BACK TO BASICS: WHY DO FIRMS INVEST IN RESEARCH?

Ashish Arora
Sharon Belenzon
Lia Sheer

Back to Basics: Why do Firms Invest in Research?
Ashish Arora, Sharon Belenzon, and Lia Sheer
NBER Working Paper No. 23187
February 2017, Revised November 2017
JEL No. O31,O32

ABSTRACT

If scientific knowledge is a public good, why do firms invest in research? This paper revisits this classic question with new data on patent citations to scientific publications by corporations. Using data on 4,736 firms for the period 1980-2006, we document that corporate investment in research is closely related to its use in internal invention. Specifically, firms that build on their scientific research in their inventive activity invest more in research than those that are less successful in using their research internally. Consistent with this, research that is internally used is valued more and is more productive. Our results are consistent with the view that to justify further investment in research corporate scientists need to demonstrate that their recent scientific work is useful for the core inventive activity of the sponsoring firm.

Ashish Arora
Fuqua School of Business
Duke University
Box 90120
Durham, NC 27708-0120
and NBER
ashish.arora@duke.edu

Sharon Belenzon
Fuqua School of Business
Duke University
100 Fuqua Drive
Durham, NC 27708
and NBER
sharon.belenzon@duke.edu

Lia Sheer
Fuqua School of Business
Duke University
100 Fuqua Drive
Durham NC 27708
lia.sheer@duke.edu

Back to Basics: Why do Firms Invest in Research?

Ashish Arora[*]   Sharon Belenzon[†]   Lia Sheer[‡]

November 6, 2017

Abstract

If scientific knowledge is a public good, why do firms invest in research? This paper revisits this classic question with new data on patent citations to scientific publications by corporations. Using data on 4,736 firms for the period 1980-2006, we document that corporate investment in research is closely related to its use in internal invention. Specifically, firms that build on their scientific research in their inventive activity invest more in research than those that are less successful in using their research internally. Consistent with this, research that is internally used is valued more and is more productive. Our results are consistent with the view that to justify further investment in research corporate scientists need to demonstrate that their recent scientific work is useful for the core inventive activity of the sponsoring firm.

Keywords: innovation, scientific research, development, role of science in corporate R&D
JEL Classification: O31, O32, O16.

1   Introduction

Corporate investment in scientific knowledge has always been a puzzle. It is substantial in magnitude and the available evidence suggests it is privately profitable, and yet, we do not fully understand how firms derive value from producing what is essentially a public good. In 2015, the business sector performed nearly 26% of all basic research in the United States and funded a similar share (NSF 2017)[1]. In absolute terms, the business sector invested over $22 billion in basic research, a substantial amount. While evidence on the profitability of such investments is scarce, studies based on data from the 1960s and 1970s suggest that private returns are substantial. In particular, Zvi Griliches (1986), based on a sample of 883 large

---

[*]Duke University, Fuqua School of Business and NBER (ashish.arora@duke.edu)

[†]Duke University, Fuqua School of Business and NBER (sharon.belenzon@duke.edu)

[‡]Duke University, Fuqua School of Business (lia.sheer@duke.edu)

[1]National Science Foundation, National Center for Science and Engineering Statistics. 2017. National Patterns of R&D Resources: 2014–15 Data Update. NSF 17-311. Arlington, VA. https://www.nsf.gov/statistics/2017/nsf17311. According to the same source, business performed over 40% of basic and applied research in the United States, and funded a similar share.

manufacturing firms in the United States, concluded that investment in basic research was associated with higher productivity and profits, implying a significant private returns to basic research. Yet, the mechanism by which these private returns accrue remains unclear once the knowledge produced is made publicly available in the form of patents and scientific publications.

Over the last quarter century, a variety of explanations have been put forward. These explanations focused on absorptive capacity to use external knowledge (Cohen and Levinthal, 1989; Rosenberg, 1990; Gambardella, 1992), enhancing reputation to attract investors and costumers (Hicks, 1995) and incentives for high-skilled scientist-inventors (Audretsch and Stephan, 1996; Stern, 2004; Gambardella et al., 2015). The various explanations are not mutually exclusive, but can have very different normative and positive implications (section 2 discusses these in more detail).

In this paper, we explore the empirical basis for a different explanation, namely that *"... most of the actual research in industry is devoted to the development of new products or processes"* (Griliches, 1986:145)[2]. We use newly developed data linking patents to scientific publications matched to firms to investigate the extent to which firms invest in research as an input into their own inventions. We measure corporate research by scientific papers authored by researchers employed by firms. We measure the use of research as inputs into inventive activity by citations in the firm's patents to its own scientific publications. Spillovers are measured by citations in the patents of other firms to the focal firm's scientific publications. We find that firms produce more scientific knowledge when they are able to use it in their own inventions, but spillovers to product market rivals are associated with lower scientific production. Our estimates indicate that the positive effect of one internal citation on the number of future publications produced by the focal firm is equivalent to the combined negative effect of four external citations. Thus, while spillovers might cause firms to underinvest in research, firms would still invest in research if they are able to use it internally.

Our principal contribution is to document that corporate production of scientific knowledge is closely related to its use in internal invention. We examine both the determinant of progress of science and

---

[2]The modern innovation literature has tended to conflate research and development. The older innovation literature was much more careful in distinguishing between the two. Nelson (1959) was among the first to examine incentives to invest in research as opposed to development. Using publications data, Adams (1990) finds that scientific knowledge absorbed by firms is associated with higher productivity growth. Simon Kuznets, in his 1971 Nobel Prize Lecture concluded that "*Mass application of technological innovations, which constitutes much of the distinctive substance of modern economic growth, is closely connected with the further progress of science, in its turn the basis for additional advance in technology*".

its application in development at the firm level. Our main methodological contribution is to match publication records from Web of Science (WoS) to front-page non-patent literature (NPL) references on a large scale. While previous research using patent citation data was mostly done for selected industries and years (e.g., Narin and Noma, 1985; Narin et al., 1997; McMillan et al., 2000; Hicks et al., 2001; Breschi and Catalini, 2010; Bikard, 2015; Popp, 2016), our research examines a broad range of companies across many industries over a quarter of a century. Our primary firm sample consists of 4,736 U.S. headquartered publicly listed, R&D-performing companies over the period 1980-2006. Collectively, these firms account for approximately 300 thousand corporate scientific publications over the sample period, of which 50,494 publications are cited at least once by patents granted up to 2014.

We present three main findings. First, we validate that our measure of patent citation to science corresponds to the use of science in invention. Using data from the Carnegie Mellon Survey of R&D performing firms (Cohen et al., 2000), we show that firms whose patents cite science are also those that report using scientific output in their R&D projects. This relationship continues to hold at a finer measurement scale such as relating citations to use in specific technology areas.

Second, we demonstrate a strong positive relationship between corporate investment in research and internal use of past research in invention. Firms invest more in research when the scientific knowledge they produce is cited in their patents. This relationship is stronger for new, basic, high quality cited research and for citing patents in core technology areas of the inventing firm. We supplement these findings by showing that internal use is associated with a higher R&D productivity and a higher stock market valuation of scientific publications stock.

Third, consistent with the view that research is an input into internal inventive activity, we find that firms produce fewer publications when their research spills over to close product market rivals and that these spillovers are negatively related to the private value of research. These patterns are inconsistent with the view that firms invest in research principally to absorb external knowledge. This is also inconsistent with firm's investing in research principally to attract talented workers or to signal to customers or regulators.

The firm behavior we analyze, namely the production and use of scientific knowledge by profit seeking companies, is complex. We are well aware that both the production and use of research are potentially

affected by common variables, some of which are unobserved and are likely to vary across firms and industries. We probe the robustness of our results by using firm-fixed effects and by directly measuring organizational features that should help firms use internally the science they produce. We present instrumental variable estimates that are based on variation across states over time in the application of the Inevitable Disclosure Doctrine. This doctrine restricts scientists and inventors' mobility between companies, leading to a higher use of research in invention through a tighter link between research and development personnel.

The rest of the paper is organized as following. Section 2 discusses related literature, Section 3 presents the data and empirical methodology, Section 4 presents the estimation results and Section 5 concludes.

2   Related literature

American industrial research activity typically takes place inside corporate research labs. These labs were initially established with only modest goals. In the late 19th century, firms in technology intensive sectors such as railroads, steel, and telegraphy relied largely on external inventions. These firms established industrial labs to evaluate the quality of inputs, such as the quality of steel for rails (Mowery, 1995; Carlson, 2013). Thereafter, in the early 20th century, firms such as AT&T, GE and DuPont invested in internal research to solve production problems and evaluate and adapt inventions acquired from other firms (Reich, 1985; Hounshell and Smith, 1988). Corporate investment in research became more significant during the inter-war years, as corporations grew larger and more anxious to manage innovation instead of having to rely on external inventions (Maclaurin, 1953). Stronger anti-trust enforcement provided an additional impetus as some farsighted managers saw in research a source of new products to fuel growth without running afoul of the anti-trust authorities.

The importance of discoveries such as vacuum tubes, radar, radio, synthetic rubber, nuclear fission, and penicillin in the conduct of World War II led to a deeper appreciation of the potential economic usefulness of research. The simplest view of the role of research in innovation was the so-called "linear model" associated with Bush (1945), who asserted that technical progress rests upon scientific advance; that inventions grew out of research[3]. This view was modified and enriched in a variety of ways (Kline

---

[3]In Vannevar Bush's own words: "*Basic research leads to new knowledge. It provides scientific capital. It creates the fund*

and Rosenberg, 1986; David, Mowery and Steinmueller, 1992). However, the underlying notion that "...*most of the actual research in industry is devoted to the development of new products or processes...*" (Griliches, 1986: 145) remained in place. In a seminal study, Griliches estimated the private return to research using the National Science Foundation (NSF) R&D-Census match, containing information on R&D expenditures, sales and employment for approximately 1000 largest manufacturing firms from 1957 through 1977. He estimated a Cobb-Douglas production function, including basic research as a fraction of total R&D in addition to R&D stock, labor and capital and found a very large return to basic research. Firms that spent a larger share of R&D on basic research were substantially more productive.[4]

While Griliches's work demonstrated the presence of significant private returns to research, the mechanism by which these private returns accrue remained unclear. Since the results of basic research were typically published and shared (e.g., Dasgupta and David, 1994), this raised the question of how firms were benefiting from their investment in research and why should they invest themselves rather than "free-ride" on the research of others (Arrow, 1962; Nelson 1959). Contributing to this puzzle was the absence of a serious attempt to explore empirically the extent to which firms invest in science to spur their own downstream inventions. This puzzle led to new explanations for why firms invest in research including absorptive capacity (Cohen and Levinthal, 1989; Rosenberg, 1990; Gambardella, 1992), incentives for high-skilled scientist-inventors (Audretsch and Stephan, 1996; Stern, 2004; Gambardella et al., 2015) and enhancing reputation to attract investors and costumers (Hicks, 1995). The various explanations are not mutually exclusive, but can have very different normative and positive implications.

Cohen and Levinthal (1989) challenged the public good nature of research, arguing that accessing outside knowledge is costly and requires absorptive capacity, which in turn requires that firms engage in R&D. Rosenberg (1990) also challenged the idea that existing knowledge, though in the public domain, was "on the shelf", available to all. Instead, he argued that finding, evaluating, and using publicly available knowledge itself presupposed some prior knowledge. He claimed therefore that firms invest in research because, in part, basic research helps the company stay up-to-date and identify scientific developments

---

from which the practical applications of knowledge must be drawn. 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." (p. 241).

[4]Mansfield (1980) found that investment in basic research was related to productivity growth in U.S. manufacturing industries in the period 1948-66, controlling for applied research and development.

in its field as well as more easily absorb external knowledge while fitting it to its own needs (Rosenberg 1990). A vast literature has found evidence consistent with absorptive capacity. Using survey data, Levin et al. (1987) find that independent R&D is most effective for learning about rivals' technology. Gambardella (1992) shows that pharmaceutical firms with better research capabilities, measured by number of publications, are able to exploit internal as well as external science more effectively.

Using data from the pharmaceuticals industry, Cockburn and Henderson (1998) argue that to take advantage of public research firms must develop internal basic research programs as a platform of interacting with public sector researchers. Several studies examined the role of corporate publications in attracting talented scientist-inventors. Hicks (1995) argues that companies invest is research because publications are an effective tool to recruit scientists[5]. Henderson and Cockburn (1994) emphasize the importance of publications as a reward system. Examining research programs of major pharmaceutical firms they find that scientists that are promoted on the basis of their publications and reputation in the wider scientific community generate more important patents. Some researchers may have a "taste for science" i.e., may be willing to accept industrial positions if allowed to spend some time on their own research and to publish it. Stern (2004) finds that scientists may be willing to accept 20% lower wages in exchange for autonomy, such as time for conducting and publishing independent research. Using survey data on PhD industrial scientists, Sauermann and Cohen (2010) study the relationship between industrial scientists' motives and their innovation activities. They find that intellectual challenge and independence have the strongest (positive) relationship with innovation output, especially in upstream research activities. Gambardella et al. (2015) argue that DuPont's invention of Nylon is the result of offering the young Harvard scientist Wallace Carothers the opportunity to publish his independent research. Gans et al. (2013) develop a theoretical model in which firms allow researchers to publish in return to lower wages as long as patents can be used to prevent the disclosed knowledge from benefiting rivals. It is worth noting that while demonstrating that performing research and producing scientific articles play an important motivating role for attracting scientists, it remains unclear why firms wish to attract scientists in the first place. Moreover, if performing research is merely a reward instrument, whether research is used in invention

---

[5]Cockburn and Henderson (1998) conclude that participation in research "*acts as a powerful recruiting tool, since the highest quality scientists in a field are often reluctant to work for private firms if they will not be able to publish and thus maintain their personal scientific reputation*".

should not influence investment decisions.

Basic research can also benefit the sponsoring firm by allowing it to signal its scientific and technical capabilities to regulators, prospective customers, employees, or financiers (Hicks, 1995; Audretsch and Stephan, 1996; Azoulay, 2002). Lichtenberg (1986) shows that through investment in research, firms can signal their capabilities to attract government contracts. Based on Compustat firm data and defense-related federal procurement data he finds that approximately half of the increase in private R&D investment between 1979 and 1984 was stimulated by increase in government demand. Audretsch and Stephan (1996) suggest that collaborative research with university scientists helps biotech firms signal their quality to investors. Azoulay (2002) finds that prescriptions for anti-ulcer drugs are influenced by publications activity by drug manufacturers. Hicks (1995) suggests that scientific publications serve as a signal of the firm's tacit knowledge and capabilities and thus enhance the technical reputation of the firm. She emphasizes that by publishing in open science firms establish a trustworthy reputation with the academic and scientific community, which helps them in turn trade information with these communities.

The various explanations for why firms invest in research predict that invention productivity would be higher in firms that invest in research. For instance, insofar as the firm invests in research to attract talented inventors, one would also expect that talented inventors publish scientific articles, and for invention productivity to be higher in such firms. The absorptive capacity view has similar predictions if one makes the auxiliary assumption that inventors that are active in research are better able to absorb external knowledge, and hence, are more productive. What is distinctive is whether the coupling between research and invention is based on the use of internally generated knowledge in the sponsoring firm's own inventions.

There are two related implications of this distinction that can be empirically tested. If corporate research is principally about building reputation capital with regulators, customers, or others, it should be more effective in doing so when others build upon it and cite it. In other words, spillovers should enhance the efficacy of the signal and internal use would merely be a welcome bonus. If, however, corporate research is primarily about producing more and higher quality inventions internally, spillovers, particularly to rivals, should lower private return and reduce the incentives to invest in research. In such case, the firm's incentives to invest are directly tied to whether its own inventors build upon the internally

generated research. To test this, in our empirical analysis we examine how a company's publication activity is related to the use of internally generated knowledge in its own inventions as well as to the use by others including its rivals in the product market.

3   Data

We combine data from three main sources: (i) company and accounting information from U.S. Compustat, (ii) scientific publications from Web of Science and (iii) patent and non-patent literature (NPL) citation from PatStat database. Building on Arora et al. (2015), we develop new data linking corporate publications to NPL citations to learn about the use of corporate science in invention and its implications for corporate investment in research, R&D productivity and stock market value.

