ($mm) for high integration share). Conversely, firms with above-median integration share have a significantly lower rate and quality of scientific discovery (an average of 0.66 citation-weighted publications to R&D ($mm) per firm-year for low integration share versus an average of 0.55 citation-weighted publications to R&D ($mm) for high integration share). The difference is even higher for more science-based technologies. These results are consistent with the idea that while integration is beneficial for internal invention, it has adverse effects on scientific discovery.

4.5.2 Determinants of integration

Table 4.5 presents mean comparison tests for integration by above and below median values of determinants of integration. It shows that firm-cohorts with above-median reliance on science in invention (measured by average citations to external science per patent) have a statistically significant higher share of integration (an average of 0.15 for low science-based versus an average of 0.28 for high science-based technologies). Similarly, integration is more prominent in firm-cohorts with early-stage technologies - measured by patents granted no more than 10 years from the related IPC inception year. When examining external sourcing, I find that one primary determinant of integration is external markets for technology. Integration is statistically significant higher in firm-cohorts with above-median MFT (an average of 0.19 for low MFT versus an average of 0.24 for high MFT).

Table 4.3: Summary Statistics for Main Variables

					Distribution		
VARIABLE	# Obs.	# Firms	Mean	Std. Dev.	10th	50th	90th
Integration	24,510	1,506	0.22	0.22	0	0.18	0.5
Authors	24,510	1,506	152	609	1	17	244
Inventors	24,510	1,506	224	828	2	32	442
Market for technology (MFT)	24,510	1,506	0.06	0.02	0	0.06	0.08
Scientific publications count	24,510	1,506	31	130	0	4	47
Scientific publications stock	24,510	1,506	156	694	2	15	208
Patents count	24,510	1,506	49	200	0	6	94
Patents stock	24,510	1,506	227	933	2	25	444
R&D expenditures ($mm)	24,510	1,506	188	708	2.74	26	319
R&D stock ($mm)	24,510	1,506	814	3,414	6.6	93	1,276
Market value ($mm)	24,510	1,506	6,515	30,218	24	474	10,466
Tobin's Q	24,510	1,506	5	6	0.46	2	20
Sales ($mm)	24,510	1,506	4,450	16,555	5	427	8,921
Assets ($mm)	24,510	1,506	3,343	14,522	4	236	5,990

Notes: This table provides summary statistics for the main variables used in the econometric analysis. The sample is restricted to publishing firms with at least 10 publications during the sample period. The sample is at the firm-year level and includes an unbalanced panel of 1,506 U.S. headquarter publicly traded manufacturing ultimate owner parent companies over the sample period 1980-2015.

Table 4.4: Integration, invention, and scientific discovery

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Diff. in means	High Integration			Low Integration		
	(3) minus (6)	No. Obs.	Mean	Std. Dev.	No. Obs.	Mean	Std. Dev.
Patent flow/R&D exp.	0.124**	747	0.568	0.687	748	0.444	0.785
Citation-weighted patent flow/R&D exp.	0.272**	747	0.755	0.906	748	0.483	0.746
Publications flow/R&D exp.	-0.212**	747	0.481	0.742	748	0.693	1.277
Citation-weighted publication flow/R&D exp.	-0.106*	747	0.553	0.835	748	0.659	1.215
Firms with high reliance on science:							
Citation-weighted patent flow/R&D exp.	0.350**	379	0.799	0.927	379	0.449	0.705
Citation-weighted publication flow/R&D exp.	-0.196**	379	0.473	0.647	379	0.670	1.048

Notes: This table presents mean comparison tests for firms with high integration vs. firms with low integration. The unit of analysis is a firm, yearly values are averaged over the sample period. High and low share of integration are defined by above and below the median sample value of average per firm share of integration, respectively. Patents are weighted by IPC-year normalized forward patent citations. Publications are weighted by journal-year normalized forward publication citations. R&D expenditures are in $mm. * p<0.01 ** p<0.05

Table 4.5: Mean Comparison: Determinants of Integration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Diff. in means	High			Low		
	(3) minus (6)	No. Obs.	Mean	Std. Dev.	No. Obs.	Mean	Std. Dev.
Reliance on science in invention	0.128**	2,815	0.277	0.201	2,815	0.148	0.196
Early stage technology	0.070**	695	0.274	0.161	4,935	0.204	0.213
Market for technology	0.048**	2,815	0.236	0.218	2,815	0.188	0.196

Notes: This table presents mean comparison tests for integration by high and low values of different determinants of integration. The unit of analysis is a firm-cohort, yearly values are averaged over the cohort period. High and low are defined by above and below the median cohort value, respectively. See main text for variable definitions. * $p<0.01$ ** $p<0.05$

4.5.3 Validating share of integrated authors as a measure of organization of scientific discovery

To validate my measure of integration, I use the 1994 Carnegie Mellon Survey (CMS) data on industrial R&D (Cohen et al., 2000). As part of the survey, lab directors in R&D performing firms were asked about the relationship of their lab with other business functions. Of the firms in my sample, 214 are also covered in the CMS. I match the integration measure for years 1991-1995 to CMS questions response related to the importance of inter-firm cross-function communication.

Table 4.6 confirms that my measure of integration is related to cross-functional interactions within firms. Specifically, Column 1 shows that integration is related to above-median communication between R&D units.[29] More importantly, Column 2 suggests that integration is related to project teams with cross-functional participation.[30]

[29] based on CMS data Q.5c: How frequently do your R&D personnel talk face to face with personal from other R&D units?

[30] based on CMS data Q.6b: During the last three years, have Project teams with cross-functional participation been used to facilitate interaction?

Table 4.6: Supporting Evidence from Carnegie Mellon Survey

	(1)	(2)
Dependent variable:	Dummy for above median communication between R&D units	Dummy for cross functional teams (CMS Q6B)
Integration	0.653**	0.202*
	(0.208)	(0.098)
ln(*Authors*)	0.025	-0.016
	(0.023)	(0.010)
ln(*Sales*)	0.010	0.035*
	(0.021)	(0.014)
Main R&D personnel degree field dummies	Yes	Yes
Observations	120	143
R-squared	0.19	0.18

Notes: This table presents OLS estimation results for the relationship between integration and the 1994 Carnegie Mellon survey (CMS) questions response (Cohen et al., 2000) related to the importance of inter-firm cross-function communication. Column 1 is based on CMS data Q.5c: "How frequently do your R&D personnel talk face to face with personal from other R&D units?". Column 2 is based on CMS data Q.6b: "During the last three years, have project teams with cross-functional participation been used to facilitate interaction?". Sample is restricted to survey firms that were matched to our sample. Robust standard errors in parentheses. Number of observations vary based on the response rate to each question.

4.6 Econometric analysis

4.6.1 Integration and determinants

Table 4.7 provides econometric evidence supporting the relationship between integration and main determinants, as presented in Table 4.5. All specifications include firm and year fixed effects, and two-year lagged R&D stock.

Column 1 shows a positive and statistically significant relationship between reliance on science in invention (measured by annual citations to external scientific publications) and integration. Column 2 examines the relationship between early-stage technology and integration. It shows that one standard deviation increase in share of new technology will increase integration by 1.6%. Column 3 further examines the relationship between MFT and integration. It shows that one standard deviation increase in MFT will increase integration by 5%. Finally, Column 4 combines the different determinants and shows that results continue to hold. Table 4.7 suggests that the optimal coupling between scientific discovery and invention is positively related to the nature of research and technology in the field as well as to the availability of external technology sourcing.

Table 4.7: Integration and Determinants

	(1)	(2)	(3)	(4)
Dependent variable:		Integration		
$ln(Citations\ to\ external\ science)_{t-2}$	0.022**			0.022**
	(0.001)			(0.001)
$Share\ new\ technology_{t-2}$		0.116**		0.087*
		(0.041)		(0.040)
MFT_{t-2}			0.446**	0.260**
			(0.086)	(0.084)
$ln(R\&D\ stock)_{t-2}$	0.007**	0.018**	0.017**	0.007**
	(0.002)	(0.002)	(0.002)	(0.002)
Firm fixed-effects	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
Number of firms	1,445	1,445	1,445	1,445
Observations	21,441	21,441	21,441	21,441
R-squared	0.56	0.55	0.55	0.56

Notes: This table presents OLS estimation results examining the relationship between integration and main determinants. Integration is defined as the share of a firm's authors who both published an article and were granted a collaborative patent with a specialized inventor during a 5-year-cohort period. New technology is defined as patents granted no more than 10 years from the related IPC class inception year. Markets for technology (MFT) is based on patent trading activity in invention classes relevant for the focal firm's patent portfolio. One is added to logged control variables. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

4.6.2 Integration and scientific discovery

The relationship between scientific discovery and integration is specified as follows:

$$\begin{aligned} \ln(Publications)_{it} = & \beta_0 + \beta_1 Integration_{it-2} + \beta_2 \ln(R\&D\ stock)_{it-2} \\ & + \boldsymbol{Z}'_{it-2}\boldsymbol{\gamma} + \boldsymbol{\eta_i} + \boldsymbol{\tau_t} + \epsilon_{it} \end{aligned} \tag{4.1}$$

In Equation 4.1, $Publications_{it}$ is the annual flow of publications by firm i in year t weighted by the number of forward citations each publication receives divided by the average number of citations received by all other publications published in the same journal-year. Integration is proxied by $Integration_{it-2}$, measured by the share of a firm's authors who both publish and patent out of all authors. $\boldsymbol{Z}_{it-2}$ is a vector of two-year lagged firm-year controls. The coefficient of interest is β_1. Following the prediction in Table 4.1, I expect $\hat{\beta}_1 < 0$.