We start with all publicly traded firms in the U.S. annual Compustat database and select companies with active records and positive R&D expenses for at least one year during our sample period, 1980-2006. We exclude companies without at least one patent based on the NBER 2006 patent dataset. We also exclude firms that are not headquartered in the United States. Our final estimation sample consists of an unbalanced panel of 4,736 firms and 57,765 firm-year observations. Of those firms, 2,413 have at least one scientific publication during the sample period.

We use scientific publications as our measure of production of new 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 or built upon the knowledge in the publication. Internal citation is a citation by a patent to a scientific publication produced by the same firm.

Corporate publications. Similar to the method discussed in Arora et al. (2015), to measure firms' participation in scientific research we match our sample firms to the Web of Science (WoS) database, which covers articles published in over 5000 journals. 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, we identify approximately 300 thousand articles with at least one author employed by our sample of Compustat firms, published from 1980 through 2006. Appendix 6.2 provides details on the matching procedure.

Patent citation to corporate science. The main methodological contribution of this paper is

matching NPL citations to WoS publications. Using all patents granted in the period 1980-2014, we perform a many-to-many match between NPL citations and WoS corporate publications (approximately 14 million citations matched to 300 thousand corporate publications), allowing for more than one publication to be matched to each citation. To exclude mismatches, we develop a specialized matching algorithm that is based on different sources of publication information: standardized authors' names, number of authors listed, 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.[6] Finally, we perform extensive manual checks to confirm matches. Details on the matching algorithm are provided in Appendix 6.4.

Following the above procedures, we obtain 266,361 patent citations to 50,494 corporate publications (17% of corporate publications), by 151,412 citing patents. Of the cited publications, 79% receive only external citations and the remaining receive at least one internal citation.

Ownership structure. Due to the complexity of measuring large firms' innovative activities, which typically take place inside numerous subsidiaries (Arora et al., 2014), we aggregate the data to the ultimate-owner-parent-company level (UO). For example, if a firm's subsidiary publishes scientific articles while the parent company is the assignee registered on the firm's patents, we record both at the UO level and a citation from a patent to a publication would be considered as an internal citation.

The construction of the firm dataset presents several challenges. For instance, a parent company and a subsidiary may have different identification numbers and records in Compustat. Furthermore, a single company may correspond to multiple firm identifiers due to changes in ownership structure and accounting changes over the sample period. We detail the challenges of constructing the dataset and the procedures we take to deal with them in Appendix 6.1.

3.1 Descriptive statistics

Our main sample and variables are at the parent company-year level. Appendix Table A1 summarizes the definition and data source for each variable. Table 1 presents descriptive statistics for our main variables over the sample period, 1980-2006. Our sample includes a wide distribution of firm sizes: market value

---

[6]An example of a front-page patent reference to a non-patent literature is presented in Appendix Figure A5.

9

ranging from 5 million dollars (10th percentile) to 2.3 billion dollars (90th percentile) and sales ranging from 2 million dollars (10th percentile) to 2 billion dollars (90th percentile).

Table 2 presents summary statistics for the main citation variables used in the econometric analysis for publishing firms. A total of 2,413 firms (51 percent of our sample firms) publish at least one scientific article. 799 firms receive at least one citation to their publications (an average of 1 internal and 7 external citations per year), 388 firms make at least one citation to their own scientific publications (an average of 1 unique firm publication cited per year) and 760 firms receive at least one external citation to their publication (an average of 6 unique patents citing a firm's publications per year).

Table 3 presents mean comparison tests for differences in characteristics between firms with high and low internal citations for the sample of publishing firms with at least one citation. Firms with above mean value share of internal citation (defined as the ratio of internal citations received and the sum of internal and external citations received) have statistically significant higher (i) number of publications, (ii) R&D intensity and (iii) inventor-author overlap, where higher overlap indicates a stronger tie between research and development personnel, measured as the share of patents that list at least one scientist (publication author) as an inventor.[7][8]

Insert Tables 1-3 here

3.2 Validating patent citations to scientific articles as a measure of use of science in invention

To validate our measure of use of science–NPL citation to WoS articles– we utilize the Carnegie Mellon Survey (CMS) data on industrial R&D (Cohen et al., 2000).[9] Our sample includes 772 patenting firms that were included in the survey with patents granted between 1991 and 1999 (a total of 29,318 patents). Figures 1A-1C present the relationship between patent citations to science and the survey answers pertaining to the role of science in corporate R&D. There is a strong correlation between the average number of patent citations to science made by the surveyed firms' patents and the responses of the same firms to

---

[7]This measure is further discussed in section 4.2.1 as part of our instrumental variable analysis.

[8]Appendix Table A3 presents a mean comparison test between patents that cite internal science and patents that do not cite internal science. We find that patents that cite internal science are of statistically significant higher quality (based on forward patent citation received) and are in the core technology of the firm (core defined as the four-digit IPC that has the majority of firm patents in a given year).

[9]We thank Michael Roach and Wesley Cohen for providing the Carnegie Mellon survey data.

questions on the importance of scientific research for their R&D projects.

Figures 1A and 1B present mean comparisons for the average number of patent citations to publications for firms with high and low survey response for the use of public science in R&D projects.[10] Classification to high and low is based on median value survey response. The figures show that firms with high self-reported use of science also cite more public science (publications by all universities and research institutions, Figure 1A) and articles publications by top 200 American universities (Figure 1B).[11][12] Figure 1C shows that the relationship between self-reported use of science and citations to science holds also within narrowly defined technology areas.[13] The figure shows that there is a tight correspondence between the specific scientific areas the firm reports to have influenced its R&D projects and the research areas cited by its patents.

Table 4 confirms that the above correlations continue to hold in a regression analysis that controls for firm size, number of backward patent citations to other patents (to ensure that the effect of NPL citations is not driven by how many backward citations the patent makes), and complete sets of four-digit industry and year dummies.

Insert Figure 1 and Table 4 here

4 Econometric analysis

4.1 Internal citation and publication output

We estimate the relationship between internal use of research, measured by patent citations to internally produced science, and investment in research, measured by numbers of publications authored by at least one corporate scientist. Our baseline specification is as follows:

---

[10] Based on 1994 CMS data Q.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?: a. Research findings".

[11] Top university publications were identified by matching a list of top 200 U.S. university names based on Shanghai Ranking to the affiliation field of each publications record.

[12] Appendix Figure A1 shows that the same pattern holds for main industries. Appendix Figures A2-A4 present additional supporting evidence. Figure A2 shows that citations to science are positively related to share of Ph.Ds or M.D. scientists of all R&D employees as reported in the survey.

[13] Based on 1994 CMS data Q.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. Then, indicate the importance of that field's findings to your R&D activities". The sample is restricted to firms that indicated their main field in Q22 as A-J (excluding category K -'others'). Publications were classified to main fields based on key related words under the WoS journal subject category field. For example, "Organometallics" Journal related subject category is "Chemistry, Inorganic & Nuclear; Chemistry, Organic" accordingly it is classified under main field of chemistry.

$$\ln(1 + Publications_{it}) = \beta_0 + \beta_1 \ln(1 + Internal\ citation_{it-1}) + \mathbf{Z}'_{it-1}\gamma + \eta_i + \tau_t + \epsilon_{it} \qquad (1)$$

$Publications_{it}$ is number of publications by firm i in year t. $Internal\_citation_{it-1}$ is lagged number of patent citations made by firm i's patents granted up to year t-1 (inclusive) to its own scientific publications published up to year t-1 (inclusive). $\mathbf{Z}_{it-1}$ is a vector of lagged firm-year controls, including patent stock, R&D stock, and sales[14]. η_i and τ_t are complete sets of firm and year dummies, respectively. ϵ_{it} is and iid error term. Our coefficient of interest is β_1 and we expect $\widehat{\beta}_1 > 0$. All specifications include a dummy variable for firm-year observations with zero publications. Table 5 presents the estimation results.

One main concern is that firms with a higher number of publications are more likely to randomly cite one of their publications, which would upward bias $\widehat{\beta}_1$. To mitigate this concern, all of our specifications include firm fixed effects as well as firm controls for scale such as patent stock, R&D stock and sales. Furthermore, our choice of the temporal structure of internal citations aims at mitigating concerns that number of publications and internal citation are affected by common shocks (e.g., shocks to research opportunity that affect both number of publications and number of patents).[15]

Column 1 presents the estimation results from a pooled specification with four-digit industry fixed effects. There is a positive and statistically significant relationship between internal citation and number of publications. Column 2 presents the same pattern of results for a between-firm specification, which collapses the panel data into a cross-section by averaging variables at the firm level. Column 3 adds firm fixed-effects. $\widehat{\beta}_1$ falls sharply from 0.68 (column 1) to 0.11, indicating that the relationship between internal citation and publications is driven largely by heterogeneity across firms rather than within firms over time. Yet, $\widehat{\beta}_1$ remains statistically significant. Based on the estimates from column 3 (firm fixed effect), one additional internal citation is associated with an additional 0.5 publication per firm-year $(0.11 \times 6.216/1.274)$.

Columns 4-8 present several robustness checks. Column 4 restricts the sample to firms with at least one publication during the sample period. There is no substantial change in $\widehat{\beta}_1$. Column 5 controls

---

[14] R&D stock is calculated using a perpetual inventory method with a 15 percent depreciation rate (Hall et al., 2005). R&D stock, GRD, in year t is $GRD_t = R_t + (1 - \delta)GRD_{t-1}$ where R_t is the R&D expenditure in year t and $\delta = 0.15$. Patent stock in year t is $Patent\ stock_t = Pat_t + (1 - \delta)Patent\ stock_{t-1}$ where Pat_t is the number of patents in year t.

[15] The temporal structure of citations and publications are illustrated in Appendix Figure A6.

for internal patent citations to own patents ("self-citations"). While $\widehat{\beta}_1$ remains stable, the coefficient estimate on self citations is statistically zero. This result is reassuring because it mitigates a concern than that $\widehat{\beta}_1$ captures a "self-citation" effect that might be driven by cumulative innovation capabilities (Hall et al., 2005; Belenzon, 2012).

Columns 6-8 present the estimation results using alternative measures of internal citation, all yielding the expected positive relationship with publications.[16][17][18]

<div align="center">**Insert Table 5 here**</div>

4.1.1 Heterogeneous effects

Not all citations to science are of equal importance. We expect internal citations to affect future investment in research when the cited publication (i) is of high scientific impact, (ii) is based on recent work that is less known by others and (iii) is related to the firm's core technologies and to its valuable inventions. These predictions are confirmed in our data and are reported in Table 6.

Column 1 distinguishes between citations to old *vs.* new science. Internal citations to new science include only citations to articles published no later than five years from the grant year of the citing patent. While the coefficient estimate on internal citation to recent science is positive and statistically significant (0.130), internal citations to old science have no effect.

Columns 2-3 distinguish between citations to basic and applied publications using journal CHI index[19] (Column 2) and Journal Impact Factor (JIF) (Column 3). In Column 2, we define basic (applied) journals as the top (bottom) two categories of the CHI index and classify publications accordingly. In Column 3, we classify journals as basic and applied based on their JIF value. A publication is classified as basic if it

---

[16] In unreported robustness checks, we ran the analysis excluding references to articles related to clinical trial phase in the pharmaceutical and biotech industry, which are not considered as research. We examine all publications cited by pharmaceutical and biotech patents and identify clinical trial publications by related phrases in the title and abstract of each publication record (e.g., clinical trial, clinical study, preclinical trial, subjects). We locate less than 100 internally cited clinical trial publications and exclude them from the analysis. Our results remain robust.

[17] As additional unreported robustness checks, we performed the same analysis excluding citations added by patent examiners. We also excluded citations where patent inventors cite their own publications. The results are robust in both cases.

[18] Columns 1-3 in Appendix Table A4 present our main results with right-hand-side variables lagged by two and three years. The coefficient estimate on internal citation remains positive and statistically significant.

[19] Narin et al. (1976) and CHI Research develop the CHI index to classify scientific journals into four research categories ranging from applied to basic.

is published in a journal with above median JIF value (using the JIF value distribution in the complete WoS database), and as applied otherwise. About 70% of internal citations are to articles published in basic science journals (this percentage is robust across both classifications) and these citations matter the most for the production of future publications. Based on Column 3 (JIF), the coefficient estimate on internal citation to basic science is positive and statistically significant (0.122), while the coefficient estimate on internal citation to applied science is statistically zero.

Column 4 distinguishes between high and low quality publications using number of citations an article receives from other publications, divided by average number of citations received by all WoS publications published in the same journal-year as the focal publication. Classification of articles into high and low quality is based on median value of normalized citations received in the corporate publications sample. Our results indicate that only citations to high quality publications matter. While the coefficient estimate on high quality internal citation is positive and significant (0.116), the coefficient estimate on citations to low quality publications is statistically zero.

Column 5 distinguishes between use of science by patents in core and non-core technology areas of the focal firm. Core citations include only citations to publications made by patents in the firm's core technology area. Core technology is defined as the IPC with the majority of the firm's patents in a given year.[20] There is a strong relationship between citations by core technology patents and future publications. Moving from core to non-core citing patents lowers the coefficient estimate on internal citation from 0.12 to 0.04.

Column 6 distinguishes between citations by high and low quality patents. Patent quality is based on number of citations a patent receives divided by 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 median value from the corporate patents sample. The relationship between internal citation and publications is stronger for high quality citing patents, but the coefficient estimates on high and low quality patents are not statistically different from each other (0.091 *vs.* 0.054).

In summary, Table 6 shows that internal patent citations to science that matter for the production of future science are citations to recent, high quality basic publications that are made by core and high

---

[20] Average number of patents in core technology areas is 8.6. About 40% of internal citations are by core patents.

quality patents. These results are consistent with the view that to justify further investment in research scientists are required to demonstrate that their recent scientific work is useful for the core inventive activity of the sponsoring firm.

Insert Table 6 here

4.2   Exploring the causal effect of internal citations

An important concern is that internal citations and investment in research can be driven by common unobserved or mismeasured time-varying effects, such as technological opportunity shocks or technology specialization that can be correlated with number of publications and internal citations. This section proposes an instrumental variable estimation strategy to mitigate this concern.

4.2.1   Inventor-author overlap

Our instrumental variable estimation is motivated by a potential determinant of internal citation: inventor-author overlap. Arguably, firms should be better positioned to use the science they produce in their downstream development if there is a tighter link between its research and development personnel (Kline and Rosenberg, 1986; Rosenberg, 1990).[21] This link should align research priorities with downstream needs and facilitate "back and forth" between upstream research and downstream development. In particular, a firm should be better positioned to capitalize on its research if some aspects of the research findings are tacit and more easily transmitted through face-to-face interactions between researchers and inventors.

Building on this logic, the next section explores legal constraints to labor mobility that potentially affect the overlap between research and development personnel as an instrument for internal citation. Stronger mobility barriers should raise the overlap between researcher and inventor teams and in turn lead to a higher internal citation rate. Importantly, there is no clear reason why mobility barriers should affect the number of publications a firm produces, conditional on how these publications are used in downstream development.

---

[21] According to Rosenberg (1990, p.170): *"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. The history of basic research in industry suggests that it is likely to be most effective when it is highly interactive with the work, or the concerns of applied scientists and engineers."*

We construct a measure of overlap between inventors and authors at the firm-year level. This measure is the share of patents with inventors who are also authors of a publication by the same firm published no later than three years from the focal patent's grant year. Inventor-author overlap is defined only for publishing firms. Appendix 6.6 provides additional details on the construction of this measure.[22]

Columns 1-2 in Table 7 present OLS estimation results of a linear probability model that examines the relationship between inventor-author overlap and internal citation. The dependent variable is one for firm-year observations with at least one internal citation and zero otherwise. Column 1 includes a complete set of four-digit industry dummies and Column 2 includes firm fixed effects. As expected, the probability of internal citation increases with inventor-author overlap. Based on the within-firm estimates (Column 2), a one-standard deviation increase in inventor-author overlap increases the probability of internal citation by about 30%, relative to the sample mean.[23]

Column 3 examines external citations and finds that they are unrelated to inventor-author overlap. This result mitigates a concern that the overlap-citation relationship is driven by a higher general propensity to cite as overlap increases.

4.2.2 Inevitable Disclosure Doctrine

We proceed by introducing a source of variation that should affect inventor-author overlap, but not publications. We exploit variation in the adoption of Inevitable Disclosure Doctrine (IDD) by U.S. state courts as an instrument for internal citation. IDD is a legal doctrine that restricts workers 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. IDD status at the focal firm's state in a given year is taken from Klasa et al. (2015) (see also Marx et al., 2009). Our instrument multiplies IDD by a firm-specific employment mobility risk, measured as number of rival publishing firms in close geographical proximity to the focal firm.

Our instrument, *IDD mobility,* is constructed as:

---

[22] Average value of inventor-author overlap is 0.2 with a median of 0.

[23] Under the assumption that overlap should affect citations to publications of current workers, as a robustness check we compute internal citation only for recent publications (published no later than five years from the grant year of the citing patent). The coefficient estimate on inventor-author overlap increases in size and remains statistically significant.

$$IDD\ mobility_{it} = IDD_{st} \times mobility\ risk_{it}$$

Where, IDD_{st}, is a dummy variable equals one if IDD is in effect in the focal firm's (i) state (s) in year t. Firm address is from the publication's "affiliation" field.[24] $mobility\ risk_{it}$ is the number of publishing firms in the same industry (4-digits SIC) within 100 mile of the focal firm in a given year. We calculate distance between firms using the NBER's ZIP Code Distance Database.

Our main identifying assumption is that mobility barriers do not directly affect incentives to invest in research. Thus, one has to assume that IDD adoption is uncorrelated with unobserved state-specific variables such as technological opportunities. We test this assumption by examining whether changes in IDD status are correlated with technological opportunities, measured by patents per R&D. As shown in Appendix Table A7, there is no relationship between changes in IDD and patents per R&D (same pattern holds for publications per R&D).

Column 4 in Table 7 presents OLS estimation results for the relationship between $IDD\ mobility$ and inventor-author overlap. As expected, higher mobility restrictions are associated with a higher inventor-author overlap. The estimates indicate that a two standard deviation increase in $IDD\ mobility$ is associated with an increase of 22% in inventor-author overlap (relative to the sample mean).

Columns 5-7 present within-firm Two-Stage Least Squares estimation results for the effect of internal citation on publications using lagged $IDD\_mobility$ as an instrument. To mitigate possible unobserved time-varying state heterogeneity in economic conditions that might be correlated with IDD we also control for state employment level.[25] Column 5 presents the first stage estimation results, where internal citation is regressed on lagged $IDD\ mobility$ controlling for patents stock, R&D stock, sales and state employment. The results confirm that higher mobility restrictions are associated with higher internal citation and that the instrument has strong explanatory power (Kleibergen-Paap F statistic=72).

Column 7 presents the second stage estimation results where internal citation is instrumented with $IDD\ mobility$. The IV coefficient estimate on internal citation is larger than the OLS estimate in

---

[24] Our algorithm accounts for cases where several publishing institutions are listed under the affiliation field and locates the relevant state in the string that is related to the focal firm. For example, "JOHNS HOPKINS UNIV DEPT CHEM BALTIMORE MD 21218 USA, **ARCO CHEM** CO NEWTOWN SQ **PA** 19073 USA."

[25] Annual state employment is from U.S. Bureau of Economic Analysis (BEA).

Column 6 (0.6 versus 0.1). One possible explanation for the smaller IV estimate is that the OLS estimate is downward biased due to unobserved firm heterogeneity, which is corrected in the IV estimation. For example, mismeasured or unobserved firm specialization might be positively correlated with internal citations (because both patents and publications are in similar fields they are more likely to be linked by a citation) and negatively correlated with number of publications (because they are more focused, firms perform research in a more narrow research domain). Another possible explanation for the smaller OLS estimate is noise. If the variation used in the IV estimation captures "true" use citations, $\widehat{\beta}_1$ would be larger.[26]

Finally, Appendix section 6.7 presents an analysis that exploits a different source of variation in internal citation–profit shocks due to foreign exchange rate fluctuations that affect the dependence of inventions on science (Cyert and March, 1963; Graham at al., 2004; Bruneel et al., 2016). We find that at the industry level, foreign currency devaluation is associated with drop in profits and fewer internal citations to science. These results are consistent with firms performing less exploratory, science-based, innovation in leaner times. Our instrumental variable estimation further shows a positive and statistically significant effect of internal citation on publications. Lastly, we include both instruments, IDD and devaluation, in a single two-stage least-squares specification. The same pattern of results remains in the first and second stage estimations and the Hansen test for overidentifying restrictions supports the validity of the instruments (results are presented in Appendix Table A5).

<div align="center">**Insert Table 7 here**</div>

4.3 Knowledge spillovers

If firms invest in research because it is an input into internal inventive activity, the use of this research by rivals would lower the return to such investment (Nelson, 1959; Arrow, 1962). In this section we investigate how investment in research is related to external citations–citations in patents filed by others to the research published by the focal firm.

Table 8 presents the estimation results. Column 1 includes the number of external citations made

---

[26] Appendix Table A4 Columns 4-6 present results for the IV estimation with right-hand-side variables lagged by two and three years. The coefficient estimate on internal citation remains positive and statistically significant.

to the focal firm's publications (by corporate and non-corporate patents). While the coefficient estimate on internal citation remains robust and similar in size to previous estimates (Table 5, Column 3), the coefficient estimate on external citation is negative and statistically indistinguishable from zero. A similar relationship is found when restricting external citation to citations received only from corporate patents (Column 2).

Not all citations represent profit-reducing spillovers. Profit-reducing spillovers are more likely when a rival uses the knowledge than when an unrelated firm uses it. Similarly, profit-reducing spillovers are more likely when recently generated knowledge is used by a rival, rather than when the knowledge is older and already diffused.

We build on Bloom et al. (2013) and Jaffe (1988) to construct SEGMENT and TECH as our measures of the proximity of citing and cited firms in product market space and technology space, respectively. Firms are close in product space if the distribution of sales across different product market segments is similar. Firms are close in technology space if the distribution of patents across technology classes is similar. Formally, the distance in technology space is the cosine of vectors representing the share of patents in 4-digit IPC classes for each pair of firms. Product market distance is measured analogously using industry segments (4-digit SIC codes level).[27] More details on the construction of SEGMENT and TECH are provided in Appendix 6.5.

We compute two external citation variables as SEGMENT-weighted and TECH-weighted number of outsider citations to the focal firm's publications (using SEGMENT and TECH as weights). Naturally, only citations by corporate patents are included in this analysis. Insofar as higher SEGMENT citations represent profit-reducing spillovers, we expect these to be negatively related to publications. We make no clear prediction for TECH citations.

Columns 3-5 present the estimation results of breaking up external citations by SEGMENT and TECH. The coefficient estimate on SEGMENT citations is negative, however statistically insignificant (Column 3).

---

[27] SEGMENT proximity for each cited–citing

firm pair is the absolute un-centered correlation between their sales segment share vectors, calculated as $\left| \frac{S_i' \times S_j}{\sqrt{S_i} \times \sqrt{S_j}} \right|$. s_i is business segment sales shares vector for firm i taken from Compustat's operating segments database. The measure ranges from zero (least correlated) to 1. Similarly, TECH proximity is computed based on firm's patent share distribution across technology fields (4-figit IPC).

Columns 4-5 restrict citations to recent publications (published no later than five years from the citing patent's grant year). As expected, external citations from close product market rivals are negatively related to publications. Conversely, citations from close technology rivals are positively related to publications (Column 4). This may capture unobserved firm-specific publications quality effect–quality is positively correlated with citations from technology rivals and with publications production. Finally, the SEGMENT and TECH estimates become larger in absolute value when restricting the sample to publishing firms (Column 5).[28]

Overall, our results are consistent with the view that a firm's investment in research depends, among other things, on how its research is used internally and externally. A firm whose research is used in its own inventive activity is likely to continue investing in research. However, a firm whose research spills over to rivals is likely to reduce its investment. An important empirical contribution of the present paper is quantifying this internal/external tradeoff. Based on the estimates from Column 4 in Table 8, one additional internal citation mitigates the negative effect of four segment-weighted external citations.

<div align="center">**Insert Table 8 here**</div>

4.4 R&D productivity

We next examine whether higher internal citation implies a greater productivity of R&D investment as measured by number of citation-weighted patents produced per R&D dollar invested. Table 9 presents the estimation results of the following specification:

$$
\ln(1 + Citation\ weighted\ patents_{it}) = \alpha_0 + \alpha_1 Share\ of\ inter\ citation_{it-1} \tag{2}
$$
$$
+ \alpha_2 Share\ of\ inter\ citation_{it-1} \times \ln\left(R\&D\ stock_{it-1}\right)
$$
$$
\alpha_3 \ln\left(R\&D\ stock_{it-1}\right) + \alpha_4 \ln\left(1 + Pub\ stock_{it-1}\right) + \eta_i + \tau_t + \epsilon_{it}
$$

Citation-weighted patents is the annual flow of patents weighted by the ratio of the number of citations each patent receives and average number of citations received by all other patents granted in the same year as the focal patent. *Share of Inter citation* is the ratio of internal citations from own patents to

---

[28] The observed pattern of result is not driven by any particular industry. See Tables A6 for variation across main industries. Appendix A2 includes a list of four-digit SIC codes included in the industry breakdown in Table A6.

number of citations received from all corporate patents.[29] The coefficients of interest are α_1 and α_2. We expect that firms with more scientific R&D programs to be more productive. Thus, we expect $\widehat{\alpha}_1 > 0$ and $\widehat{\alpha}_2 > 0$. As in previous analysis, all specifications control for firm fixed effects.

Column 1 shows a strong positive relationship between internal citation and R&D productivity ($\widehat{\alpha}_1 > 0$). Columns 2-3 add the interaction term between R&D stock and share of internal citation. As expected, $\widehat{\alpha}_2 > 0$ and is statistically significant. Column 3 restricts the sample to publishing firms with no substantial change in the estimates.

Insert Table 9 here

4.5 Stock market value

If internal citation increases private returns to research, whereas spillovers to rivals reduce private returns, this should be reflected not only in the level of publication output, but also in its value. We examine next the relationship between use of research and firm stock market value[30]. Following Griliches (1986) and Hall et al. (2005), we estimate the following specification:

$$\ln(Market\ value_{it}) = \alpha_0 + \alpha_1 \ln(1 + Internal\ pub\ stock_{it-1}) + \alpha_2 \ln(1 + External\ pub\ stock_{it-1})$$

$$+ \alpha_3 \ln Assets_{it-1} + \mathbf{Z}'_{it-1}\gamma + \eta_i + \tau_t + \epsilon_{it} \tag{3}$$

Internal pub stock and *External pub stock* are publications stock weighted by number of internal and external citations each publication receives, respectively[31]. Assets is the book value of physical capital[32] and $\mathbf{Z}$ is a vector of controls including lagged sales, R&D stock and patents stock. The coefficient estimates are amenable to different interpretations. We interpret these coefficients as reflecting the imputed value attributable to the relevant asset, or the "shadow price" of the asset (Hall et al., 2005). Our interest is at the coefficients α_1 and α_2. We expect $\widehat{\alpha}_1 > 0$ and $\widehat{\alpha}_2 < 0$. Table 10 presents the estimation results.

---

[29] Average value of *Share of Inter citation* is 0.013 with a standard deviation of 0.093.

[30] Market value is the sum of common stock, preferred stock and total debt net of current assets.

[31] Publications stock is computed as *Internal pub stock_t* = *Internal pub_t*+ *Internal pub stock_{t-1}*, where *Internal pub_t* is the number of internal citations publications receive in year t. *External pub stock_t* is computed in an equivalent way with external citations received. For example, if a publication receives one internal citation and two external citation, it adds one to *Internal pub stock_t* and two to *External pub stock_t*.

[32] Calculated as the sum of net plant, property and equipment, inventories, investments in unconsolidated subsidiaries, and intangibles other than R&D.

Consistent with previous research, Column 1 shows a positive relationship between publications stock and market value (Arora et al., 2015). Column 2 breaks up publications stock into internally- and externally-cited publications. As expected, $\widehat{\alpha}_1 > 0$, but contrary to our expectation $\widehat{\alpha}_2 > 0$ (significant at the 5% level).

Column 3 distinguishes between external citations received from rivals in product markets and those from other firms in the same technical domains. Thus, *External pub stock* is broken up into two separate measures, one where external citations are weighted by product market proximity (SEGMENT) and another where external citations are weighted by technology market proximity (TECH). This decomposition of external citations leads to a negative and statistically significant coefficient estimate on SEGMENT, as expecred. The coefficient estimate on TECH is positive, similar to our findings from Column 4, Table 8.

The estimates from Table 10, Column 3 indicate that one additional internal citation weighted publication mitigates the negative market value effect of approximately 3 external SEGMENT-weighted publications. Taken together, the evidence from Table 10 supports the view that private return to research is positively related to its internal use in invention, but negatively related to its use in invention by close product market rivals.

Insert Table 10 here

5 Conclusion

Using data on 4,736 publicly traded American firms over the period 1980-2006, this paper studies the relationship between use of corporate research in invention and corporate production of science. We systematically match all NPL (non-patent literature) references to publication records from Web of Science to learn about how corporate research is used in invention. Our primary contribution is providing systematic evidence of the private economic value of corporate research as an input into internal inventive activity and demonstrating that investment in research is strongly tied to how research benefits technology development.

We make several empirical contributions. First, we show that patent citations to scientific publications are a good measure of use of science in invention. We utilize the Carnegie Mellon Survey to show that firms whose patents cite science also report greater use of science in their R&D projects. Second, we show

that firms that are able to use their research in their inventions produce more of it. This relationship is stronger for new, high quality basic research and for citing patents in the core technology area of the inventing firm. We supplement these findings by showing that internal use is associated with higher R&D productivity and stock market valuation of scientific publications stock. Third, spillovers reduce private value and firms publish less when their research spills over to product market rivals.

Our findings support the view that firms invest in research because its scientific output feeds into downstream technology development. Our paper contributes to the growing discussion of why American firms are withdrawing from investment in science (Arora et al., 2015). To understand the causes and implications of the decline in corporate research, we must first develop a better understanding of why firms invest in research in the first instance. Over time, firms will invest less in research if the output of their research becomes relatively less important for the technology they develop, and spills over to product market rivals.

References

[1] Adams, J.D., 1990. Fundamental stocks of knowledge and productivity growth. Journal of Political Economy, 98(4), pp.673-702.

[2] Arora, A., Belenzon, S. and Patacconi, A., 2015. Killing the golden goose? The decline of science in corporate R&D (No. w20902). National Bureau of Economic Research.

[3] 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(3), pp.317-337.

[4] 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). Princeton University Press.

[5] Audretsch, D.B. and Stephan, P.E., 1996. Company-scientist locational links: The case of biotechnology. The American Economic Review, 86(3), pp.641-652.

[6] Azoulay, P., 2002. Do pharmaceutical sales respond to scientific evidence?. Journal of Economics & Management Strategy, 11(4), pp.551-594.

[7] Belenzon, S., 2012. Cumulative innovation and market value: evidence from patent citations. The Economic Journal, 122(559), pp.265-285.

[8] Bikard, M., 2015. Peer-Based Knowledge Validation: A Hurdle to the Flow of Academic Science to Inventors. Available at SSRN 2333413.

[9] Bloom, N., Schankerman, M. and Van Reenen, J., 2013. Identifying technology spillovers and product market rivalry. Econometrica, 81(4), pp.1347-1393.

[10] Breschi, S. and Catalini, C., 2010. Tracing the links between science and technology: An exploratory analysis of scientists' and inventors' networks.Research Policy, 39(1), pp.14-26.

[11] Bruneel, J., D'Este, P. and Salter, A., 2016. The impact of financial slack on explorative and exploitative knowledge sourcing from universities: evidence from the UK. Industrial and Corporate Change, 25(4), pp.689-706.

[12] Bush, V., 1945. Science: The endless frontier. Transactions of the Kansas Academy of Science (1903-), 48(3), pp.231-264.

[13] Carlson, W.B., 2013. Innovation and the Modern Corporation. Companion Encyclopedia of Science in the Twentieth Century, p.203.

[14] 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(2), pp.157-182.

[15] Cohen, W.M. and Levinthal, D.A., 1989. Innovation and learning: the two faces of R & D. The economic journal, 99(397), pp.569-596.

[16] 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) (No. w7552). National Bureau of Economic Research.

[17] Cyert, R.M. and March, J.G., 1963. A behavioral theory of the firm. Englewood Cliffs, NJ, 2

[18] Dasgupta, P. and David, P.A., 1994. Toward a new economics of science.Research policy, 23(5), pp.487-521.