The organization of scientific discovery may vary across firms and industries. I thus include firm fixed effects as well as time-varying firm characteristics for scale, such as R&D stock. Furthermore, the choice of the temporal structure aims at mitigating concerns that the relationship between the number of yearly publications and integration is merely due to common shocks.

Table 4.8 presents the estimation results. Column 1 presents results from a pooled specification with four-digit SIC dummies and 2-digit main IPC class dummies. There is a negative and statistically significant relationship between integration and number of yearly publications. In Column 2, which adds firm fixed effects to the specification in Column 1, $\hat{\beta}_1$ slightly decreases, indicating that the relationship between integration and publications partly reflects a degree of heterogeneity across firms. Yet, $\hat{\beta}_1$ remains positive, both substantively and statistically: a one standard deviation increase in integration is associated with a 3.9% decrease in yearly publications - approximately 1.3 publications per year. As a robustness check, Column 3 shows that results also hold when controlling for lagged stock of patents.

Integration does not only condition the quantity of publications but, more importantly, the quality of publications. Columns 4-6 confirm that results hold for a variety of scientific quality measures. In Column 4, publications are weighted by journal-year normalized forward publication citations. In Column 5, the sample of patents is restricted to the top 2 percentile of corporate publication by journal-year weighted forward citations. In Column 6, the sample is restricted to publications with above-median Journal Impact Factor.

The level of integration the firm chooses varies by the nature of technology in the field. Columns 7 and 8 divide the firm sample based on below and above-median reliance on science in invention, respectively. Results indicate that the observed relationship is driven by science-based technologies. The coefficient estimate for integration for high reliance on science is negative and statistically significant, while the estimate for low reliance on science is statistically zero. For the sample of science-based firms, a one standard deviation increase in integration is associated with a 12% decrease in yearly publications - approximately 10 publications per year.

It is possible that both integration and the production of research are potentially affected by common unobserved variables, which would bias the OLS estimation. For example, a change in scientific opportunities that affects both the rate of investment in scientific discovery and the collaboration opportunities between researchers and inventors. Another concern is that changes in firm strategy will affect both integration and investment in science - for example, a firm might be more inclined towards integration if it intends to focus on more applied invention in the future. This could coincide with the firm's decision to invest less in science, which would mean that the observed correlation between integration and research output is not causal. To mitigate these concerns, I present results when integration is instrumented using a potential external determinant of integration –the supply of external technology sources. More precisely, I use as an instrument changes in state policy that regulate workers' ability to move across employers.

The reduction in scientific discovery quality associated with integration, which in return lowers long-term invention quality, suggests that firms would need to rely on external sourcing in the long-term. Firms can obtain external technology through MFT, MFF, and by hiring scientists. Moreover, there is evidence that internalizing and diffusing a new hire's knowledge requires a high degree of within-firm integration (Tzabbar et al., 2015; Singh and Agrawal, 2011).

I follow Klasa et al. (2018) and exploit variation over time in the adoption of Inevitable Disclosure Doctrine (IDD) by U.S. state courts as a source of exogenous variation in integration. IDD has been shown to impose legal restrictions on labor mobility (Marx et al., 2009; Klasa et al., 2018). IDD restricts worker mobility from one organization to another in cases where they might "inevitably disclose" trade secrets. It is applicable even if the employee did not sign a non-compete or non-disclosure agreement, if there is no evidence of actual disclosure, or if the rival is located in another state.[31] The dummy variable, IDD_{st}, is equal to one when the Inevitable Disclosure Doctrine is in effect in the focal firm's state (s), in a given year (t), and zero otherwise.[32] External hiring can either substitute or complement external sourcing through MFT (Arora and Gambardella, 1990; Arora, 1996; Cassiman and Veugelers, 2006, 2007; Bei, 2018). To capture the firm-specific external technology sourcing options, IDD is multiplied by the firm-specific external market for technology measure, MFT_{it} to form $IDD \times MFT_{sit}$.[33]

The key identifying assumption is that policy-imposed barriers to researcher's mo-

[31] As a robustness check, in unreported results, I compute a measure of authors' tenure for each firm-cohort and find a statistically significant positive relationship between IDD and authors' tenure, suggesting that IDD is related to restricted mobility

[32] IDD effective years for relevant states are based on Klasa et al. (2018). The relevant state for each firm-year is determined using the majority publishing-state in each 5-year cohort, based on the publication's affiliation field.

[33] 43% of the firm-year observations in the sample have effective IDD. The average value of $IDD \times MFT$ measure is 0.056 with a standard deviation of 0.024

bility do not directly affect the incentives to invest in research, but only through their affect on integration. Further, one has to assume that the instrument $IDD \times MFT_{it-2}$ shifts the cost of integration, but is uncorrelated with unobserved state-specific variables (e.g., technological or scientific opportunities) that may also affect incentives to invest in research. To mitigate concerns, I further restrict my instrumental variable estimation only to the publication equation as there is a higher chance that IDD may affect incentives to patent, or that technology opportunities may directly affect inventive output.

Columns 9-10 presents the estimates from instrumenting $Integration\ share$ with IDD_{st-2}, MFT_{it-2}, as well as the interaction between the two, $IDD \times MFT_{sit-2}$, in a single two-stage least-squares specification, at the main state-technology class level.[34] Column 10 presents the first stage estimation results of regressing $Integration\ share$ against $IDD \times MFT_{it-2}$. The results confirm that lower mobility (i.e., a higher value of IDD) is associated with lower Integration and that higher value of MFT and the interaction between the two, $IDD \times MFT_{sit-2}$, are associated with more integration. The Hansen test for overidentifying restrictions is consistent with the instruments being valid (p-value for overidentifying restrictions=1.827, Hansen J statistic=0.401). Column 9 presents the second stage estimation results. The coefficient estimate of $Integration\ share$ is negative, yet larger in absolute magnitude than in the pooled OLS estimation results in Column 1. One explanation for the higher coefficient is that the IV is correcting for a measurement problem (i.e., integrated authors is a noisy measure of the true level of integration. For example, my measure might omit cases of actual integration that did not result in a granted patent, but affected scientific outcome), which causes a downward bias to the OLS estimation.

Lastly, to confirm that results are robust to a change in specification, Column

[34] the IV variation is cross-technology and cross-state, but not much over time. Therefore, I do not include firm fixed effects.

11 presents results with Inverse hyperbolic sine transformation. The main results remain robust. Supplementary Table 4.14 presents additional robustness checks for the relationship between integration and publication.[35]

Overall, Table 4.8 shows that the simple patterns reported in Table 4.4 are not just due to differences in the nature of technology and research, firm characteristics, or time. Whereas previous research has focused on the relationship between integration and invention outcomes, Table 4.8 suggests that integration significantly conditions the firm's investment in scientific publications, especially for science-based technology firms.

[35] Supplementary Table 4.14 Columns 1 and 2 distinguish between established publicly traded scientific firms and more recent firms by splitting the sample to firms that entered before and after the year 1990, respectively. To address concerns that publication and patenting patterns might have changed throughout the sample period, in Columns 3 and 4, the panel is split by firm-years prior and post the year 2000, respectively. To examine variation by firm size, in Columns 5 and 6 the sample is split by below and above median R&D, respectively. To address concerns that the results are driven by life science industry, Column 7 excludes life science firms based on related SIC codes. Lastly, Column 8 lags the integration measure by 5 years (1 cohort). Results are robust to all specifications.

Table 4.8: Integration and Discovery

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Dependent variable:	ln(1+Number of publications)									Integration	Inverse hyperbolic sine
	Pub. Count Pooled	Pub. Count Firm FE	Patent stock	Citation Weighted Publications	Top 2 per-centile cited pub.	High Impact Factor Pub.	Low reliance on science in invention	High reliance on science in invention	2nd Stage IV IDD $\times MFT$	1st Stage IV IDD $\times MFT$	ASINH
$Integration_{t-2}$	-0.255**	-0.176**	-0.175**	-0.190**	-0.054**	-0.093*	0.005	-0.449**	-0.696**		-0.208**
	(0.024)	(0.041)	(0.040)	(0.048)	(0.014)	(0.041)	(0.051)	(0.087)	(0.258)		(0.049)
$ln(R\&D\ stock)_{t-2}$	0.085**	0.071**	0.071**	0.076**	0.032**	0.059**	0.027	0.086**	0.060**	0.010**	0.071**
	(0.004)	(0.013)	(0.015)	(0.016)	(0.006)	(0.014)	(0.020)	(0.023)	(0.005)	(0.001)	(0.015)
$ln(Authors)_{t-2}$	0.727**	0.476**	0.477**	0.441**	0.046**	0.351**	0.316**	0.573**	0.817**	0.008**	0.569**
	(0.005)	(0.013)	(0.013)	(0.016)	(0.007)	(0.015)	(0.019)	(0.025)	(0.006)	(0.001)	(0.015)
$ln(Patent\ stock)_{t-2}$			-0.001								
			(0.015)								
MFT_{t-2}										1.504**	
										(0.102)	
IDD_{t-2}										-0.031**	
										(0.007)	
$(IDD \times MFT)_{t-2}$										0.269*	
										(0.118)	
Firm fixed-effects	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-	-	Yes
Tech-class dummies	Yes	No	No	No	No	No	No	No	Yes	Yes	No
Industry dummies	Yes	No	No	No	No	No	No	No	No	No	No
State Dummies	No	No	NO	No	No	No	No	No	Yes	Yes	No
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DV Average	33.39	33.42	33.42	42.14	0.32	19.37	3.70	80.92	33.39	0.22	2.19
Number of firms	1,445	1,445	1,445	1,445	1,445	1,445	727	718	1,445	1,445	1,445
Observations	21,480	21,441	21,441	21,441	21,441	21,441	10,767	10,674	21,480	21,480	21,441
R-squared	0.83	0.88	0.88	0.82	0.65	0.87	0.49	0.84	-	0.12	0.87