[19] David, P.A., Mowery, D. and Steinmueller, W.E., 1992. Analysing the economic payoffs from basic research. Economics of innovation and New Technology, 2(1), pp.73-90.

[20] Gambardella, A., 1992. Competitive advantages from in-house scientific research: The US pharmaceutical industry in the 1980s. Research Policy, 21(5), pp.391-407.

[21] Gambardella, A., Panico, C. and Valentini, G., 2015. Strategic incentives to human capital. Strategic Management Journal, 36(1), pp.37-52.

[22] Gans, J.S., Murray, F.E. and Stern, S., 2013. Contracting over the disclosure of scientific knowledge: Intellectual property and academic publication (No. w19560). National Bureau of Economic Research.

[23] Graham, JR., CR Harvey and Rajgopal, S., 2005 The economic implications of corporate financial reporting. Journal of accounting and economics, 40 (1), 3-73.

[24] Griliches, Z., 1986. Productivity, R and D, and Basic Research at the Firm Level in the 1970's. The American Economic Review, 76(1), pp.141-154.

[25] Griliches Z. 1981. Market value, R&D, and patents. Economics Letters 7(2): 183–187.

[26] Hall, B.H., Jaffe, A. and Trajtenberg, M., 2005. Market value and patent citations. RAND Journal of economics, pp.16-38.

[27] Henderson, R. and Cockburn, I., 1994. Measuring competence? Exploring firm effects in pharmaceutical research. Strategic management journal,15(S1), pp.63-84.

[28] Hicks, D., 1995. Published papers, tacit competencies and corporate management of the public/private character of knowledge. Industrial and corporate change, 4(2), pp.401-424.

[29] Hicks, D., Breitzman, T., Olivastro, D. and Hamilton, K., 2001. The changing composition of innovative activity in the US—a portrait based on patent analysis. Research policy, 30(4), pp.681-703.

[30] Hounshell, D.A. and Smith, J.K., 1988. Science and Corporate Strategy: Du Pont R and D, 1902-1980. Cambridge University Press.

[31] Jaffe, A.B., 1988. Demand and supply influences in R & D intensity and productivity growth. The Review of Economics and Statistics, pp.431-437.

[32] Klasa, S., Ortiz-Molina, H., Serfling, M. and Srinivasan, S., 2015. Protection of trade secrets and capital structure decisions. Available at SSRN 2439216.

[33] Kline, S.J. and Rosenberg, N., 1986. An overview of innovation. The positive sum strategy: Harnessing technology for economic growth, 14, p.640.

[34] Kuznets, S., 1971. Prize Lecture: Modern Economic Growth: Findings and Reflections. Nobelprize.org.

[35] Levin, R.C., Klevorick, A.K., Nelson, R.R., Winter, S.G., Gilbert, R. and Griliches, Z., 1987. Appropriating the returns from industrial research and development. Brookings papers on economic activity, 1987(3), pp.783-831.

[36] Lichtenberg, F.R., 1986. Private Investment in R&D to Signal Ability to Perform Government Contracts. NBER Working Paper, (w1974).

[37] Maclaurin, W.R., 1953. The sequence from invention to innovation and its relation to economic growth. The Quarterly Journal of Economics, pp.97-111.

[38] Mansfield, E., 1980. Basic research and productivity increase in manufacturing. The American Economic Review, 70(5), pp.863-873.

[39] Marx, M., Strumsky, D. and Fleming, L., 2009. Mobility, skills, and the Michigan non-compete experiment. Management Science, 55(6), pp.875-889.

[40] McMillan, G.S., Narin, F. and Deeds, D.L., 2000. An analysis of the critical role of public science in innovation: the case of biotechnology. Research policy, 29(1), pp.1-8.

[41] Mowery, D.C., 1995. International Computer Software Industry. Oxford University Press, Inc.

[42] Narin, F., Pinski, G., & Gee, H. H. (1976). Structure of the biomedical literature. Journal of the American Society for Information Science, 27(1), 25-45.

[43] Narin, F. and Noma, E., 1985. Is technology becoming science? Scientometrics, 7(3-6), pp.369-381.

[44] Narin, F., Hamilton, K.S. and Olivastro, D., 1997. The increasing linkage between US technology and public science. Research policy, 26(3), pp.317-330.

[45] Nelson, R.R., 1959. The simple economics of basic scientific research. Journal of political economy, 67(3), pp.297-306.

[46] Popp, D., 2016. From Science to Technology: The Value of Knowledge From Different Energy Research Institutions (No. w22573). National Bureau of Economic Research.

[47] Reich, L. S., 1985. The Making of American Industrial Research. New York: Cambridge University Press.

[48] Rosenberg, N., 1990. Why do firms do basic research (with their own money)?. Research policy, 19(2), pp.165-174.

[49] Sauermann, H. and Cohen, W.M., 2010. What makes them tick? Employee motives and firm innovation. Management Science, 56(12), pp.2134-2153.

[50] Stern, S., 2004. Do scientists pay to be scientists? Management science,50(6), pp.835-853.

6   Appendix.

6.1   Sample construction

We implement various matching procedures to construct two main datasets, one at the firm-year level and another at the citation-publication level, including (i) matching scientific publications to Compustat companies; (ii) mapping patent citations to publications; (iii) matching patent data to Compustat; and (iv) dynamic matching of Compustat accounting information. We discuss these procedures below.

A parent company and a subsidiary may have different identification numbers and records within the Compustat data. Also, a single company may correspond to multiple firm identifiers (CUSIPs or GVKEYs) within the Compustat database due to changes in ownership structure and accounting changes over the sample period that can lead to a change in its identification number. Appendix Table B1 illustrates the challenges in assigning a unique id over time based on the firm id in Compustat. To deal with these challenges, we implement several procedures.

First, we rely on the NBER 2006 patent data project, which corresponds to our subsample years, to identify multiple Compustat records that are associated with a single company in our subsample.

Second, we further merge parent Compustat companies and independent Compustat subsidiaries and related joint-ventures that appear in our initial subsample using the "CGS Associated Issuers" database, which links related issuers in the Compustat database and other online sources. For example, we merge *ARMSTRONG WORLD INDUSTRIES* under its holding company *ARMSTRONG HOLDINGS INC* as well as merge the joint venture, *DOW CORNING CORP*, under both his parent companies, *DOW CHEMICAL* and *CORNING INC*. We manually modify the NBER data according to these changes. Third, we uniquely identify each parent company by a 9-digit CUSIP (our UO-COMPANY variable) and up to five associated CUSIPs in case of multiple related firm CUSIPs and subsidiary CUSIPs. In addition to our UO-COMPANY identifier, we assign each company a unique NBER "PDPCO" code. This enables us to dynamically match our dataset with Compustat accounting data and with NBER patent data. We exclude from our sample non-patenting firms. Lastly, we exclude firms that are not headquartered in the United States, based on Compustat current records. Appendix Table B2 presents an example of the dynamic match.

The above procedures leave us with our final estimation sample of an unbalanced panel including 4,736 publicly traded US headquartered companies and 57,765 firm-year observations over the period 1980-2006. These firms have at least one year of positive R&D expenditures and at least one patent from 1980 through 2006.

6.2   Matching scientific publications to Compustat companies

After obtaining our initial subsample of firms we proceed to match our firm sample to publications data to capture their investment in science. We obtain publications data from the Web of Science database (previously known as ISI Web of Knowledge). We include articles from journals covered in the "Science Citation Index" and "Conference Proceedings Citation Index - Science," excluding social sciences, arts and humanities articles.

Each publication record contains detailed information including title of the publication, authors, journal info and our main variable of interest, an affiliation field with name and address of the publishing institute or company in case of a corporate publication. This field can include more than one listing in case of a collaborative publication, for example, "*DARLEY M (reprint author), TEXAS INSTRUMENTS INC, DEPT DATAPATH VLSI PROD SEMICOND GRP 8330 LBJ FREEWAY, POB 655303, DALLAS, TX 75265 USA SUN MICROSYST INC, MT VIEW, CA USA*".

Companies appear in the Compustat file under their most current name with no records of previous names. Since company names may change over the course of our sample years (e.g., due to mergers and acquisitions), we manually search online sources (e.g., Bloomberg, Opencorporates, Crunchbase websites) and check older Compustat datasets and the "CGS Associated Issuers" in order to identify previous names relevant to our sample period. For example, EG&G, which changed its name to "PerkinElmer Inc" in 1999 after buying the Analytical Instrument division from Perkin-Elmer will appear under both names in our dataset. Similarly, "*APPLIED MOLECULAR GENETICS INC*", which changed its name in 1983 to "*AMGEN INC*" will appear under both names in our dataset, as well as the pharmaceutical company "*WYETH*", which up to 2002 was known as "*AMERICAN HOME PRODUCTS*", "*3M CO*", formerly known as the "*MINNESOTA MINING AND MANUFACTURING COMPANY*", and

27

"*SPECTRA DIODE LAB*", which changed its name to "*SDL, INC*". In addition to the parent company name we also check related subsidiary names using SDC Platinum (M&A data) and other online sources. Lastly, we allocate abbreviations that are commonly used by companies instead of their official name. For example, "*INTERNATIONAL BUSINESS MACHINES CORP*", will also appear under its common abbreviation "*IBM*" and "*GENERAL ELECTRIC CO*" under "*GE.*"

Company name was first standardized by cleaning all non- alphabetic characters as well as Compustat related indicators and converting all strings to uppercase characters. Where possible, we omitted legal entity endings and other common words (e.g. INC, CORP, LTD, PLC, THE, LAB, PHARMA-CEUTICAL) to maximize the matching rates (e.g., "*XEROX CORP*" was standardized to "*XEROX*", "*ABBOTT LABORATORIES*" to "*ABBOTT*"). However, in cases where the company's name is too short, generic or can match to other strings within the address field, we preserved the original name to avoid mismatches. For example, omitting the legal entity from "*QUANTUM CORP*" would mismatch it to various research institutions such as "*TEXAS STATE UNIV CTR APPL QUANTUM ELECTR DEPT*". Similarly, "*MALLINCKRODT INC*" can mismatch with "*EDWARD MALLINCKRODT INST*" or with "*HARVARD UNIV MALLINCKRODT CHEM LAB*" and "*KELLOGG CO*" can mismatch with "*M.W. KELLOGG*". For cases where we cannot omit the legal entity, we try adding additional relevant names to improve the match, for example for "*MERCK & CO*" we also include "*MERCK RESEARCH LAB*" and for "*GTE CORP*" we include "*GTE LAB*". To further improve the quality of match, we obtained the list of most frequent company names of the first publisher within the affiliation string and adjusted our company name list accordingly.

One last step in standardizing the company names is to fit it to the publication affiliation field format that contains many abbreviated words. For this process, we formed a list that includes over 80 abbreviated words matched to their various origins as well as to other forms of abbreviations. For example, LABORA-TORIES, LABORATORY, LABS, LABO, LABORATORIE, LABORATARI, LABORATARIO, LABO-RATARIA, LABORATORIET and LABORATORIUM were all abbreviated to "LAB". The list was compiled from the most frequent abbreviated words in the Address field (accordingly, the list is targeted to our sample). Appendix Table B3 presents a list of the most frequent abbreviated words.

Finally, we apply a many-to-many match between each standardized company name and the affiliation field for each publication (approximately 14mm publications and 5.5K names), while allowing for more than one firm to be matched to each publication (to allow for collaborative publications). We use STATA's "regexm" command[33] to detect whether the affiliation field contains each company name. In addition to our automated algorithm, we perform extensive manual checks to detect cases that can cause mismatches and verify matches by comparing the address listed within Compustat to the address in the publication data. For example, to distinguish between "*THERATECH INC / UTAH*" and "*THERATECH INC*" we verify that the address of the firm under the affiliation field is in Salt Lake City.

We find approximately 300 thousand articles from more than 5000 different journals that were published from 1980 through 2006, with at least one author employed by our sample of Compustat firms. At the end of this procedure, we obtain a match between a unique publication id and a UO company id.

---

[33]Stata's "strdist" command is based on the Levenshtein distance score, which measures the distance between two strings by the minimum number of character edits required to gain an exact match.

6.3 Matching patent data

We apply the strategy used in the NBER 2006 patent data project (Hall et al., 2001) to perform a dynamic match of patents to our subset of Compustat firms. The NBER dynamic reassignment of patents accounts for changes in Compustat identification numbers and M&A reassignment of patents based on SDC data. We make manual adjustments to the NBER data based on our aggregation of the data under a parent "UO" company. For example, while *BOEING CAPITAL CORP* and *BOEING CO* are treated as separate companies in the NBER database, we merge *BOEING CAPITAL CORP* under *BOEING CO* as it is a wholly owned subsidiary of the company. We adjust the NBER data in such way that all of *BOEING CAPITAL CORP* patents are assigned to *BOEING CO*. Using this process we identify the original UO firm of each patent and also account for reassignment of patents over time. In case a patent has several assignees, we match the patent to several firms and assign fractional patent ownership to each assignee (i.e., 1/number of assignees).

6.4 Matching NPL citations to Web of Science

Patent citations to science are obtained from the Non-Patent Literature (NPL) citations section located at the front page of patents taken from PatStat database. An example of a front-page patent citation to non-patent literature is provided in Appendix Figure A5. Using all patents granted in the period 1980-2014 (including corporate and non-corporate patents), we match NPLs to corporate publications from Web of Science (approximately 14mm citations and 300k corporate publications). This is a central match and the most challenging one, due to differences in structure between NPLs and publications. We begin with a many-to-many match, allowing more than one publication to be matched to each citation. For each possible records pair, we construct a score that captures the degree of textual overlap between the title, journal, authors and publication year. To exclude mismatches, we use a more detailed matching algorithm that is based on different sources of publication information: standardized authors' names, number of authors, article title, journal name and year of publication. The matching algorithms accounts for misspelling, unstructured text, incomplete references, and other issues that may cause mismatches.[34]

The first step is to match between the publication's "Title" field and the title that is located within the citation string. There are two main problems: (i) the position of the title within the citation is not fixed (ii) there may be small variation in the title (e.g., "GIVE" *vs.* "GIVES") and thus an exact match may not perform well. To overcome these problems we implement a fuzzy match algorithm. After we standardize and clean the different fields, we measure the length-difference between the citation string and the publication title string. Then, using STATA's "strdist" command we calculated the distance between the two strings. We use the difference between the length difference and distance as a measure of proximity of the titles. We supplement this measure with an exact match of the first part of the title. In some cases the title is missing from the citation string. In such cases we rely more on other available

---

[34] The following example (from the first line in Appendix Table B4) illustrates the matching challenge. NPL citation: LIN, KUN SHAN, ET AL., SOFTWARE RULES GIVES PERSONAL COMPUTER REAL WORD POWER , INTERNATIONAL ELECTRONICS, VOL. 53, NO. 3, FEB. 10, 1981, PP. 122 125. Matched Publication: Title: SOFTWARE RULES GIVE PERSONAL-COMPUTER REAL WORD POWER, Authors: LIN KS, FRANTZ GA, GOUDIE K, Journal information: ELECTRONICS 54 (3): 122-125 1981.

features to determine the final match.

Second, we match between the publication's "Authors" field and the authors listed within the citation string. As with the title, we cannot identify the exact location where the authors are listed within the citation string since the location varies from one citation to another. In addition, there are several differences in how names are written: (i) Last name only vs. full names; (ii) names vs. initials (e.g., LIN KS vs. LIN KUN SHAN); (iii) listing of all authors vs. one author followed (or not) by "et al."; (iv) order of last and first names within the string. To verify a match by authors we first count the number of authors listed in the publication record. We then check whether the citation string contains "et al.". To mitigate the name variation problem, we implement an algorithm that matches different variations of the authors' name to the citation (including transformation of last and/or first and/or middle name to initials and changes in order listed). In cases where several authors are listed under the publication and "et al." does not appear within the citation we perform a one-to-many match between the citation and each author and assume that at least 80% of the authors must be matched to the citation to determine a match. For cases where several authors are listed in the publication and only one is matched within the citation while "et al." is omitted, we rely more on match results in other features to determine the final match.

Next, we match journal information including standardized journal's name, year published, page numbers and volume, while accounting for typos, abbreviations and differences in format between the datasets (e.g., "INTERNATIONAL ELECTRONICS" vs. "ELECTRONICS"; "VOL. 53, NO. 3" vs. "54 (3)").

Finally, we combine the match results for the different features (title, authors and journal information) using different weights according to their relative importance, in order to determine a final match. We perform extensive manual checks to confirm matches. At the end of this procedure, we obtain unique identification numbers for the citation, the citing patent and the cited publication.

6.5   SEGMENT and TECH proximity measures

We build on Bloom et al. (2013) and Jaffe (1988) to construct the measures SEGMENT and TECH as the correlations between firm pairs. SEGMENT proximity is computed based on the distribution of line of business listed within the Compustat operating segments database. We use the dynamic match as explained previously to match our UO firms to Compustat segment database and calculate their sales share over the complete sample period (1980-2006) in each segment. There are a total of 60 segments (out of 69 available) related to our citing and cited firms. We generate a vector for each UO firm based on the distribution of sales share in each of these 60 segments. The SEGMENT proximity for each cited–citing firm pair is the absolute un-centered correlation between their vectors, $\frac{S_i' \times S_j}{\sqrt{S_i} \times \sqrt{S_j}}$ where S_i denotes the vector of the shares of firm i's sales in different segments. In case of several assignee firms matched to the cited patent, we compute the average distance between the different firms. The measure ranges from zero (least correlated) to one (fully correlated).

Similarly, TECH is computed based on each firm's patent share distribution across different technology fields. We generate a vector for each UO firm based on its granted patents share in each 3-digit main IPC

over the complete sample period (1980-2006). The TECH proximity for each cited–citing firm pair is the absolute un-centered correlation distance between their IPC-share vectors, $\frac{T_i' \times T_j}{\sqrt{T_i} \times \sqrt{T_j}}$ where T_i denotes the vector of the shares of firm i's patents in the different technology fields.

6.6 Inventor-author overlap

Inventor-author overlap is measured as the share of patents per firm-year that include an inventor who is also an author of a corporate publication by the same firm published no later than 3 years from the patent's grant year.

We perform the following steps to construct the overlap measure. First, we standardize inventor and author names for all non-collaborative patents and corporate publications related to our sample firms during the sample period (1980-2006). For patent inventor names, we use the HBS Patent Inventor Database that identifies individual inventors for each patent and conveniently lists their first-name and last-name in separate fields. For publication author names, we use the publication's author list, which lists names by last name followed by initials (e.g., "MALIK RJ , HAYES JR , CAPASSO F , ALAVI K , CHO AY"). The standardized name format we implement for the match is last name followed by first name initial, all in uppercase letters (e.g., MALIK R).

Next, for each firm-grant year we perform a many to many match between each patent's inventor-standardized names and author-standardized names of publications published by the focal firm up to 3 years prior to the patent's grant year. For example, IBM patent US6013336 (granted in 2000) includes an inventor, Peter Baumgart, who also published a paper in 1997: "Wang, R.H., Raman, V., Baumgart, P., Spool, A.M. and Deline, V., 1997. Tribology of laser textured disks with thin overcoat. IEEE Transactions on Magnetics, 33(5), pp.3184-3186."

Finally, we compute for each firm-grant year the share of non-collaborative patents that their inventor list includes at least one author of a corporate publication published by the firm up to 3 years prior to the patent's grant year. The average value of inventor-author overlap is 0.21 median is 0, standard deviation is 0.33 and the 10th and 90th percentile values are 0 and 0.9, respectively. Figures B1-B2 show that there is heterogeneity in inventor-author overlap across our sample firms.

6.7 Instrumental variable II: Profit shocks due to exchange rate fluctuations

A different source of variation in internal use arises from changes in the mix of downstream innovation activities. Whereas incremental innovation is less likely to use science, including internally generated science, exploratory innovation is more likely to build on scientific advances and be guided by them. Exploration has more uncertain outcomes and is less likely during lean times. Put differently, exploratory innovation requires financial slack, because "*slack provides a source of funds for innovations that would not be approved in the case of scarcity*" (Cyert and March, 1963, p. 279). The opposite relationship has been argued between slack and exploitation. Exploitative innovation, also called 'problemistic search', is directed towards finding an immediate solution to a specific problem (Levinthal and March, 1981; March, 1991; Greve, 2003, 2007; Bruneel et al., 2016). For example, Bruneel et al. (2016), building on survey data from 2002-2006, find that high levels of financial slack (measured by cash flow) are associated

with British firms' engagement in explorative knowledge sourcing from universities, whereas low levels of slack are associated with exploitative knowledge sourcing. Moreover, consistent with the 'slack search' view, Graham at al. (2004) report that 80% of the 401 executives they surveyed would decrease their discretionary R&D spending and delay starting a new project in order to meet an earning target.

We exploit financial shocks using foreign currency devaluations for export-oriented American firms. When dollar-denominated profits drop due to a stronger US dollar (USD), firms would engage in more exploitation and less exploration. Assuming that exploitation builds less on science, we expect less internal use of research (that is, patent citations to own science) as firm patents become more exploitative and less exploratory. The main idea is that exchange rates affect profits of firms with foreign subsidiaries, and profits in turn affect the decision of firms to exploit science. Our identifying assumption is that short-term profit shocks affect the decision of firms to use science, but short-term profit shock does not affect the value of scientific research independently from its effect on use. Specifically, we assume that (i) the future profits depend upon the realized level of exchange rates, so that conditioning on current exchange rates, the long run value of research is unaffected for a given level of internal use, and (ii) shocks to exchange rates lead to changes in internal use. We demonstrate that negative shocks to exchange rates lead to a decrease in profitability and that negative exchange rate shocks shift downstream innovation away from exploration and towards exploitation, as reflected in reduced patent citations to science. We assume that in turn this will also reduce internal use of the firm's own science.

Our instrument uses the yearly change in foreign exchange rates weighted by firm-specific weights. We use two sets of weights: (i) foreign subsidiaries by firm i in each country and (ii) the industry-level export of goods between the US and each foreign country in the main industry of firm i. We include only manufacturing firms (SIC 20-39). The data are for the years 1990-2006 (which allow us to include subsidiaries in former USSR countries). The weighted change in exchange rates is computed as:

$$\Delta d_{it} = \sum_c \sum_j S_{ic} \times Export_{jct} \times \Delta d_{tc}$$

Where, Δd_{it} is firm-year change in the weighted-average value of foreign currency relative to USD. It includes only manufacturing industries and covers the years 1990 to 2006. Higher Δd_{it} means USD becomes stronger indicating a negative shock to USD-denominated profits. Δd_{tc} is the change in the USD denominated value of country's c currency between years t and t-1. Annual exchange rates are official exchange rates from the World Bank's World Development Indicators. We compute an annual average based on monthly averages (local currency units relative to the U.S. dollar). For countries adopting the Euro currency, we adjusted Δd_{tc} to zero, for the year of the currency change. Changes in annual exchange rate vary from a 10th percentile value of -0.09 to 90th percentile value of 0.24 (mean of 0.22 and standard deviation of 1.4). Examples of extreme devaluation include the Brazilian Real that depreciated by more than 1600% during 1993 and 1994, the Indian Rupee depreciated by 30% between 1990-1991, the Chinese Renminbi that depreciated by 50% between 1993-1994 and the British Pound Sterling that depreciated by 17% between 1992-1993.

S_{ic} is the share of subsidiaries firm i has in country c of all subsidiaries owned by firm i. Subsidiaries information for our sample firm are from the 2014 Orbis database, maintained by Bureau VanDyke. The

average firm has 54 foreign subsidiaries in 8 different countries. Example of affiliates' countries include: Great Britain (9%), Germany (7%), France (6%), Netherlands (5%), Italy (4%), China (4%), Brazil (3%), India (3%), Russia (1%). $Export_{jct}$ is the share of export by industry j (where firm i operates) to country c in year t. Annual industry export flows between US and foreign countries are based on Schott (2010)[35]. Δd_{it} varies from a 10th percentile value of -0.38 to 90th percentile value of 0.45 (mean of 0.02 and standard deviation of 0.5) for our estimation sample.

Appendix Table A5 presents the estimation results. Our sample is conditioned on firm with at least one publication stock. It includes 1901 publishing firms, out of which 981 firms have subsidiaries in 52 different countries. We compute a dummy variable based on the measure, which receives the value of 1 for devaluation ($\Delta d_{it}>0$). We lag Δd dummy by two periods as our instrument for one-period lagged internal use and control for the level of exchange rate, d_{it}.[36]

Columns 1-4 present the relationship between Δd dummy with profits (EBIDTA) and the use of science–the average number of patent citations to non-patent literature (NPL), per patent. Consistent with our proposed mechanism, foreign currency devaluation is associated with drop in profits. Based on the estimates from Column 1 and evaluated at the sample average, devaluation is associated with 8% drop in EBIDTA. In other words, exchange rates shocks affect short-term profitability.

Columns 2-4 further show foreign currency devaluation is associated with drop in use of science. Column 2 shows that devaluation is associated with reduction of 0.5 citations per patent to the non-patent literature (NPL) which is equivalent to a 9.7% decrease at the mean. Column 3 shows that results hold when restricting the sample to firm-years with patents. Based on the estimates from Column 3, devaluation is associated with 12.5% decrease in average NPL per patent. For Column 4 the dependent variable is share of patents per year with at least one citation to NPL. Evaluated at the sample average, devaluation is associated with 12% drop in share of patents citing NPL. These results are consistent with firms conducting less exploratory inventive activity in leaner times. It is plausible that this would also imply less use of internal science.

Columns 5-7 present the results using devaluation as an instrument for internal use. The first stage estimation instruments internal citation with a dummy variable for $\Delta d_{it}>0$. As expected, devaluation of foreign currencies is negatively associated with internal use. Based on the estimates from Column 5, devaluation is associated with 8.1% decrease in average internal use (a decrease of 0.65 internal cites). We reject the test for weak instruments with a Kleibergen-Paap F statistic=59 (Staiger and Stock, 1997).

Column 7 presents the second stage estimation results, regressing the log of number of publications against the predicted lagged use of science due to profitability shocks. The coefficient estimate on internal citation increases from 0.9 (OLS estimation, Column 6) to 1.4. Based on this estimate, a 10 percent increase in internal citation is associated with approximately 14% increase in annual publications.

Lastly, Columns 8-10 include jointly our two instruments, IDD and devaluation, in a single two-stage least-squares specification. The same pattern of results remains in the first and second stage estimations. The Hansen test for overidentifying restrictions is consistent with the instruments being

---

[35]Data are available for download at: http://faculty.som.yale.edu/peterschott/sub_international.htm

[36]$d_{it} = \sum_c \sum_j S_{ic} \times Export_{jct} \times d_{tc}$, where d_{tc} is the exchange rate of country's c currency in USD at time t.

valid; we are unable to reject the null hypothesis that the instruments are uncorrelated with the error term and correctly excluded from the estimated specification (p-value for overidentifying restrictions=0.44, Hansen J statistic=0.59).

References

[1] Bruneel, J., D'Este, P. and Salter, A., 2016. The impact of financial slack on explorative and exploitative knowledge sourcing from universities: evidence from the UK. Industrial and Corporate Change, 25(4), pp.689-706.

[2] Cyert, R.M. and March, J.G., 1963. A behavioral theory of the firm. Englewood Cliffs, NJ, 2.

[3] Graham, JR., CR Harvey and Rajgopal, S., 2005 The economic implications of corporate financial reporting. Journal of accounting and economics, 40 (1), 3-73.

[4] Greve, H.R., 2003. A behavioral theory of R&D expenditures and innovations: Evidence from shipbuilding. Academy of management journal, 46(6), pp.685-702.

[5] Greve, H.R., 2007. Exploration and exploitation in product innovation. Industrial and Corporate Change, 16(5), pp.945-975.

[6] Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." NBER Working Paper 8498.

[7] Klasa, S., Ortiz-Molina, H., Serfling, M. and Srinivasan, S., 2015. Protection of trade secrets and capital structure decisions. Available at SSRN 2439216.

[8] Levinthal, D. and March, J.G., 1981. A model of adaptive organizational search. Journal of Economic Behavior & Organization, 2(4), pp.307-333.

[9] March, J.G., 1991. Exploration and exploitation in organizational learning. Organization science, 2(1), pp.71-87.

[10] Schott, P., 2010. US manufacturing exports and imports by SIC or NAICS category and partner country, 1972 to 2005. Notes.

[11] Staiger, D. and Stock, J.H., 1997. Instrumental Variables Regression with Weak Instruments. Econometrica, 65(3), pp.557-586.

Table 1. Summary Statistics for Main Variables

VARIABLE	# Obs.	# Firms	Mean	Std. Dev.	Distribution		
					10th	50th	90th
Publications count	32,923	2,413	9	51	0	0	9
Publications stock	32,923	2,413	104	762	0	2	75
Patents stock	57,765	4,736	65	379	0	3	75
Patents count	57,765	4,736	12	74	0	0	14
R&D expenditures($mm)	57,765	4,736	55	319	0.25	5	67
R&D stock($mm)	57,765	4,736	231	1,467	0.5	17	260
Market value ($mm)	57,765	4,736	2,105	15,280	5	87	2,271
Sales ($mm)	57,765	4,736	1,394	8,063	2	66	1,997
Assets ($mm)	57,765	4,736	954	6,281	1	30	1,237
Inventor-author overlap	16,538	2,081	0.2	0.3	0	0	0.9

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 4,736 US HQ publicly traded companies (out of which 2,413 are publishing companies) over the sample period, 1980-2006. These firms have at least one year with positive R&D expenditures and at least one patent during the sample period. The sample for all publication variables is restricted to publishing firms. Inventor-author overlap is the share of patents per year with inventors who include at least one author of a publication published by the same firm in the three-year window prior to the patent's grant year. For Inventor-author overlap, the sample is conditional on at least one publication stock and on firm-years with granted patents.

Table 2. Summary Statistics for Citations Variables (only firms with cites to own publications)

	(1)	(2)	(3)	(4)
	Number of firms with positive	Average value	Number of citing patents per firm-	Number of cited publications per
VARIABLE	values	per firm-year	year	firm-year
Total patent citations to own publications	799	8	7	5
Internal patent citations to own publications	388	1	1	1
External patent citations to own publications	760	7	6	4

Notes: This table provides summary statistics for the main citation variables used in the econometric analysis. The sample is at the firm-year level and is conditional on firms with positive patent citations to publications.

Table 3. High Internal Use *vs.* Low Internal Use (only firms with cites to own publications)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		High Share of internal citations (> median)			Low Share of internal citations (≤ median)		
VARIABLE	(3) minus (6)	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
Publications flow/R&D expenditures	0.1**	388	0.3	0.5	411	0.2	0.4
Patents stock/R&D expenditures	0.06	388	0.43	0.4	411	0.37	0.5
R&D expenditures/Sales	1.3**	388	2.3	5.2	411	1.0	2.8
Inventor-author overlap	0.2**	388	0.4	0.3	411	0.2	0.3

Notes: This table presents mean comparison tests for firms with high share of internal citations *vs.* firms with low share of internal citations. Share of internal citations is the ratio of citations the firm's publications receive from its own patents to total citations received. The sample is conditional on firms with positive citations. The unit of analysis is a firm, yearly values are averaged over the period 1980-2006. * and ** denote that the difference in means is significant at the 5% and 1% level, respectively.

Table 4. Supporting Evidence from 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 the main research field's findings (Q.22)	Importance 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 citation to publications per patent and the 1994 Carnegie Mellon survey (CMS) questions response (Cohen et al., 2000) related to the importance of research findings as an input to the firm's R&D projects. The relevant CMS questions are mentioned in the main text. The sample includes only patenting firms. In Column 3, the sample is restricted to firms that indicated their main research field in Q22 (excluding 'Others' category). For *Citations to articles in main research field*, publications were classified to research fields based on WoS journal subject category. *Citations to corporate articles* include citation to publications by our main sample of Compustat firms. Citations to patents include backward citations to patents. Robust standard errors in parentheses.