Notes: This table presents OLS estimation results for the relationship between integration and annual publications. Integration is defined as the share of a firm's authors who both published an article and were granted a collaborative patent with a specialized inventor during a 5-year-cohort period. Columns 7 and 8 are classified by below and above median use of external science in invention, respectively. Markets for technology (MFT) is based on patent trading activity in invention classes relevant for the focal firm's patent portfolio. IDD dummy is based on a state level recognition of Inevitable Disclosure Doctrine, and is equal to one for firm-years where IDD is effective. Kleibergen-Paap F statistic: F=129>Stock-Yogo CV 5%=14 (J-statistic=0.401). One is added to logged control variables. All specifications include lagged dummies for zero publications per year. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms. ** p<0.01 * p<0.05

4.6.3 Integration and invention

Next, I estimate a patent production function to assess the hypotheses that both research and integration increase downstream invention and R&D productivity:

$$\begin{aligned}\ln(Patents)_{it} = &\omega_0 + \omega_1 \ln\left(Publications\ stock\right)_{it-2} + \omega_2 Integration_{it-2} + \omega_3 \ln(R\&D\ stock)_{it-2} \\ &+ \boldsymbol{Z}'_{it-2}\boldsymbol{\gamma} + \boldsymbol{\eta_i} + \boldsymbol{\tau_t} + \epsilon_{it}\end{aligned} \tag{4.2}$$

In Equation 4.2, $Patents_{it}$ is the annual flow of patents weighted by the number of citations each patent receives divided by the average number of citations received by all other patents granted in the same ipc-year. The main variables of interest are $Publications\ stock_{it-2}$ and $Integration_{it-2}$. Other controls include the stock of R&D and author flow, both lagged by two years.

Internal research can enhance downstream invention both as a direct input to invention as well as indirectly by guiding invention (Fleming and Sorenson, 2004). I thus expect firms with more scientific research stock to be more productive ($\hat{\omega}_1 > 0$). Furthermore, following the prediction in Section 4.3, if integration leads to more downstream invention, I expect $\hat{\omega}_2 > 0$. As shown in Table 4.9, both predictions are confirmed in the data.

Table 4.9 Column 1 presents results from a pooled specification with four-digit SIC dummies and 2-digit main IPC class dummies. There is a positive and statistically significant relationship between number of yearly patents and both integration and publication stock. Column 2 presents the same pattern of results for a within-firm specification: one standard deviation increase in integration is associated with a 16.7% increase in yearly patents - approximately 9 patents per year. Moreover, the marginal effect of an additional publication, evaluated at the sample mean, is equal to approximately 4 patents. The results support the idea that scientific discovery complements the innovation process.

Integration conditions not only the quantity of inventions but also the quality of inventions. Columns 3-5 confirm that results hold for a variety of invention quality measures. In Column 3, patents are weighted by IPC-year normalized forward patent citations. Column 4 includes citation-weighted publication stock as control. Column 5 measures patent originality based on the uniqueness of technology classes combination reported in the patent data.

Columns 6-9 examine how the observed relationship varies by firm characteristics. Columns 6 and 7, divide the firm sample based on below and above-median reliance on science in invention, respectively. Similar to Table 4.8, results indicate that the observed relationship is driven by science-based firms. The coefficient estimate for integration for high reliance on science is statistically significant higher than for low reliance on science ($\hat{\omega}_2$ is 1.041 for high reliance on science, while 0.438 for low reliance on science). For the sample of science-based firms, a one standard deviation increase in integration is associated with a 28% increase in yearly citation-weighted-patents - approximately 32 patents per year. Furthermore, the coefficient estimate for publication stock for high reliance on science is positive and statistically significant, while the estimate for low reliance on science is statistically zero.

Columns 8 and 9 further show that the relationship between integration and yearly citation-weighted patents is stronger for firm-cohorts with early-stage technologies.

Lastly, to confirm that results are robust to a change in specification, Column 10 presents results with Inverse hyperbolic sine transformation. The main results remain robust.

Taken together, Table 4.8 and Table 4.9 support the conjecture that integration is related to a tradeoff between short-term and long-term R&D initiatives - because integration is negatively related to publication (as presented in Table 4.8), which in turn will have an adverse effect on inventions (as presented in Table 4.9). The direct increase in patents due to a one standard deviation increase in integration drops by approximately 60% (from 9 patents to 3.8 patents)[36] due to the indirect negative relationship between integration and publications.

[36] the net increase equals to $9 - 1.3 \times 4 = 3.8$ patents

Table 4.9: Integration and Invention

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable:	ln(1+Number of patents)									Inverse hyperbolic sine
	Patent Count Pooled	Patent Count Firm FE	Citation Weighted Patents	Citation Weighted Patents & Publica-tions	Original Patents (IPC com-bination)	Low reliance on science in invention	High reliance on science in invention	Old Tech	New Tech	ASINH
$Integration_{t-2}$	0.998**	0.757**	0.739**	0.742**	0.615**	0.438**	1.041**	0.639**	1.445**	0.855**
	(0.031)	(0.058)	(0.064)	(0.064)	(0.053)	(0.061)	(0.114)	(0.062)	(0.222)	(0.064)
$ln(Publication\ stock)_{t-2}$	0.061**	0.114**	0.095**	0.044**	0.100**	-0.000	0.118**	0.037**	0.048	0.114**
	(0.009)	(0.029)	(0.031)	(0.014)	(0.024)	(0.014)	(0.025)	(0.014)	(0.032)	(0.033)
$ln(R\&D\ stock)_{t-2}$	0.397**	0.247**	0.225**	0.229**	0.208**	0.219**	0.173**	0.216**	0.235**	0.282**
	(0.007)	(0.028)	(0.028)	(0.027)	(0.023)	(0.032)	(0.042)	(0.027)	(0.068)	(0.031)
$ln(Authors)_{t-2}$	0.287**	0.190**	0.198**	0.214**	0.162**	0.146**	0.243**	0.180**	0.237**	0.216**
	(0.009)	(0.017)	(0.018)	(0.019)	(0.015)	(0.020)	(0.034)	(0.019)	(0.058)	(0.019)
Firm fixed-effects	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Tech-class dummies	Yes	No	No	No	No	No	No	No	No	No
Industry dummies	Yes	No	No	No	No	No	No	No	No	No
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DV Average	53.36	53.44	61.70	61.70	24.99	8.66	115.20	30.44	261.67	2.70
Number of firms	1,445	1,445	1,445	1,445	1,445	727	718	1385	338	1,445
Observations	21,480	21,441	21,441	21,441	21,441	10,767	10,674	18,532	2,898	21,441
R-squared	0.77	0.88	0.85	0.85	0.86	0.71	0.84	0.80	0.91	0.87

Notes: This table presents OLS estimation results of a patent equation- examining the relationship between R&D productivity and integration. Integration is defined as the share of a firm's authors who both published an article and were granted a collaborative patent with a specialized inventor during a 5-year-cohort period. Columns 6 and 7 are classified by below and above median use of external science in invention, respectively. All specifications include lagged dummies for zero publications per year and zero patents per year. One is added to logged control variables. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms. ** p<0.01 * p<0.05

4.6.4 Integration and market value

If integration is positively related to invention quality and negatively related to scientific discovery quality, this should be reflected not only in the level of patent and publication output but, more importantly, in the firm's value. I examine next the relationship between integration and firm stock market value and estimate the following Tobin's Q specification following Griliches (1986) and Hall et al. (2005) as well as more recent work by Simeth and Cincera (2016) and Arora et al. (2021a).[37]

$$\begin{aligned}\ln\frac{Value_{it}}{Assets_{it}} =&\alpha_0\frac{G_{it-2}}{Assets_{it}} + \alpha_1 Integration_{it-2} * \frac{\ln(Publication\ stock)_{it-2}}{Assets_{it}} \\ &+ \alpha_2 Integration_{it-2} * \frac{\ln(Patent\ stock)_{it-2}}{Assets_{it}} \\ &+ \alpha_3 Integration_{it-2} + \boldsymbol{Z}'_{it-1}\gamma + \boldsymbol{\eta_i} + \boldsymbol{\tau_t} + \epsilon_{it}\end{aligned} \tag{4.3}$$

G is knowledge assets, measured as the perpetual stocks of publications and patents. The main interest is at coefficients α_1 and α_2, which estimate the interaction between integration and publication stock and patent stock, respectively.

Consistent with the results for publication and patent equation, I expect the positive relationship between the market value of a firm and its stock of invention to be stronger with integration. Conversely, the positive relationship between market value and scientific knowledge stock should be weaker with integration. Thus, $\hat{\alpha_1} < 0$ and $\hat{\alpha_2} > 0$.