Table 5. Internal Use and Publication Output

	Dependent variable: ln(1+number of publications)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Pooled	Between-firms	Within-firms	Publishing firms only	Backward patent citation to own patents	Internal citation, average per patent	Share internal citation of total citation received by own pubs	Share internal citation of total citation made to corp pubs
ln(1+Internal citation to publications)$_{t-1}$	0.681	1.692	0.110	0.098	0.103			
	(0.042)	(0.113)	(0.026)	(0.025)	(0.026)			
ln(1+Self-citation)$_{t-1}$					0.014			
					(0.007)			
Internal patent citation, average per patent						0.035		
						(0.018)		
Internal citation/Total citations received$_{t-1}$							0.099	
							(0.035)	
Internal citation/Total citations made$_{t-1}$								0.068
								(0.032)
ln(R&D stock)$_{t-1}$	0.094	0.076	0.065	0.096	0.06	0.066	0.066	0.066
	(0.007)	(0.007)	(0.009)	(0.014)	(0.009)	(0.004)	(0.004)	(0.004)
ln(1+Patent stock)$_{t-1}$	0.142	0.133	0.073	0.085	0.065	0.080	0.079	0.080
	(0.012)	(0.009)	(0.011)	(0.014)	(0.012)	(0.005)	(0.005)	(0.005)
ln(Sales)$_{t-1}$	0.028	0.023	0.047	0.088	0.047	0.047	0.047	0.047
	(0.005)	(0.005)	(0.006)	(0.010)	(0.006)	(0.003)	(0.003)	(0.003)
Firm fixed-effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies (4 digit)	Yes	Yes	-	-	-	-	-	-
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample average Publication	5.2	2.9	5.2	9.1	5.2	5.2	5.2	5.2
Std. Internal citation variable	3.6	1.8	3.6	4.7	3.6	0.2	0.093	0.107
Number of firms	4,634	4,634	4,634	2,380	4,634	4,634	4,634	4,634
Observations	53,029	4,634	53,029	30,510	53,029	53,029	53,029	53,029
R-squared	0.64	0.66	0.87	0.85	0.87	0.87	0.87	0.87

Notes: This table presents OLS estimation results for the relationship between past internal patent citation to own publications and future annual publications, for the period 1980-2006. *Internal citation to publications* include patent citations up to year t-1 to publications published up to the same year. Column 2 averages variables at the firm level and performs a cross section analysis. In Column 5, *Self-citation* is defined as average number of patent citations to own patents per firm-year. In Column 6, internal citations are measured as average citations to own publications, per corporate patent. In Column 7, share of internal citations is defined as ratio of citations the firm's publications receive from own patents to total citations received by the focal firm's publications. In Column 8, share of internal citations is defined as ratio of citations the firm's publications received from own patents to total citations to corporate science made by the focal firm's patents. All specifications include a dummy variable that receives the value of one for firms that never published up to the focal year. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

Table 6. Internal Use and Publication Output: Heterogeneous Effects

	(1)	(2)	(3)	(4)	(5)	(6)
		Basic vs. Applied Publications (CHI)	Basic vs. Applied Publications (JIF)	High vs. Low quality Publications	Core vs. Non-Core Tech	High vs. Low quality Patens
	New vs. Old science					
ln(1+Internal citation to publications, NEW)$_{t-1}$	0.130					
	(0.017)					
ln(1+Internal citation to publications, OLD)$_{t-1}$	0.028					
	(0.015)					
ln(1+Internal citation to publications, BASIC)$_{t-1}$		0.132	0.122			
		(0.015)	(0.014)			
ln(1+Internal citation to publications, APPLIED)$_{t-1}$		-0.003	-0.007			
		(0.021)	(0.022)			
ln(1+Internal citation to publications, High Quality)$_{t-1}$				0.116		0.091
				(0.014)		(0.016)
ln(1+Internal citation to publications, Low Quality)$_{t-1}$				0.001		0.054
				(0.024)		(0.018)
ln(1+Internal citation to publications, CORE)$_{t-1}$					0.120	
					(0.017)	
ln(1+Internal citation to publications, NON-CORE)$_{t-1}$					0.044	
					(0.016)	
ln(R&D stock)$_{t-1}$	0.065	0.065	0.065	0.065	0.065	0.065
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
ln(1+Patent stock)$_{t-1}$	0.074	0.075	0.074	0.074	0.074	0.074
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
ln(Sales)$_{t-1}$	0.047	0.047	0.047	0.047	0.047	0.047
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Difference of coefficients (significance level)	0.000	0.000	0.000	0.001	0.005	0.212
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable sample average	5.2	5.2	5.2	5.2	5.2	5.2
Number of firms	4,634	4,634	4,634	4,634	4,634	4,634
Observations	53,029	53,029	53,029	53,029	53,029	53,029
R-squared	0.87	0.87	0.87	0.87	0.87	0.87

Notes: This table presents OLS estimation results for the relationship between past patent citations to own publications and future annual publications, while distinguishing between different types of internal citations. All Internal citation variables include patent citations up to year t-1 to publications published up to the same year. See main text for exact variable definition. All specifications include a dummy variable that receives the value of one for firms that never published up to the focal year. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

Table 7. Instrumental Variable Estimation: Inventor-Author Overlap, Inevitable Disclosure Doctrine and Publication Output (Sample: Publishing Firms)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Inventor-Author Overlap				IV: Inevitable Disclosure Doctrine		
Dependent variable:	Dummy for internal citation		Dummy for external citation	Inventor-author overlap	ln(1+Internal citation)$_{t-1}$	ln(1+Number of publications)	
	Pooled	Firm FE	Firm FE	Firm FE	First Stage	OLS	2SLS
Inventor-Author overlap$_t$	0.152	0.079	-0.009				
	(0.012)	(0.011)	(0.012)				
IDD-Mobility$_{t-2}$				0.003	0.017		
				(0.001)	(0.002)		
ln(Internal citation to own publications)$_{t-1}$						0.111	0.670
						(0.012)	(0.104)
ln(1+Publication stock)$_{t-1}$	0.061	0.065	0.142				
	(0.004)	(0.006)	(0.007)				
ln(1+Patent stock)$_{t-1}$	0.023	0.031	0.032		0.020	0.130	0.116
	(0.003)	(0.005)	(0.006)		(0.004)	(0.010)	(0.010)
ln(R&D stock)$_{t-1}$	0.005	0.007	-0.001		0.076	0.055	0.011
	(0.003)	(0.005)	(0.007)		(0.005)	(0.007)	(0.011)
ln(Sales)$_{t-1}$	-0.002	0.006	-0.002		0.012	0.127	0.120
	(0.002)	(0.004)	(0.005)		(0.004)	(0.008)	(0.009)
ln(Total employment at state level)$_{t-1}$					-0.001	0.021	0.030
					(0.009)	(0.023)	(0.023)
ln(1+Publication stock)$_{t-2}$				0.042			
				(0.004)			
ln(Patent stock)$_{t-2}$				-0.014			
				(0.002)			
ln(R&D stock)$_{t-2}$				0.004			
				(0.003)			
ln(Sales)$_{t-2}$				-0.012			
				(0.003)			
Firm fixed-effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	-	-	-	-	-	-
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weak identification(Kleibergen-Paap)					F=72		
Dependent variable sample average	0.08	0.08	0.19	0.14	0.65	12	12
Number of firms	2,259	2,259	2,259	2,199	2,199	2,199	2,199
Observations	23,466	23,466	23,466	22,226	22,226	22,226	22,226
R-squared	0.36	0.54	0.61	0.60	0.64	0.87	-0.01

Notes: for this table the sample is conditional on at least one publication stock over the sample period, 1980-2006. Columns 1-3 present OLS estimation results for the relationship between inventor-author overlap and citation to science. Dummy for internal (external) citation is equal to one if the firm receives at least one internal (external) citation at the focal year to any of its publication published up to the focal year. Inventor-author overlap is measured by the share of patents per year with inventors who include at least one author of a publication published by the same firm in the three-year window prior to the patent's grant year. IDD-Mobility is equal to the status of the Inevitable Disclosure Doctrine (IDD) per the focal firm's state-year (i.e., for effective IDD equals to one) multiplied by a mobility risk measure that is based on the number of publishing firms within 100 miles in the same industry. Specifications include a dummy variable that receives the value of one for firms with no patents at the focal year. Columns 5-7 present Two-Stage Least Squares estimation results for the effect of patent citations to own publications on the number of future publications, using IDD-Mobility as an instrumental variable. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

	(1)	(2)	(3)	(4)	(5)
Dependent variable: ln(1+Number of publications)					
		All cites		5 years citing lag	
	All patents	Corporate patents	Citation received, by SEGMENT and TECH	Citation received, by SEGMENT and TECH	Publishing firms only
ln(1+Internal citation to publications)$_{t-1}$	0.121	0.117	0.118	0.111	0.100
	(0.027)	(0.026)	(0.014)	(0.016)	(0.015)
ln(1+External citation to own publications)$_{t-1}$	-0.018	-0.013			
	(0.027)	(0.026)			
ln(1+External citation to own publications, SEGMENT)$_{t-1}$			-0.047	-0.080	-0.106
			(0.030)	(0.040)	(0.038)
ln(1+External citation to own publications, TECH)$_{t-1}$			0.015	0.135	0.144
			(0.022)	(0.025)	(0.024)
ln(R&D stock)$_{t-1}$	0.065	0.065	0.065	0.064	0.095
	(0.009)	(0.009)	(0.004)	(0.004)	(0.006)
ln(1+Patent stock)$_{t-1}$	0.074	0.074	0.075	0.073	0.085
	(0.011)	(0.011)	(0.005)	(0.005)	(0.007)
ln(Sales)$_{t-1}$	0.047	0.047	0.047	0.047	0.088
	(0.006)	(0.006)	(0.003)	(0.003)	(0.005)
Difference of coefficients (significance level)	0.000	0.000	0.213	0.001	0.000
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes
Dependent variable sample average	5.2	5.2	5.2	5.2	9.1
Number of firms	4,634	4,634	4,634	4,634	2,380
Observations	53,029	53,029	53,029	53,029	30,510
R-squared	0.87	0.87	0.87	0.87	0.85

Notes: This table presents OLS estimation results for the relationship between external and internal patent citation to own publication and future annual publications, for the period 1980-2006. All the citation variables include patent citations up to year t-1 to publications published up to the same year. SEGMENT and TECH measure the product market proximity and the technology market proximity, respectively. In Column 1, external citation include corporate and non-corporate patent citations. Columns 2-5 include only external citation from corporate patents. Columns 4-5 include citations to publications published no earlier than five years prior to the citing patent. All specifications include dummy variable that receives the value of one for firms that never published up to the focal year. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

Table 9. Internal Use and Patent Production

	(1)	(2)	(3)
		Interaction Within	Publishing firms
	Within firms	firms	only
Dependent variable: ln(1+Number of citation-weighted patents)			
Share of internal citation$_{t-1}$	0.198	-0.603	-0.509
	(0.047)	(0.168)	(0.155)
ln(R&D stock)$_{t-1}$ × Share of internal citation$_{t-1}$		0.154	0.131
		(0.031)	(0.029)
ln(R&D stock)$_{t-1}$	0.166	0.165	0.256
	(0.006)	(0.006)	(0.013)
ln(1+Publication stock)$_{t-1}$	0.137	0.135	0.135
	(0.008)	(0.008)	(0.013)
Firm fixed-effects	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes
Std. Share of internal citation	0.093	0.093	0.138
Number of firms	4,634	4,634	2,259
Observations	53,029	53,029	23,466
R-squared	0.86	0.86	0.87

Notes: This table presents results OLS estimation results of a patent equation, for the period 1980-2006. Patents are weighted by citations. Share of internal citations is defined as ratio of citations the firm's publications receive from own patents to citations received from all patents. All specifications include a dummy variable that receives the value of one for firm-years without patents and a dummy variable that receives the value of one for firm-years without citations. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

Table 10. Stock Market Value and Use

	(1)	(2)	(3)
Dependent variable: ln(Market value)			
	Pub stock	Pub stock weighted by inter & exter citation	Exter-cited pub stock weighted by SEGMENT and TECH
ln(1+Publication stock)$_{t-1}$	0.083		
	(0.014)		
ln(1+Publication stock weighted by internal citation)$_{t-1}$		0.053	0.075
		(0.017)	(0.018)
ln(1+Publication stock weighted by external citation)$_{t-1}$		0.031	
		(0.013)	
ln(1+Publication stock weighted by external citation, SEGMENT)$_{t-1}$			-0.081
			(0.035)
ln(1+Publication stock weighted by external citation, TECH)$_{t-1}$			0.069
			(0.029)
ln(1+Patent stock)t-1	-0.004	0.001	0.005
	(0.012)	(0.012)	(0.012)
ln(R&D stock)t-1	0.066	0.070	0.071
	(0.012)	(0.012)	(0.012)
ln(Sales)t-1	0.008	0.009	0.009
	(0.007)	(0.007)	(0.007)
ln(Assets)t-1	0.201	0.201	0.201
	(0.010)	(0.010)	(0.010)
Difference of coefficients (significance level)			0.049
Firm fixed-effects	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes
Dependent variable sample average	2,080	2,080	2,080
Number of firms	4,223	4,223	4,223
Observations	39,734	39,734	39,734
R-squared	0.81	0.81	0.81

Notes: This table presents OLS estimation results for the relationship between citation-weighted publication stock and stock market value, for the period 1980-2006. In Column 2, publication stock is weighted by its internal and external citations from patents granted up to 2014. In Column 3, external cited publications are weighted by product market proximity (SEGMENT) and the technology market proximity (TECH), between the citing patent and cited publication. Publication stock includes WoS publications of our sample firms, published over the sample period (1980-2006). Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

Table A1. Main Variables Definition

Variable	Description	Data Source
Publications count	Publication count for firm I in year t, including all publications with at least one author employed by the focal firm.	Web of Science articles, covered in "Science Citation Index" and "Conference Proceedings Citation Index-Science", 1980-2006
Publication stock	Publication stock in year t for firm I is calculated by: $Publication\_stock_t = Pub_t + Publications\_stock_{t-1}$, where Pub_t is the focal firm's publication count in year t.	Web of Science
Patent count	Patent count in year t for firm i	NBER 2006 patent data project
Patent Stock	Patent stock in year t for firm i is calculated by: $Patent\_stock_t = Patent_t + Patent\_stock_{t-1}$, where $Patent_t$ is the focal firm's patent count in year t.	NBER 2006 patent data project
Internal citation to publications	Annual flow of internal patent citations to firm's i own publications	PatStat database and citation match for patents granted at the focal year and publications published from 1980 until the focal year.
External citations to firm's own publications	Annual flow of external patent citations to firm's i publications. Includes citations by corporate and non-corporate patents.	PatStat database and citation match for patents granted at the focal year and publications published from 1980 until the focal year.
External citations to firm's own publications, SEGMENT	Annual flow of external patent citations to firm's i publications, weighted by product market proximity of the citing and cited firms. Product market proximity is computed based on each firm's sales share distribution across line of business listed within the Compustat operating segments database.	Compustat operating segments database, PatStat database and citation match for patents granted at the focal year and publications published from 1980 until the focal year.
External citations to firm's own publications, TECH	Annual flow of external patent citations to firm's i publications, weighted by technology market proximity of the citing and cited firms. Technology market proximity is computed based on each firm's patent share distribution across different technology fields.	PatStat database and citation match for patents granted at the focal year and publications published from 1980 until the focal year.
Share of internal citations	Share of internal citations to science is defined as ratio of internal-citations from own patents to internal and external citations received by corporate and non-corporate patents, per year.	PatStat database and citation match for patents granted at the focal year and publications published from 1980 until the focal year.
Inventor-author overlap	The share of patents per year with inventors who include at least one author of a publication published by the same firm in the three-year window prior to the patent's grant year.	Web of Science, NBER 2006, HBS Patent Inventor Database. Including all non-collaborative patents and publications related to our sample firms during the sample period (1980-2006),
Market value	Following Griliches (1981), market value per firm-year is defined as the sum of the values of common stock, preferred stock, and total debt net of current assets. *Tobin's-Q* is defined as the ratio of market value to assets.	U.S. Compustat
R&D stock	R&D stock per firm-year is calculated using a perpetual inventory method with a 15 percent depreciation rate (Hall et al., 2005), such that the R&D stock, GRD, in year t is $GRD_t = R_t + (1-\delta)GRD_{t-1}$ where R_t is the focal firm's R&D expenditure in year t based on Compustat data and $\delta=0.15$.	U.S. Compustat
Assets	The book value of capital includes net plant, property and equipment, inventories, investments in unconsolidated subsidiaries, and intangibles other than R&D.	U.S. Compustat

Table A2. SIC Classification by Main Industries

Category	Description	Related 4-digit sic codes in our sample of firms
Telecommunication	Telecom, Communication- system, equipment, services	3661 3663 3669 4812 4813 4822 4832 4833 4841 4899
IT & Software	IT & Software - Development, Provider, Sale & Services	5040 5045 5734 7370 7371 7372 7373 7374
Machinery/equipment/system	Manufacture /sale/ rent - Machinery, Systems, Equipment, Instruments, Components and Tools not elsewhere included (e.g., medical, lab, heating, transportation, construction, measurement)	3420 3430 3433 3510 3523 3524 3530 3531 3532 3533 3537 3540 3541 3550 3555 3559 3560 3561 3562 3564 3567 3569 3580 3585 3590 3711 3713 3714 3715 3716 3728 3743 3760 3790 3812 3821 3822 3823 3824 3825 3826 3827 3829 3841 3842 3843 3844 3845 3873 5047 5070 5080 5084 7359
Energy	Electricity, Oil, Gas, Power station- including: utility, exploration, equipment, services, machinery, tools, etc.	1311 1381 1382 1389 1600 1623 1700 2911 2990 4922 4923 5172
Chemicals	Chemicals- Manufacture & Sale	1000 1040 1220 1221 2800 2810 2820 2821 2840 2842 2851 2860 2870 2890 2891 3320 3330 3334 3341 3350 3357 3360 3390 5160
Electronics & Semiconductor	Electronic products and equipments including components; semiconductor; computers including system and components -Manufacture, Sale & Rent	3570 3571 3572 3575 3576 3577 3578 3579 3600 3612 3613 3620 3621 3630 3634 3640 3651 3670 3672 3674 3677 3678 3679 3690 3695 3861 5063 5064 5065
Drugs, Pharmaceuticals and Biotechnology	Drugs, pharmaceuticals & biotech- Manufacture, Sale & Services	2833 2834 2835 2836 5122 5912 8731

Table A3. Corporate patents that cite science *vs.* do not cite science, 1980-2006

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
			Patents that cite science			Patents that cite internal science			Patents that do not cite science		
Variable	(4) minus (10)	(7) minus (10)	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
Forward citations	5.6**	4.7**	32,310	25.1	45	10,460	24.3	51	656,164	19.6	33
Core patents	0.06**	0.02**	32,310	0.46	0.5	10,460	0.43	0.5	656,164	0.40	0.5

Notes: This table presents mean comparison tests for corporate patents that cite vs. do not cite (internal) science. The sample includes all patents related to our sample firms granted over the period 1980-2006. Patent that cite internal science include patents with at least one citation to the firm's own publications.** denotes that the difference in means is significant at the 1 percent level.

Table A4. Internal Use and Publication Output- Different Lags

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: ln(1+number of publications)						
	OLS			IV IDD (Sample: Publishing Firms)		
	Lag 1	Lag 2	Lag 3	Lag 1	Lag 2	Lag 3
ln(1+Internal citation to publications)$_{t-x}$	0.110	0.080	0.063	0.670	0.521	0.480
	(0.026)	(0.013)	(0.013)	(0.104)	(0.108)	(0.112)
ln(R&D stock)$_{t-x}$	0.065	0.062	0.058	0.011	0.095	0.071
	(0.009)	(0.004)	(0.004)	(0.011)	(0.011)	(0.012)
ln(1+Patent stock)$_{t-x}$	0.073	0.056	0.044	0.116	0.001	-0.004
	(0.011)	(0.005)	(0.006)	(0.010)	(0.012)	(0.013)
ln(Sales)$_{t-x}$	0.047	0.054	0.054	0.120	0.130	0.117
	(0.006)	(0.003)	(0.003)	(0.009)	(0.010)	(0.010)
ln(Total employment at state level)$_{t-x}$				0.030	0.033	0.032
				(0.023)	(0.025)	(0.028)
Weak identification(Kleibergen-Paap)				F=72	F=70	F=63
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable sample average	5	6	6	12	13	14
Number of firms	4,634	4,430	4,104	2,199	1,938	1,717
Observations	53,029	48,395	43,965	22,226	17,969	15,124
R-squared	0.87	0.87	0.88	-0.01	-0.01	-0.03

Notes: The table presents OLS estimation result (Columns 1-3) and Two-Stage Least Squares estimation results (Columns 4-6) for the relationship between past internal patent citations to own publications and future annual publications, for the period 1980-2006, for different lags of right hand side variables. Internal citations to own publications include patent citations up to year t-1 to publications published up to the same year. For the IV estimation, the sample is conditioned on at least one publication stock. The endogenous variable, internal citation to publications, is instrumented by the IDD-Mobility measure. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

Table A5. Instrumental Variable Estimation II: Changes in Foreign Exchange Rates (Sample: Publishing Firms)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS				IV: Devaluation			IVs: IDD and Devaluation		
Dependent variable:	Profits (EBIDTA)	Avg. NPL citations per patent		Share of patents with NPL>0	ln(1+Internal citation)$_{t-1}$	ln(1+Number of publications)		ln(1+Internal citation)$_{t-1}$	ln(1+Number of publications)	
	Publishing firms	Publishing firms	Exc. zero patents	Publishing firms	First Stage	OLS	IV	First Stage	OLS	IV
ln(Internal citation to publications)$_{t-1}$						0.886	1.408		0.885	1.434
						(0.040)	(0.232)		(0.041)	(0.223)
Devaluation dummy$_{t-2}$	-33.266	-0.097	-0.125	-0.044	-0.081			-0.079		
	(11.577)	(0.031)	(0.043)	(0.010)	(0.010)			(0.011)		
Inevitable disclosure doctrine dummy$_{t-2}$								0.023		
								(0.008)		
ln(R&D stock)$_{t-1}$					0.068	0.244	0.209	0.067	0.241	0.205
					(0.003)	(0.019)	(0.017)	(0.003)	(0.019)	(0.016)
ln(1+Patent stock)$_{t-1}$					0.093	0.066	0.019	0.093	0.067	0.017
					(0.004)	(0.016)	(0.022)	(0.004)	(0.016)	(0.021)
ln(Assets)$_{t-1}$	88.434	-0.074	-0.100	-0.028	-0.020	-0.006	0.006	-0.019	-0.004	0.007
	(9.969)	(0.007)	(0.009)	(0.002)	(0.002)	(0.010)	(0.007)	(0.002)	(0.010)	(0.006)
Exchange rate level$_{t}$					-0.001	0.001	0.001	-0.001	0.001	0.001
					(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ln(Total employment)$_{t-1}$								0.020	0.031	0.021
								(0.004)	(0.021)	(0.010)
Firm fixed-effects	Yes	No	No	No	No	No	No	No	No	No
Industry dummies	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weak identification(Kleibergen-Paap)					F= 59			F=29.168 > Stock-Yogo CV 5%= 13.46		
Overidentification (Hansen test)								t-statistic=0.59		
Dependent variable sample average:	403	5	7	0.36	0.83	11	11	0.83	11	11
Number of firms	1901	1901	1751	1901	1901	1901	1901	1901	1901	1901
Observations	14,565	14,565	10,883	14,565	14,565	14,565	14,565	14,565	14,565	14,565
R-squared	0.91	0.32	0.10	0.39	0.28	0.54	0.50	0.28	0.54	0.50

Notes: This table presents instrumental variable estimation results for the effect of patent citations to own science on firm's future publication, for the period 1990-2006. The sample is conditional on at least one publication stock. The endogenous variable, internal citation in year t-1, is instrumented at the firm-year level by a devaluation dummy that is based on weighted-changes in exchange rates in countries where the firm has subsidiaries. Profit is measured by EBIDTA. NPL citations are cites in year t by the focal firm's patents to any article. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

Table A6. Publications and Citations to Science, by Industry

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	ln(1+No. of Publications)						
	Electronics & Semiconductor	Pharma&Biotech	Chemicals	Energy	IT & Software	Telecom	Machinery / equipment / system
ln(1+*Internal citation to publications*)$_{t-1}$	0.096	0.073	0.123	0.054	0.186	0.120	0.087
	(0.034)	(0.025)	(0.052)	(0.051)	(0.081)	(0.055)	(0.042)
ln(1+ *External citation to own publications*)$_{t-1}$	-0.081	0.069	-0.077	-0.126	-0.132	-0.120	-0.134
	(0.026)	(0.020)	(0.034)	(0.044)	(0.059)	(0.038)	(0.034)
ln(*R&D stock*)$_{t-1}$	0.074	0.184	0.153	-0.026	0.036	0.061	0.038
	(0.010)	(0.014)	(0.018)	(0.040)	(0.011)	(0.012)	(0.009)
ln(*1+Patent stock*)$_{t-1}$	0.106	0.059	0.047	0.147	0.085	0.103	0.095
	(0.011)	(0.017)	(0.022)	(0.048)	(0.019)	(0.017)	(0.010)
ln(*Sales*)$_{t-1}$	0.040	0.071	0.064	0.027	0.047	-0.003	0.051
	(0.006)	(0.010)	(0.016)	(0.017)	(0.007)	(0.007)	(0.007)
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9,229	6,246	2,871	1,065	4,978	3,122	13,600
R-squared	0.86	0.88	0.89	0.94	0.86	0.94	0.82

Notes: This table presents OLS estimation results for the relationship between internal and external citations to firm's own publication and future annual publications, by industry, for the period 1980-2006. Industry classification is based on four-digit main SIC code. Citation variables include patent citations up to year t-1 to publications published up to the same year. External cites to own publications include corporate and non-corporate patent citations. All specifications include dummy variable that receives the value of one for firms that never published up to the focal year. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

Table A7. IDD Effective *vs.* Non-Effective: Patents to R&D, Per Firm-Year

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Effective-IDD			Non-Effective IDD	
Variable	(3) minus (6)	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
MA (1994)							
Patents to R&D	-0.376	2,322	0.5	7.2	1,596	0.9	5.5
TX (1993)							
Patents to R&D	0.037	1,422	0.8	4.6	1,408	0.7	3.6
NJ (1987)							
Patents to R&D	-0.947	2,278	0.8	10.8	717	1.7	13.6
IL (1989)							
Patents to R&D	-0.484	1,477	0.9	12.0	780	1.3	9.8
MN (1986)							
Patents to R&D	-0.212	1,707	0.8	4.9	363	1.0	3.6

Notes: This table presents mean comparison tests for Patent-count to R&D for firm-years before and after the recognition of the Inevitable Disclosure Doctrine (IDD) in main effective states, over the sample period (1980-2006). The table includes only affected states with above 2000 firm-year sample observations. For this analysis, the firm's state is based on the HQ of each firm.

Figure 1. Patent citations to science are related to Carnegie Mellon survey response on the importance of research

Figure A. Patent citations to Web of Science articles, all publications
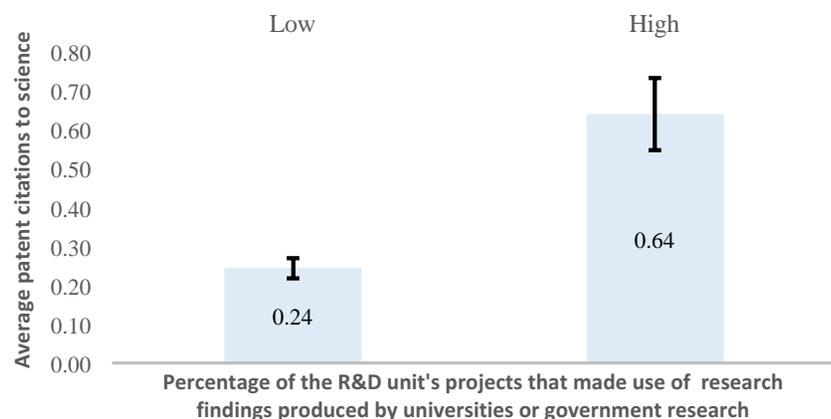
Figure B. Patent citations to Web of Science articles by top 200 U.S. universities
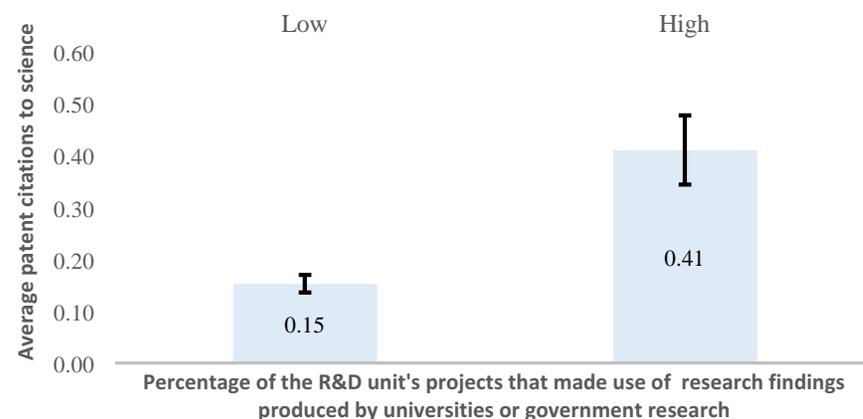
Figure C. Patent citations to Web of Science articles in main research field
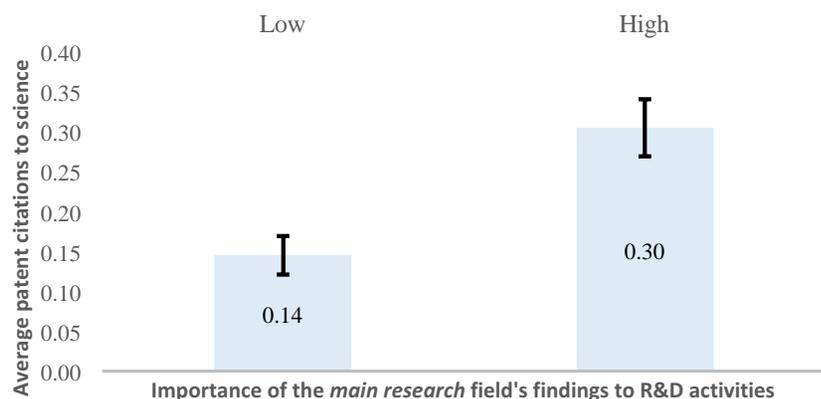
Note: figures A & B present mean comparisons for average patent citation to publications by low and high percentage of the R&D unit's projects that made use of research findings produced by universities or government. Based on 1994 Carnegie Mellon survey data (Cohen et al., 2000) Q.18: "During the last 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 : a. Research findings". *High and low* are defined by the mid category rank in the survey- Low -636 firms; High-110 firms. Figure C presents mean comparisons for average patent citation to publications by low and high importance of the *main research* field's findings to the firm's R&D activities. Based on CMS Q.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. Then, indicate the importance of that field's findings to your R&D activities". The sample is restricted to firms that indicated their main field in Q22 as A-J (excluding category K -'others'). Publications were classified to main fields by their subject category. High is defined by the top rank and low by the lowest 3 ranks in the survey. Low -354 firms; High-308 firms. The lines represent standard error bars. The sample includes only patenting firms. Citations are restricted to publications published no earlier than five years prior to the citing patent. Related patents were granted between 1991 to 1999.

Figure A1. Patent citations to science are related to Carnegie Mellon survey response on the importance of research, by main fields

Figure A1. Patent citations to science are related to Carnegie Mellon survey response on the importance of research, by main fields.