Table 4.10 presents the estimation results. Building on Simeth and Cincera (2016) and Arora et al. (2021a), Column 1 presents the break up into publication and patent stocks, indicating a positive value for both patents and publication stocks. Column 2 adds the interaction between integration and citation-weighted patents

[37] Market value is the sum of common stock, preferred stock, and total debt net of current assets. Tobin's Q is market value over assets

and publication stocks. As expected, the coefficient estimate of the interaction with publication stock is negative ($\hat{\alpha_1} < 0$), and the estimate of interaction with patent stock is positive $\hat{\alpha_2} > 0$. Both estimates are statistically different from zero. In Column 3, the relationship endures even after controlling for firm fixed effects. Lastly, as a robustness check, Column 4 presents a within-firm estimation result for a market value specification. Results hold, and both interaction estimates are statistically different from zero.

Overall, the results in Table 4.10 are consistent with the idea that value creation is conditioned by organizational structure (Arora et al., 2014). In particular, the private value of publications decreases, and the private value of patents increases when firms integrate scientific discovery with invention.

Table 4.10: Integration and Market Value

	(1)	(2)	(3)	(4)
Dependent variable:	ln(Tobin's Q)			ln(Market Value)
	Base (Pooled)	Integration (Pooled)	Integration (Firm FE)	Integration (Firm FE)
$Integration_{t-2}$ ×				
$Publication\ stock_{t-2}/Assets$		-0.118**	-0.067**	
		(0.021)	(0.021)	
$Patent\ stock_{t-2}/Assets$		0.091**	0.042*	
		(0.020)	(0.020)	
$ln(Publication\ stock)_{t-2}$				-0.097**
				(0.028)
$ln(Patent\ stock)_{t-2}$				0.109**
				(0.027)
$Integration_{t-2}$		0.100*	-0.172**	-0.286**
		(0.040)	(0.040)	(0.101)
$Publication\ stock_{t-2}/Assets$	0.025**	0.038**	0.015	
	(0.005)	(0.007)	(0.008)	
$Patent\ stock_{t-2}/Assets$	0.047**	0.017*	0.022**	
	(0.005)	(0.007)	(0.009)	
$R\&D\ stock_{t-2}/Assets$	0.080**	0.067**	0.051**	
	(0.005)	(0.005)	(0.007)	
$Authors_{t-2}/Assets$		0.047**	0.056**	
		(0.005)	(0.006)	
$ln(Publication\ stock)_{t-2}$				-0.063**
				(0.015)
$ln(Patent\ stock)_{t-2}$				-0.043**
				(0.015)
$ln(R\&D\ stock)_{t-2}$				-0.003
				(0.014)
$ln(Authors)_{t-2}$				0.134**
				(0.012)
$ln(Assets)_{t-2}$				0.278**
				(0.013)
Firm fixed-effects	-	-	Yes	Yes
Tech-class dummies	Yes	Yes	No	No
Industry dummies	Yes	Yes	No	No
Year dummies	Yes	Yes	Yes	Yes
DV Average	5	5	5	7,234
Number of firms	1,424	1,424	1,424	1,424
Observations	20,107	20,107	20,061	20,044
R-squared	0.51	0.52	0.71	0.87

Notes: This table presents OLS estimation results for the relationship between integration and value. Tobin's-Q is the ratio of market value to assets. Integration is defined as the share of a firm's authors who both published an article and were granted a collaborative patent with a specialized inventor during a 5-year-cohort period. Patents are weighted by IPC-year normalized forward patent citations. Publications are weighted by journal-year normalized forward publication citations. One is added to logged control variables. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms. ** p<0.01 * p<0.05

4.6.5 The determinants of integration and market value

Table 4.11 further examines how the determinants of integration condition the private value of publications and patents. I start by examining the nature of research in the field. Columns 1 and 2, divide the firm sample based on below and above-median use of external science in invention, respectively. The results are consistent with the idea that the opportunity cost of integration for more fundamental-science-based firms is higher; integration requires them to depart further away from the traditional research in their field, but the benefits of integration, as reflected in the private value of patents, are also higher (i.e., by connecting two distinct practices).

Next, I examine the stage of technology in the field. Columns 3 and 4 divide the firms based on their investments in new technology in each cohort. The results are consistent with the idea that the benefit of integration in terms of patent value is stronger when firms invest in early-stage technology that is more tightly connected to scientific research ($\hat{\alpha_2}$ is 0.241 for new technology, while the estimated coefficient for old technology is statistically zero)

Lastly, I explore how the market for technology conditions the results. Columns 5 and 6 divide firms into two groups based on the median value of MFT in each cohort. The results indicate that the cost of integration in terms of scientific discovery value is prominent for low MFT ($\hat{\alpha_1}$ is 0.106 for low MFT, while the coefficient estimate for high MFT is statistically zero). This is consistent with the idea that when MFT is low, firms need to rely on internal development for breakthroughs in the long run. That is, integration becomes more harmful, as scientific discovery is essential for breakthroughs.

Overall, Table 4.11 suggests that firms relying on internal scientific discovery for value creation should find specialization more compatible with their objective. In contrast, firms that focus on more applied, early-stage technology initiatives, possibly combined with external technology sourcing in the long-term, are best served by an integrated structure.

Table 4.11: Determinants of Integration and Market Value

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	ln(Tobin's Q)					
	Low reliance on science in invention	High reliance on science in invention	Established Technology	Early Stage Technology	Low MFT	High MFT
$Integration_{t-2}$ ×						
$Publication\ stock_{t-2}/Assets$	-0.040	-0.085**	-0.057**	-0.211*	-0.106*	-0.032
	(0.030)	(0.030)	(0.022)	(0.093)	(0.044)	(0.027)
$Patent\ stock_{t-2}/Assets$	-0.003	0.079**	0.034	0.241*	0.054	0.024
	(0.030)	(0.028)	(0.021)	(0.098)	(0.036)	(0.026)
$Integration_{t-2}$	-0.111*	-0.239**	-0.158**	-0.502**	-0.253**	-0.091
	(0.054)	(0.061)	(0.042)	(0.153)	(0.065)	(0.056)
$Publication\ stock_{t-2}/Assets$	0.033**	-0.005	0.012	0.065	0.026	0.005
	(0.012)	(0.012)	(0.009)	(0.034)	(0.015)	(0.012)
$Patent\ stock_{t-2}/Assets$	0.017	0.027*	0.022*	0.003	0.020	0.024*
	(0.013)	(0.012)	(0.009)	(0.037)	(0.015)	(0.011)
$R\&D\ stock_{t-2}/Assets$	0.053**	0.054**	0.052**	0.026	0.059**	0.048**
	(0.012)	(0.009)	(0.008)	(0.030)	(0.014)	(0.009)
$Authors_{t-2}/Assets$	0.032**	0.072**	0.057**	0.081**	0.058**	0.053**
	(0.009)	(0.008)	(0.006)	(0.023)	(0.010)	(0.008)
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
DV Average	3.62	5.46	4.73	3.46	3.60	5.46
Observations	9,852	10,209	17,257	2,790	9,766	10,229
R-squared	0.68	0.72	0.72	0.79	0.72	0.76

Notes: This table presents OLS estimation results for the cross partial relationship between value and, integration and various determinants. Tobin's-Q is the ratio of market value to assets. Integration is defined as the share of a firm's authors who both published an article and were granted a collaborative patent with a specialized inventor during a 5-year-cohort period. Columns 1 and 2 are classified by below and above median use of external science in invention, respectively. Columns 3 and 4, divide the firms based on their investment in new technology in each cohort. New technology is defined as patents granted no more than 10 years from the related IPC class inception year. Markets for technology (MFT) is based on patent trading activity in invention classes relevant for the focal firm's patent portfolio. Columns 5 and 6 are classified by below and above the median value of MFT, respectively. Patents are weighted by IPC-year normalized forward patent citations. Publications are weighted by journal-year normalized forward publication citations. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms. ** p<0.01 * p<0.05

4.7 Conclusion and discussion

Using a novel firm-level measure of integration, the paper bridges two streams of literature, innovation and organization, to try and get a better understanding of the tradeoffs and determinants related to the organization of scientific discovery at the firm level. While past research mainly highlighted the benefits of integration, in the current paper, I examine both the benefits and the costs of integration. I show that while integration guides the firm's short-term invention search process, specialization supports its long-term fundamental R&D initiatives. I present three main determinants that condition this tradeoff: MFT, reliance on science, and stage of technology.

The results suggest that firms might be losing their scientific capabilities if they tie their internal science too tightly to the firm's short-term invention needs. This has adverse effects on long-term scientific capabilities, possibly pushing away top scientists with a taste for basic science (Stern, 2004), which eventually leads to deterioration long tun invention quality.

Though specialization is important for long-term significant breakthroughs, firms cannot immediately appropriate the benefits of their investment in basic science (Nelson, 1959; Arrow, 1962). Thus, it might be the case that the tradeoff that I document in this paper is not immediately apparent to firms, as the feedback loop from basic scientific discovery to invention is only manifested in the long-term. With no clear effect in the short-term, firms might over integrate, or sequentially jump from one organization structure to another (Hounshell and Smith, 1988).

Therefore, managers must understand how to organize scientific discovery while correctly balancing short-term and long-term R&D initiatives (Laverty, 1996; Davila et al., 2006). In fact, it has been argued that the creation of Nylon at Dupont would not have been possible without the right blend of integrated research, connected to the immediate needs of the product units, the view led by the research director at that time, Elmer K. Bolton, and that of the specialized basic research, led by Charles Stine (Davila et al., 2006). This balanced view is also embedded in the concept of ambidextrous organizations that pursue both exploration and exploitation. The research on ambidexterity proposed several approaches for balancing exploration and exploitation,including, simultaneously engaging in exploration and exploitation, temporal sequential engagement, and balance of exploration and exploitation across different organization modes (March, 1991; Tushman and O'Reilly III, 1996; Lavie et al., 2010; O'Reilly III and Tushman, 2013; Stettner and Lavie, 2014). Future work can similarly design optimal structure for scientific discovery organization that would maximize the impact of corporate science on invention, and at the same time, protect its long-term properties.[38]

Over the years, there has been an increase in the availability of external sources of invention from small firms and universities (Pisano, 2010; Arora et al., 2020a). There has also been a change in the research focus of universities towards more applied fields that might affect the preferences of scientists who join the industry and the availability of external knowledge sources. Furthermore, the nature of research in several fields have changed (e.g., the emergence of bio-science and biotechnology). These changes may partly explain the relationships I document for integration. Future research should explore how these and other changes affect trends in the organization of scientific discovery.

In the current paper, I present external market for technology as an important external determinant of integration. Interestingly, Arora et al. (2021a) find that while internal investment in science is declining, firms are increasingly building on external knowledge over the years. If external sourcing is, in fact, substituting for corporate

[38] For example, Marginson and McAulay (2008: p.274) suggest that: "Balance may be achieved through diversity"- a mix of short-term oriented teams and long-term oriented teams that will challenge the firm to perform well both in the short and long-term.

research, it is important for managers to understand how to organize their firm's scientific discovery to best capitalize on the declining internal science while optimally capturing external opportunities. In future work, I further explore this relationship between integration and external sourcing and show evidence that integration not only guides the firm's internal search process but also its external technology search (Sheer, 2021).

Finally, it is interesting to compare the results in the current paper to the findings on academic scientists' engagement with commercialization. While originally there were concerns that engagement in commercialization may dilute basic research in academia by imposing conflicting norms (Merton, 1973; Dasgupta and David, 1994b; Argyres and Liebeskind, 1998), more recent research suggests that is not always the case (Murray and Stern, 2006; Breschi et al., 2008; Goldfarb et al., 2009; Azoulay et al., 2009; Thursby and Thursby, 2011; Banal-Estañol et al., 2015; Bikard et al., 2019). For example, Azoulay et al. (2009) show that patenting by academic scientists is positively related to the quantity and quality of their publications. Thursby and Thursby (2011) further suggest that both basic and applied research is greater when faculty have an interest in the commercialization of their research effort. Nevertheless, comparing simultaneous discoveries of scientists who collaborate with the industry with those who do not, Bikard et al. (2019) find that it is scientists who collaborate with industry on projects with both scientific and commercial potential who increase their research output. Their result suggests that academic scientists achieve greater levels of specialization in their basic research when leaving the commercial aspects to their industry partner. This latter finding echoes the idea of specialization that I present in the current paper - when there is a clear separation between research and invention activity, research output increases. Future research can further examine the differences between the division of labor within versus across institutions.

I acknowledge the limitation of the paper in terms of relying on patent and scientific publications data. First, these data are only sufficient for identifying instances of integration where an author-inventor successfully files patents and publishes scientific papers. As previously mentioned, these instances vary across industries and technology fields. To mitigate this concern, my measure of integration does not rely on the quantity of publications but identifies an author as an individual who had at least one publication during a 5-year cohort period.

Second, as acknowledged in the innovation literature, using patent data as a proxy of invention is not without problems. For example, some firms may choose to keep their inventions as trade secrets, and there is also variation in the use of patents across industries (Cohen et al., 2000). Lastly, patents are also different from commercial success.

4.8 Supplementary Results

Table 4.12: Integration Case Studies

COHORT	(1) IBM	(2) AT&T
1980-1985	0.25	0.21
1986-1990	0.24	0.21
1991-1995	0.35	0.30
1996-2000	0.50	0.41
2001-2005	0.58	0.45
2006-2010	0.63	N/A
2011-2015	0.69	N/A

Notes: The table presents the measure of integration for each firm-cohort.

Table 4.13: Robustness Checks for Integration Measure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample:	Excluding new journals		Excluding low JIF journals		Co-location (Pooled)		Co-location (Firm FE)	
Dependent variable:	ln(1+No. of publications)	ln(1+No of patents)	ln(1+No. of publications)	ln(1+No. of patents)	ln(1+No. of publications)	ln(1+Number of patents)	ln(1+No. of publications)	ln(1+No. of patents)
$Integration_{t-2}$	-0.231**	0.693**	-0.187**	0.613**				
	(0.037)	(0.060)	(0.041)	(0.060)				
$Co-location_{t-2}$				-0.031*	0.134**	-0.006	0.067	
				(0.016)	(0.018)	(0.032)	(0.039)	
$ln(Publication\ stock)_{t-2}$		0.141**		0.159**		0.039**		0.099**
		(0.031)		(0.034)		(0.011)		(0.032)
$ln(R\&D\ stock)_{t-2}$	0.060**	0.261**	0.055**	0.259**	0.092**	0.401**	0.073**	0.235**
	(0.013)	(0.030)	(0.014)	(0.035)	(0.006)	(0.007)	(0.017)	(0.029)
$ln(Authors)_{t-2}$	0.446**	0.151**	0.438**	0.158**	0.765**	0.337**	0.441**	0.203**
	(0.013)	(0.018)	(0.014)	(0.019)	(0.006)	(0.010)	(0.016)	(0.019)
Firm fixed-effects	Yes	Yes	Yes	Yes	-	-	Yes	Yes
Tech-class dummies	No	No	No	No	Yes	Yes	No	No
Industry dummies	No	No	No	No	Yes	Yes	No	No
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DV Average	37.18	53.44	25.83	68.58	42.78	62.65	42.83	62.76
Number of firms	1,266	1,266	1,055	1,055	1,408	1,408	1,408	1,408
Observations	19,214	19,214	16,023	16,023	21,126	21,126	21,072	21,072
R-squared	0.88	0.87	0.87	0.88	0.74	0.70	0.82	0.84

Notes: This table presents robustness checks for integration measure. In Columns 1 and 2 Integration is computed excluding scientific publications from new journals post 1990. In Columns 3 and 4 Integration is computed excluding scientific publications with below median JIF. In Columns 5-8 integration is defined based on co-location of inventors and authors. All specifications include lagged dummies for zero publications per year and/or zero patents per year. One is added to logged control variables. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms. ** p<0.01 * p<0.05

Table 4.14: Robustness Checks for Integration and Scientific Discovery

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Dependent variable:				ln(1+No. of publications)
Sample:	Established scientific firms (pre-1990)	Young scientific firms (post-1990)	Sample pre-2000	Sample post-2000	Small firms	Large firms	Excluding life-science	Lag 5
$Integration_{t-2}$	-0.168**	-0.190**	-0.189**	-0.176**	-0.132**	-0.240**	-0.165**	
	(0.058)	(0.053)	(0.063)	(0.049)	(0.044)	(0.074)	(0.045)	
$ln(R\&D\ stock)_{t-2}$	0.086**	0.069**	0.068**	0.068**	0.008	0.072**	0.065**	0.229**
	(0.018)	(0.024)	(0.018)	(0.018)	(0.020)	(0.018)	(0.015)	(0.031)
$ln(Authors)_{t-2}$	0.548**	0.380**	0.473**	0.327**	0.344**	0.560**	0.489**	
	(0.017)	(0.018)	(0.018)	(0.015)	(0.015)	(0.019)	(0.014)	
$Integration_{t-5}$								-0.140*
								(0.060)
$ln(Authors)_{t-5}$								0.201**
								(0.019)
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DV Average	49.67	12.49	35.35	31.85	3.68	64.10	27.12	46.14
Number of firms	585	860	987	1099	913	819	1,004	1,250
Observations	12,066	9,375	10,038	11,329	9,959	10,550	16,581	17,088
R-squared	0.91	0.81	0.91	0.89	0.67	0.92	0.88	0.82

Notes: This table presents robustness checks for the relationship between publication and integration. Columns 1 and 2 distinguishes between firms that enter the sample before and after 1990, respectively. In Columns 3 and 4 the panel is split by firm-years prior and post the year 2000, respectively. In Columns 5 and 6 the sample is split by below and above median R&D, respectively. Column 7 excludes life science related firms based on SIC codes. All specifications include lagged dummies for zero publications per year. One is added to logged control variables. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms. ** p<0.01 * p<0.05

Figure 4.1: Trend in Integration - Technology

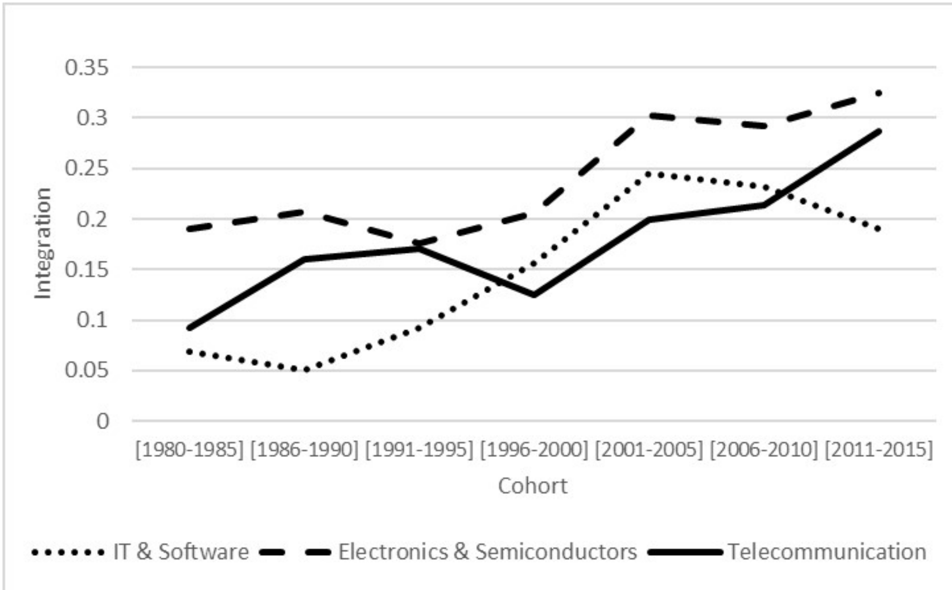
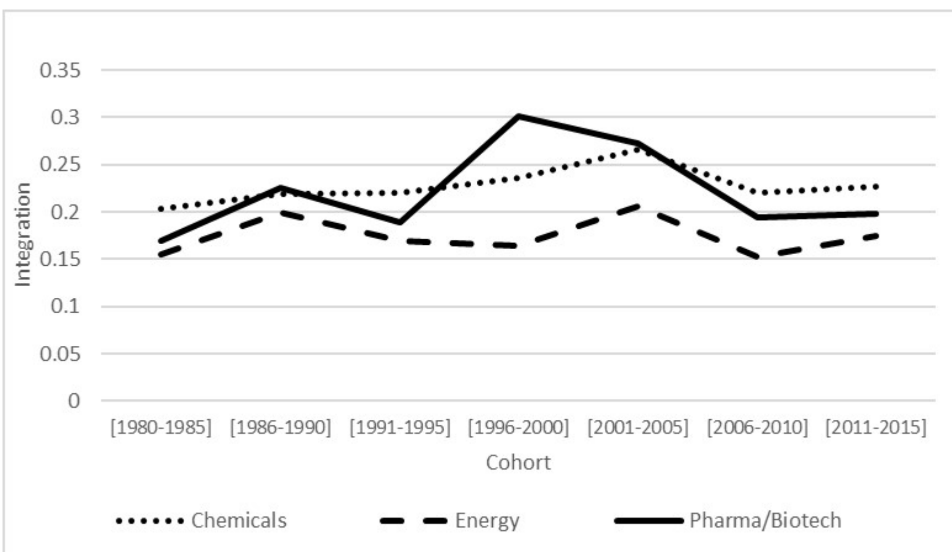


Figure 4.2: Trend in Integration - Life Science



5

Conclusions

Building on a novel dataset that I construct, I trace above 4,000 U.S. publicly traded firms' investment in science and invention for 35 years (1980-2015). Over the sample period, I observe a clear reduction in investment in corporate scientific research, changes in patterns of use of science in invention (chapter three), and changes in the organization of research and invention practices within firms (chapter four).

My research contributes to the literature on organizing for innovation (Kay, 1988; Argyres and Silverman, 2004; Arora et al., 2014; Argyres et al., 2019; Aggarwal et al., 2020). The findings from chapter three and chapter four suggest that as firms make greater use of external scientific knowledge, and as they rely more on external inventions, while becoming internally more integrated, they are less likely to invest in internal scientific research. This shift, though likely privately profitable, is not without social costs. The declining corporate engagement in research may be contributing to the reported decline in R&D productivity and the associated decline in productivity growth. Furthermore, it is unclear whether corporations can sustain their breakthroughs in the long run by relying on external sources. The type of research conducted in large firms is different in its nature than that undertaken in universities and small firms. Large firms have access to complementary resources and can tackle multidisciplinary problems more easily than universities and small firms. Furthermore, university research needs further development and integration to produce inventions that can be commercialized (Arora et al., 2020a). Future studies should further examine the changing role of science throughout the complete innovation ecosystem (large firms, smaller firms, and universities) to better understand the long-term implications of the reduction in corporate science.

Lastly, this dissertation features an important extension and improvement to data on corporate patents and scientific publications. The data are publicly available to researchers and can aid future studies explore corporate investment in scientific discovery and invention.

Bibliography

Aggarwal, V. A., Hsu, D. H., and Wu, A. (2020), “Organizing knowledge production teams within firms for innovation,” *Strategy Science*, 5, 1–16.

Aghion, P. and Jaravel, X. (2015), “Knowledge spillovers, innovation and growth,” *Economic Journal*, 125, 533–573.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., and Howitt, P. (2005), “Competition and innovation: An inverted-U relationship,” *The Quarterly Journal of Economics*, 120, 701–728.

Akcigit, U., Hanley, D., and Serrano-Velarde, N. (2013), “Back to Basics: Basic Research Spillovers, Innovation Policy and Growth,” Working Paper 19473, National Bureau of Economic Research.

Akcigit, U., Grigsby, J., and Nicholas, T. (2017), “The rise of American ingenuity: Innovation and inventors of the golden age,” NBER Working Papers: No. 23047.

Allen, T. J. (1966), “Managing the flow of scientific and technological information.” Ph.D. thesis, Massachusetts Institute of Technology.

Allen, T. J. (1969), “Roles in technical communication networks,” .

Allen, T. J. (1984), “Managing the flow of technology: Technology transfer and the dissemination of technological information within the R&D organization,” *MIT Press Books*, 1.

Allen, T. J. and Cohen, S. I. (1966), “Information flow in an R & D laboratory,” .

Allen, T. J., Tushman, M. L., and Lee, D. M. (1979), “Technology transfer as a function of position in the spectrum from research through development to technical services,” *Academy of management journal*, 22, 694–708.

Argyres, N., Rios, L. A., and Silverman, B. S. (2019), “Organizational change and the dynamics of innovation: Formal R&D structure and intrafirm inventor networks,” *Strategic Management Journal*.

Argyres, N. S. and Liebeskind, J. P. (1998), “Privatizing the intellectual commons: Universities and the commercialization of biotechnology,” *Journal of Economic Behavior & Organization*, 35, 427–454.

Argyres, N. S. and Silverman, B. S. (2004), “R&D, organization structure, and the development of corporate technological knowledge,” *Strategic Management Journal*, 25, 929–958.

Arora, A. (1996), “Testing for complementarities in reduced-form regressions: A note,” *Economics letters*, 50, 51–55.

Arora, A. and Gambardella, A. (1990), “Complementarity and external linkages: the strategies of the large firms in biotechnology,” *The journal of industrial economics*, pp. 361–379.

Arora, A. and Gambardella, A. (1994), “The changing technology of technological change: general and abstract knowledge and the division of innovative labour,” *Research policy*, 23, 523–532.

Arora, A., Fosfuri, A., and Gambardella, A. (2001), “Markets for technology and their implications for corporate strategy,” *Industrial and corporate change*, 10, 419–451.

Arora, A., Belenzon, S., and Rios, L. A. (2014), “Make, buy, organize: The interplay between research, external knowledge, and firm structure,” *Strategic Management Journal*, 35, 317–337.

Arora, A., Belenzon, S., and Patacconi, A. (2018), “The Decline of Science in Corporate R&D,” *Strategic Management Journal*, 39, 3–32.

Arora, A., Belenzon, S., Patacconi, A., and Suh, J. (2020a), “The changing structure of American innovation: Some cautionary remarks for economic growth,” *Innovation Policy and the Economy*, 20, 39–93.

Arora, A., Belenzon, S., and Suh, J. (2020b), “Science and the Market For Technology,” *Working Paper*.

Arora, A., Belenzon, S., and Sheer, L. (2021a), “Knowledge spillovers and corporate investment in scientific research,” *American Economic Review*, 111, 871–898.

Arora, A., Belenzon, S., and Sheer, L. (2021b), “Matching Patents to Compustat Firms, 1980-2015: Dynamic Reassignment, Name Changes, and Ownership Structures,” *Research Policy*, 50, 104217.

Arrow, K. (1962), “Economic welfare and the allocation of resources for invention,” in *The rate and direction of inventive activity: Economic and social factors*, pp. 609–626.

Audretsch, D. B. P. E. S. (1996), "Company-scientist locational links: The case of biotechnology," *American Economic Review*, 86, 641–652.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Reenen, J. V. (2020), "The fall of the labor share and the rise of superstar firms," *The Quarterly Journal of Economics*, 135, 645–709.

Azoulay, P. (2002), "Do pharmaceutical sales respond to scientific evidence?" *Journal of Economics Management Strategy*, 11, 551–594.

Azoulay, P., Ding, W., and Stuart, T. (2009), "The impact of academic patenting on the rate, quality and direction of (public) research output," *The Journal of Industrial Economics*, 57, 637–676.

Azoulay, P., Zivin, J. G., Li, D., and Sampat, B. N. (2019), "Public R&D investments and private-sector patenting: evidence from NIH funding rules," *The Review of economic studies*, 86, 117–152.

Banal-Estañol, A., Jofre-Bonet, M., and Lawson, C. (2015), "The double-edged sword of industry collaboration: Evidence from engineering academics in the UK," *Research Policy*, 44, 1160–1175.

Bei, X. (2018), "Boundary Choices in Innovation: How does the Availability of Hiring Affect Firms' Technology Sourcing Strategies?" Tech. rep., Working paper.

Belenzon, S. (2012), "Cumulative innovation and market value: evidence from patent citations," *Economic Journal*, 122, 265–285.

Bena, J., Ferreira, M. A., Matos, P., and Pires, P. (2017), "Are foreign investors locusts? The long-term effects of foreign institutional ownership," *Journal of Financial Economics*, 126, 122–146.

Bessen, J. (2009), "NBER PDP Project User Documentation," Data available at: `https://sites.google.com/site/patentdataproject/Home/downloads?authuser=0` (Accessed: April, 2016).

Bhaskarabhatla, A. and Hegde, D. (2014), "An organizational perspective on patenting and open innovation," *Organization Science*, 25, 1744–1763.

Bikard, M., Vakili, K., and Teodoridis, F. (2019), "When collaboration bridges institutions: The impact of university–industry collaboration on academic productivity," *Organization Science*, 30, 426–445.

Bloom, N., Schankerman, M., and Reenen, J. V. (2013), "Identifying technology spillovers and product market rivalry," *Econometrica*, 81, 1347–1393.

Bloom, N., Jones, C. I., Reenen, J. V., and Webb, M. (2017), "Are ideas getting harder to find?" NBER Working Papers: No. 23782.

Blundell, R., Griffith, R., and van Reenen, J. (1999), "Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms," *The Review of Economic Studies*, 66, 529–554.

Bonaccorsi, A. and Thoma, G. (2007), "Institutional complementarity and inventive performance in nano science and technology," *Research policy*, 36, 813–831.

Breschi, S. and Catalini, C. (2010), "Tracing the links between science and technology: An exploratory analysis of scientists' and inventors' networks," *Research Policy*, 39, 14–26.

Breschi, S., Lissoni, F., and Montobbio, F. (2008), "University patenting and scientific productivity: a quantitative study of Italian academic inventors," *European Management Review*, 5, 91–109.

Bryan, K. A., Ozcan, Y., and Sampat, B. (2020), "In-text patent citations: A user's guide," *Research Policy*, 49, 103946.

Bureau van Dijk (2018), "ORBIS Ownership files, 2002-2015," Provided via a Duke University subscription service.

Bush, V. (1945), "Science: The endless frontier," *Transactions of the Kansas Academy of Science (1903-1945)*, 48, 231–264.

Carnabuci, G. and Operti, E. (2013), "Where do firms' recombinant capabilities come from? Intraorganizational networks, knowledge, and firms' ability to innovate through technological recombination," *Strategic management journal*, 34, 1591–1613.

Cassiman, B. and Veugelers, R. (2006), "In search of complementarity in innovation strategy: Internal R&D and external knowledge acquisition," *Management science*, 52, 68–82.

Cassiman, B. and Veugelers, R. (2007), "Are External Technology Sourcing Strategies Substitutes or Complements?: The Case of Embodied versus Disembodied Technology Acquisition," .

Ceccagnoli, M. (2005), "Firm heterogeneity, imitation, and the incentives for cost reducing R&D effort," *The Journal of Industrial Economics*, 53, 83–100.

Center for Research in Security Prices (2018a), "CRSP Compustat Merged (Monthly), 1980-2015," Available to Duke University through Wharton Research Data Services (WRDS). `https://wrds-web.wharton.upenn.edu/wrds/ds/crsp/ccm_m/linktable/index.cfm?navId=137` (Accessed: August 2018).

Center for Research in Security Prices (2018b), "CRSP Stock (Monthly), 1980-2015," Available to Duke University through Wharton Research Data Services (WRDS). https://wrds-web.wharton.upenn.edu/wrds/ds/crsp/stock_m/msf.cfm?navId=145 (Accessed: August 2018).

Clarivate Analytics (2016), "Web of Science (WoS) Core Collection XML, 1900-2016," Clarivate Analytics, Philadelphia, PA. Obtained from Clarivate Analytics by license in 2017.

Clark, K. B., Chew, W. B., Fujimoto, T., Meyer, J., and Scherer, F. (1987), "Product development in the world auto industry," *Brookings Papers on economic activity*, 1987, 729–781.

Cockburn, I. M. and Henderson, R. M. (1998), "Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery," *The Journal of Industrial Economics*, 46, 157–182.

Cohen, W. M. and Klepper, S. (1996), "Firm size and the nature of innovation within industries: the case of process and product R&D," *The review of Economics and Statistics*, pp. 232–243.

Cohen, W. M. and Levinthal, D. A. (1989), "Innovation and learning: the two faces of R & D," *Economic Journal*, 99, 569–596.

Cohen, W. M. and Levinthal, D. A. (1990), "Absorptive capacity: A new perspective on learning and innovation," *Administrative science quarterly*, pp. 128–152.

Cohen, W. M., Nelson, R. R., and Walsh, J. P. (2000), "Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not)," Tech. rep., National Bureau of Economic Research.

Cohen, W. M., Sauermann, H., and Stephan, P. (2020), "Not in the job description: The commercial activities of academic scientists and engineers," *Management Science*.

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh (2000), "Carnegie Mellon Survey Data," unpublished data.

Dasgupta, P. and David, P. A. (1994a), "Toward a new economics of science," *Research policy*, 23, 487–521.

Dasgupta, P. and David, P. A. (1994b), "Toward a new economics of science," *Research policy*, 23, 487–521.

David, P. A., Mowery, D., and Steinmueller, E. W. (1992), "Analysing the economic payoffs from basic research," *Economics of Innovation and New Technology*, 2, 73–90.

Davila, T., Epstein, M. J., and Shelton, R. D. (2006), *The creative enterprise: Managing innovative organizations and people*, Greenwood Publishing Group.

Duguet, E. and MacGarvie, M. (2005), "How well do patent citations measure flows of technology? Evidence from French innovation surveys," *Economics of Innovation and New Technology*, 14, 375–393.

European Patent Office (2016), "The EPO Worldwide Patent Statistical Database, 1980-2015," PATSTAT Global single edition 2016. Obtained from EPO by license in 2016.

Faulkner, W. (1994), "Conceptualizing knowledge used in innovation: A second look at the science-technology distinction and industrial innovation," *Science, technology, & human values*, 19, 425–458.

Figueroa, N. and Serrano, C. J. (2019), "Patent trading flows of small and large firms," *Research Policy*, 48, 1601–1616.

Fleming, L. (2001), "Recombinant uncertainty in technological search," *Management science*, 47, 117–132.

Fleming, L. and Sorenson, O. (2004), "Science as a map in technological search," *Strategic management journal*, 25, 909–928.

Galasso, A. and Schankerman, M. A. (2014), "Patents and cumulative innovation: Causal evidence from the courts," *The Quarterly Journal of Economics*, 130, 317–369.

Gambardella, A. (1992), "Competitive advantages from in-house scientific research: The US pharmaceutical industry in the 1980s," *Research Policy*, 21, 391–407.

Gittelman, M. and Kogut, B. (2003), "Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns," *Management Science*, 49, 366–382.

Goldfarb, B., Marschke, G., and Smith, A. (2009), "Scholarship and inventive activity in the university: complements or substitutes?" *Economics of Innovation and New Technology*, 18, 743–756.

Gomory, R. E. (1989), "Moving IBM's technology from research to development," *Research-Technology Management*, 32, 27–32.

Griffith, R., Redding, S., and Reenen, J. V. (2004), "Mapping the two faces of R&D: Productivity growth in a panel of OECD industries," *Review of Economics and Statistics*, 86, 883–895.

Griffith, R., Harrison, R., and Reenen, J. V. (2006), "How special is the special relationship? Using the impact of US R&D spillovers on UK firms as a test of technology sourcing," *American Economic Review*, 96, 1859–1875.

Griliches, Z. (1986), "Productivity, R and D, and Basic Research at the Firm Level in the 1970's," *American Economic Review*, 76, 141–154.

Gutiérrez, G. and Philippon, T. (2017), "Declining Competition and Investment in the US," Tech. rep., National Bureau of Economic Research.

Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2001), "The NBER patent citation data file: Lessons, insights and methodological tools," NBER Working Papers: No. 8498, National Bureau of Economic Research. Data available at: `http://data.nber.org/patents/` (Accessed: April 2016).

Hall, B. H., Jaffe, A., and Trajtenberg, M. (2005), "Market value and patent citations," *RAND Journal of Economics*, pp. 16–38.

Henderson, R. and Cockburn, I. (1994), "Measuring competence? Exploring firm effects in pharmaceutical research," *Strategic Management Journal*, 15, 63–84.

Henderson, R. M. and Clark, K. B. (1990), "Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms," *Administrative science quarterly*, pp. 9–30.

Hicks, D. (1995), "Published papers, tacit competencies and corporate management of the public/private character of knowledge," *Industrial and Corporate Change*, 4, 401–424.

Hochberg, Y. V., Serrano, C. J., and Ziedonis, R. H. (2018), "Patent collateral, investor commitment, and the market for venture lending," *Journal of Financial Economics*, 130, 74–94.

Holbrook, D., Cohen, W. M., Hounshell, D. A., and Klepper, S. (2000), "The nature, sources, and consequences of firm differences in the early history of the semiconductor industry," *Strategic Management Journal*, 21, 1017–1041.

Hounshell, D. A. and Smith, J. K. (1988), *Science and Corporate Strategy: Du Pont R and D, 1902-1980*, Cambridge University Press.

Iansiti, M. (1997), *Technology integration: making critical choices in a dynamic world*, Harvard Business school press.

Jaffe, A. B. (1986), "Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value," *NBER Working Papers: No. 1815.*

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

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993), "Geographic localization of knowledge spillovers as evidenced by patent citations," *the Quarterly journal of Economics*, 108, 577–598.

Katila, R. and Ahuja, G. (2002), "Something old, something new: A longitudinal study of search behavior and new product introduction," *Academy of management journal*, 45, 1183–1194.

Kay, N. (1988), "The R&D function: corporate strategy and structure," *Technical change and economic theory*, pp. 283–294.

Klasa, S. J., Hernán Ortiz-Molina, M. S., and Srinivasan, S. (2018), "Protection of trade secrets and capital structure decisions," *Journal of Financial Economics*, 128, 266–286.

Kline, S. J. and Rosenberg, N. (1986), "An overview of innovation," in *The positive sum strategy: Harnessing technology for economic growth*, p. 640.

Laverty, K. J. (1996), "Economic "short-termism": The debate, the unresolved issues, and the implications for management practice and research," *Academy of Management Review*, 21, 825–860.

Lavie, D., Stettner, U., and Tushman, M. L. (2010), "Exploration and exploitation within and across organizations," *Academy of Management annals*, 4, 109–155.

Lawrence, P. R. and Lorsch, J. W. (1967), "Differentiation and integration in complex organizations," *Administrative science quarterly*, pp. 1–47.

Li, D., Azoulay, P., and Sampat, B. N. (2017), "The applied value of public investments in biomedical research," *Science*, 356, 78–81.

Li, G.-C., Lai, R., D'Amour, A., Doolin, D. M., Sun, Y., Torvik, V. I., Amy, Z. Y., and Fleming, L. (2014), "Disambiguation and co-authorship networks of the US patent inventor database (1975–2010)," *Research Policy*, 43, 941–955.

Lieberman, M. B. (1978), "A literature citation study of science-technology coupling in electronics," *Proceedings of the IEEE*, 66, 5–13.

Lucking, B., Bloom, N., and Reenen, J. V. (2018), "Have R&D Spillovers Changed?" NBER Working Papers: No. 24622, National Bureau of Economic Research, Cambridge, MA. Data available at: `https://stanford.app.box.com/s/iexvj0qjjnuxlnr5ot9x3brmhr0oa319` (dataset dated February 2018).

Mansfield, E. (1980), “Basic research and productivity increase in manufacturing,” *The American Economic Review*, 70, 863–873.

March, J. G. (1991), “Exploration and Exploitation in Organizational Learning,” *Organization Science*, 2, 71–87.

Marginson, D. and McAulay, L. (2008), “Exploring the debate on short-termism: A theoretical and empirical analysis,” *Strategic Management Journal*, 29, 273–292.

Marquis, D. G. and Allen, T. J. (1966), “Communication patterns in applied technology.” *American Psychologist*, 21, 1052.

Marx, M. and Fuegi, A. (2020a), “Reliance on Science by Inventors: Hybrid Extraction of In-text Patent-to-Article Citations,” Tech. rep., National Bureau of Economic Research.

Marx, M. and Fuegi, A. (2020b), “Reliance on science: Worldwide front-page patent citations to scientific articles,” *Strategic Management Journal*, 41, 1572–1594.

Marx, M., Strumsky, D., and Fleming, L. (2009), “Mobility, skills, and the Michigan non-compete experiment,” *Management Science*, 55, 875–889.

Merton, R. K. (1973), *The sociology of science: Theoretical and empirical investigations*, University of Chicago press.

Moser, P. (2005), “How do patent laws influence innovation? Evidence from nineteenth-century world’s fairs,” *American Economic Review*, 95, 1214–1236.

Motohashi, K. (2020), “Science and Technology Co-evolution in AI: Empirical Understanding through a Linked Dataset of Scientific Articles and Patents,” *Science and Technology.*

Mowery, D. C. (2009), “Plus ça change: Industrial R&D in the “third industrial revolution”,” *Industrial and Corporate Change*, 18, 1–50.

Murray, F. and Stern, S. (2006), “When ideas are not free: The impact of patents on scientific research,” *Innovation policy and the economy*, 7, 33–69.

Nagle, F. and Teodoridis, F. (2020), “Jack of all trades and master of knowledge: The role of diversification in new distant knowledge integration,” *Strategic Management Journal*, 41, 55–85.

Narin, F., Hamilton, K. S., and Olivastro, D. (1997), “The increasing linkage between US technology and public science,” *Research policy*, 26, 317–330.

National Science Foundation, National Center for Science and Engineering Statistics (2019), "National Patterns of RD Resources: 2017–18 Data Update. NSF 20-307," Alexandria, VA. Available at `https://ncses.nsf.gov/pubs/nsf20307` (Accessed: July 20, 2020).

Nelson, R. R. (1959), "The simple economics of basic scientific research," *Journal of Political Economy*, 67, 297–306.

Nelson, R. R. (1982), "The role of knowledge in R&D efficiency," *The Quarterly Journal of Economics*, 97, 453–470.

Odlyzko, A. (1995), "The decline of unfettered research," *Unpublished, University of Minnesota, http://www. dtc. umn. edu/˜ odlyzko/doc/decline. txt.*

O'Reilly III, C. A. and Tushman, M. L. (2013), "Organizational ambidexterity: Past, present, and future," *Academy of management Perspectives*, 27, 324–338.

Pisano, G. P. (2010), "The evolution of science-based business: innovating how we innovate," *Industrial and corporate change*, 19, 465–482.

Poege, F., Harhoff, D., Gaessler, F., and Baruffaldi, S. (2019), "Science quality and the value of inventions," *Science advances*, 5, eaay7323.

Ritti, R. R. (1971), *The engineer in the industrial corporation*, Columbia University Press New York.

Roach, M. and Cohen, W. M. (2013), "Lens or prism? Patent citations as a measure of knowledge flows from public research," *Management Science*, 59, 504–525.

Robbins, C. A. (2009), "Measuring payments for the supply and use of intellectual property," in *International trade in services and intangibles in the era of globalization*, pp. 139–171, University of Chicago Press.

Rosenberg, N. (1990), "Why do firms do basic research (with their own money)?" *Research policy*, 19, 165–174.

Rosenberg, N. and Nathan, R. (1994), *Exploring the black box: Technology, economics, and history*, Cambridge University Press.

Sauermann, H. and Cohen, W. M. (2010), "What makes them tick? Employee motives and firm innovation," *Management Science*, 56, 2134–2153.

Sauermann, H. and Roach, M. (2012), "Taste for science, taste for commercialization, and hybrid scientists," *Organisation, Economics and Policy of Scientific Research, Torino, Italy.*

Securities Data Company Platinum (2018), "Mergers & Acquisitions module, 1980-2015," Refinitiv, London, GB. Provided via a Duke University subscription service.

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

Simeth, M. and Cincera, M. (2016), "Corporate science, innovation, and firm value," *Management Science*, 62, 1970–1981.

Singh, J. (2005), "Collaborative networks as determinants of knowledge diffusion patterns," *Management science*, 51, 756–770.

Singh, J. and Agrawal, A. (2011), "Recruiting for ideas: How firms exploit the prior inventions of new hires," *Management Science*, 57, 129–150.

Singh, J. and Fleming, L. (2010), "Lone inventors as sources of breakthroughs: Myth or reality?" *Management science*, 56, 41–56.

Smith, A. (1776), "The wealth of nations," *New York: The Modern Library.*

Spector, A., Norvig, P., and Petrov, S. (2012), "Google's hybrid approach to research," *Communications of the ACM*, 55, 34–37.

Standard & Poor's (2018a), "Compustat Segments dataset, 1980-2015," Available to Duke University through Wharton Research Data Services (WRDS). `https://wrds-web.wharton.upenn.edu/wrds/ds/comp/seghistd/index.cfm?navId=87` (Accessed: August 2018).

Standard & Poor's (2018b), "North America Annual Compustat, 1980-2015," Available to Duke University through Wharton Research Data Services (WRDS). `https://wrds-web.wharton.upenn.edu/wrds/ds/comp/funda/index.cfm?navId=80` (Accessed August 2018).

Stern, S. (2004), "Do scientists pay to be scientists?" *Management Science*, 50, 835–853.

Stettner, U. and Lavie, D. (2014), "Ambidexterity under scrutiny: Exploration and exploitation via internal organization, alliances, and acquisitions," *Strategic management journal*, 35, 1903–1929.

Stokes, D. (1997), "Pasteur's Quadrant: Basic Science and Technological Innovation (The Brookings Institution, Washington DC)," .

Thursby, J. G. and Thursby, M. C. (2011), "Has the Bayh-Dole act compromised basic research?" *Research Policy*, 40, 1077–1083.

Tushman, M. L. (1977), "Special boundary roles in the innovation process," *Administrative science quarterly*, pp. 587–605.

Tushman, M. L. and Katz, R. (1980), "External communication and project performance: An investigation into the role of gatekeepers," *Management science*, 26, 1071–1085.

Tushman, M. L. and O'Reilly III, C. A. (1996), "Ambidextrous organizations: Managing evolutionary and revolutionary change," *California management review*, 38, 8–29.

Tzabbar, D., Silverman, B. S., and Aharonson, B. S. (2015), "Learning by hiring or hiring to avoid learning?" *Journal of Managerial Psychology*, 30, 550.

Vincenti, W. G. (1990), *What Engineers Know and How They Know It: Analytical Studies from Aeronautical History (Johns Hopkins Studies in the History of Technology)*, The Johns Hopkins University Press, Baltimore, MD.

Williams, H. L. (2013), "Intellectual property rights and innovation: Evidence from the human genome," *Journal of Political Economy*, 121, 1–27.

Wise, G. (1985), *Willis R. Whitney, General Electric and the Origins of US Industrial Research*, Columbia University Press.

Wu, Y. (2010), "What's in a name? What leads a firm to change its name and what the new name foreshadows," *Journal of Banking & Finance*, 34, 1344–1359.