Pharmaceuticals / Biotech / Chemicals

Figure A. Patent citations to Web of Science articles by public science
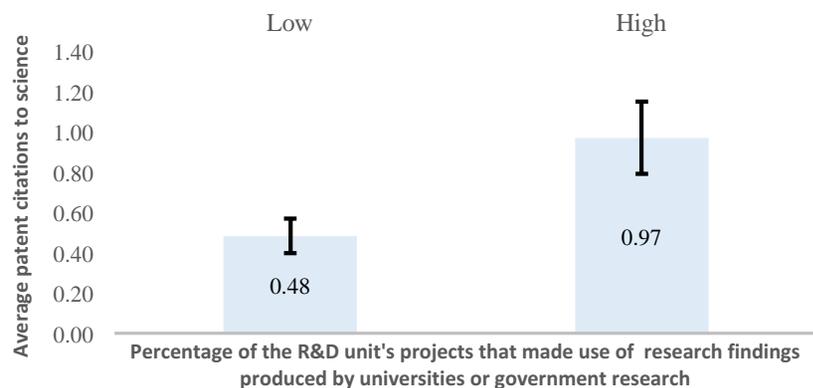
Figure B. Patent citations to Web of Science articles by top 200 U.S. universities
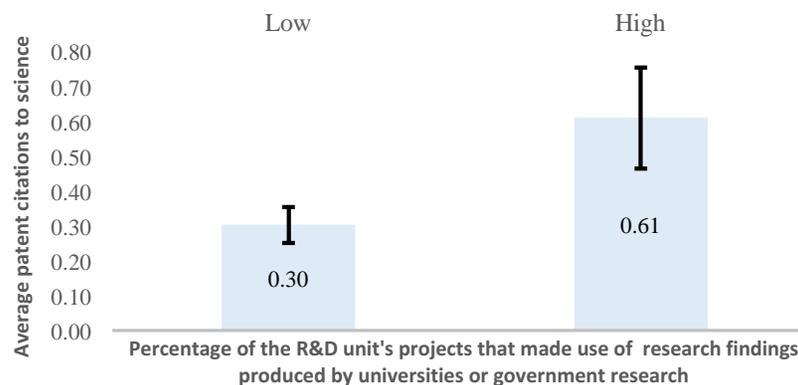
Electronics & Semiconductor / IT & software / Telecom / Machinery

Figure C. Patent citations to Web of Science articles by public science
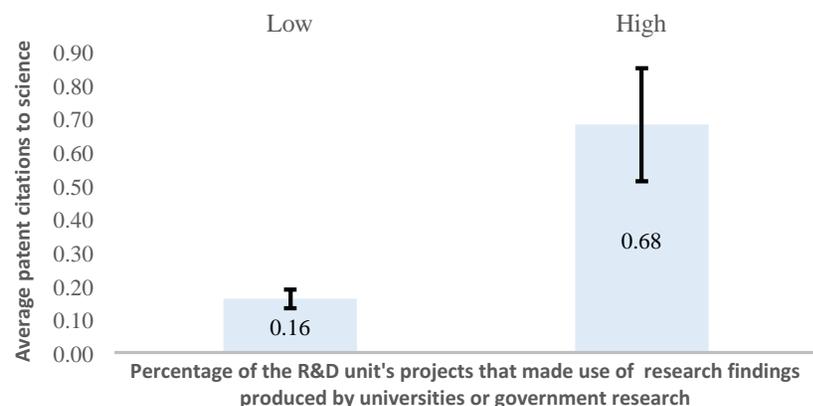
Figure D. Patent citations to Web of Science articles by top 200 U.S. universities
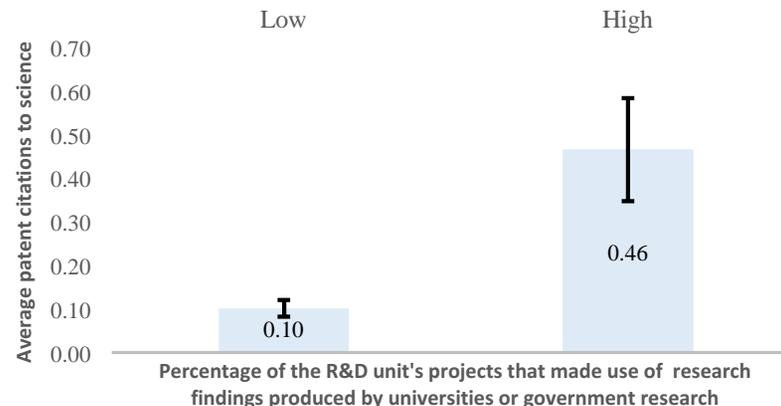
Note: the figures present mean comparisons for average patent citation to publications by low and high percentage of the R&D unit's projects that made use of research findings produced by universities or government. Based on 1994 Carnegie Mellon survey data (Cohen et al., 2000) Q.18: "During the last 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 : a. Research findings". *High and low* are defined by the mid category rank in the survey. Figure A & B: Low -131 firms; High-31 firms. Figures C & D: Low -293 firms; High-41 firms. The lines represent standard error bars. The sample includes only patenting firms. Citations are restricted to publications published no earlier than five years prior to the citing patent. Related patents were granted between 1991 to 1999.

Figure A2. Citations to science are positively related to share of scientists from CMS

Figure A. Patent citations to Web of Science articles

Figure B. Patent citations to Web of Science articles by top 200 U.S. universities

Figure C. Patent citations to Web of Science articles by corporations (Compustat)
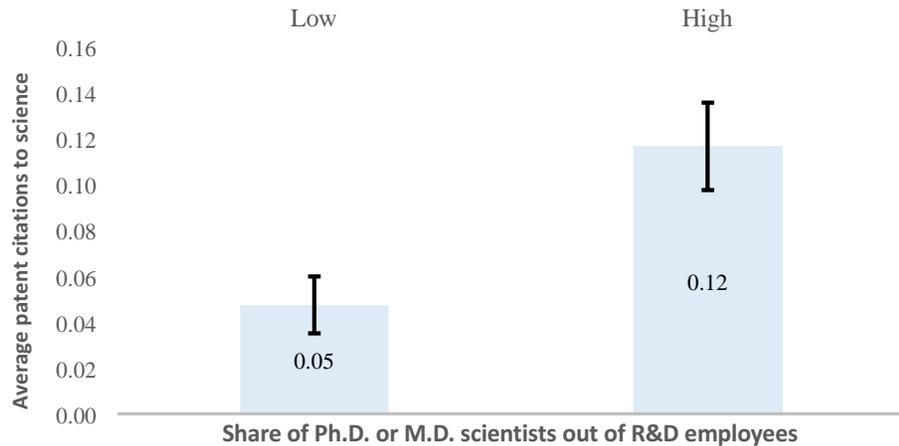
Note: The figures present mean comparisons for average patent citation to publications by low and high share of Ph.D. or M.D. scientists out of R&D employees. Based on 1994 Carnegie Mellon survey data (Cohen et al., 2000) Q.53-54: "Approximately how many professional and technical R&D employees does your firm have working in your focus industry (include all facilities working in that industry)? Of the above number, approximately how many are Ph.D. or M.D. scientists?". *High and low* are defined by above and below the median value Share of Ph.D. or M.D. scientists, respectively Low -350 firms; High-358 firms. The sample includes only patenting firms. Citations are restricted to publications published no earlier than five years prior to the citing patent. The lines represent standard error bars.

Figure A3 : Cumulative distribution for the relationship between citations to science and CMS

Figure A. Patent citations to Web of Science articles by public science
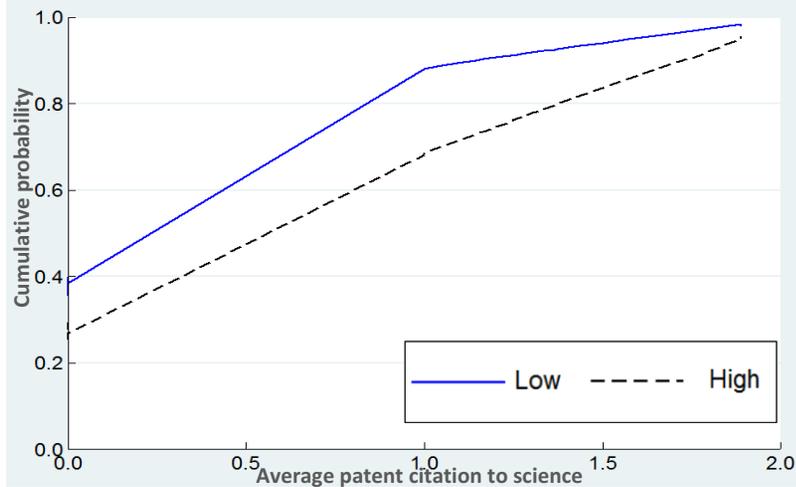
Figure B. Patent citations to Web of Science articles by top 200 U.S. universities
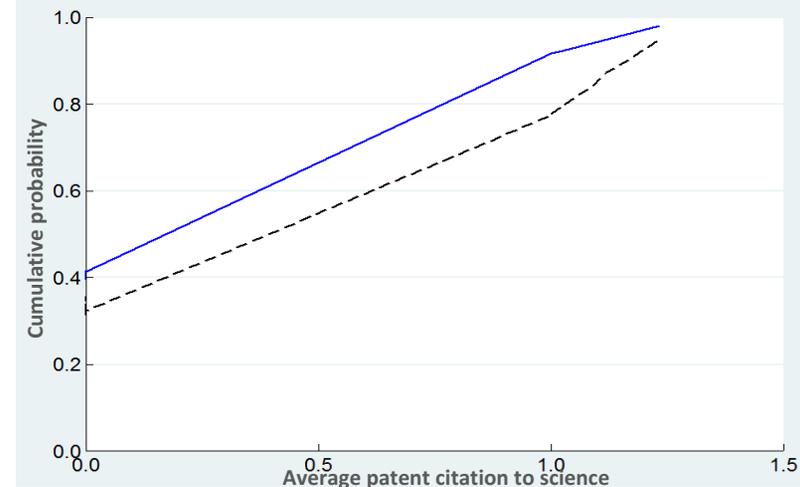
Figure C. Patent citations to Web of Science articles in main research field
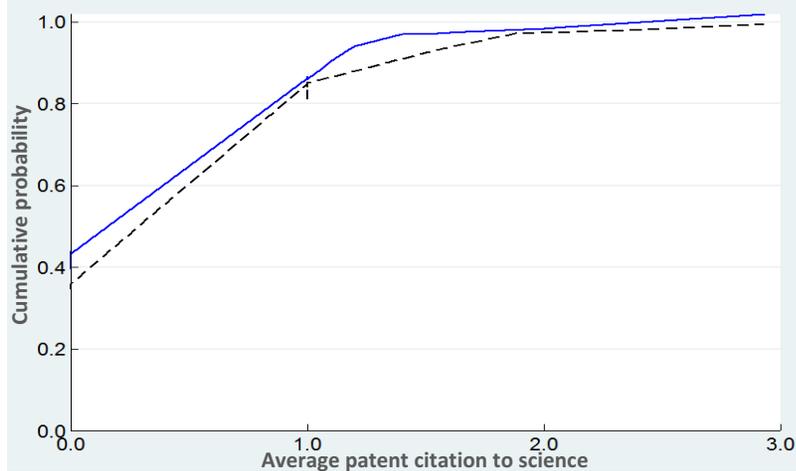
Note: Figures A & B plot the cumulative distribution of patent citations to science by low and high percentage of the R&D unit's projects that made use of research findings produced by universities or government.. Based on 1994 Carnegie Mellon survey data (Cohen et al., 2000) Q.18: "During the last three 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 : a. Research findings". High and low are defined by the mid category rank in the survey: Low -636 firms; High-110 firms. Number of patent citations per patent is presented with a proximity value in the 95th percentile of the sample. Figure C plots the cumulative distribution of patent citations to science by low and high importance of the *main research* field's findings to the firm's R&D activities. Based on CMS Q.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. Then, indicate the importance of that field's findings to your R&D activities". The sample is restricted to firms that indicated their main field in Q22 as A-J (excluding category K -'others'). Pubs were classified to main fields by their subject category. High is defined by the top rank and low by the lowest 3 ranks in the survey. Low -354 firms; High-308 firms. Number of patent citations per patent is presented with a proximity value in the 99th percentile of the sample. The sample includes only patenting firms. Citations are restricted to publications published no earlier than five years prior to the citing patent.

Figure A3 : Cumulative distribution for the relationship between citations to science and CMS

Figure A. Patent citations to Web of Science articles
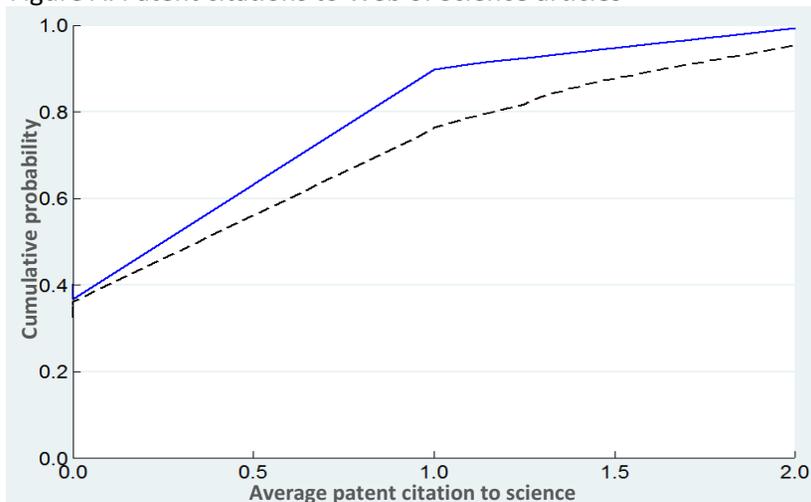
Figure B. Patent citations to Web of Science articles by top 200 U.S. universities
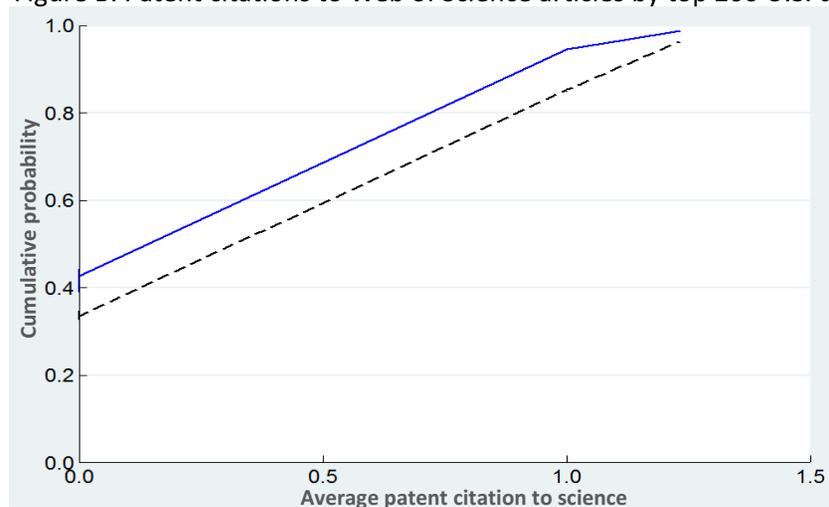
Figure C. Patent citations to Web of Science articles by corporations (Compustat)
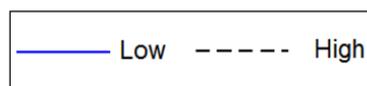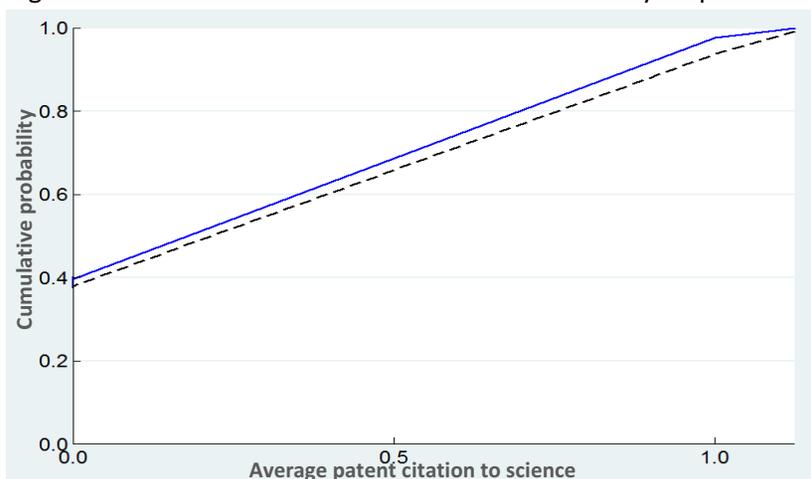
Note: The figures plot the cumulative distribution of patent citations to science, by low and high share of Ph.D. or M.D. scientists out of R&D employees. Based on 1994 Carnegie Mellon survey data (Cohen et al., 2000) Q.53-54: "Approximately how many professional and technical R&D employees does your firm have working in your focus industry (include all facilities working in that industry)? Of the above number, approximately how many are Ph.D. or M.D. scientists?". *High and low* are defined by above and below the median value Share of Ph.D. or M.D. scientists, respectively Low - 350 firms; High-358 firms. Number of patent citations per patent is presented with a proximity value in the 95th percentile of the sample. The sample includes only patenting firms. Citations are restricted to publications published no earlier than five years prior to the citing patent.

Figure A5. External and internal citation, matching process

(i) Example of an external citation to IBM's publication : the patent owner and cited corporate publication are different

(ii) Example of an internal citation to IBM's publication : the patent owner and cited corporate publication are the same

Figure A6. Timeline- Production and Use of Research
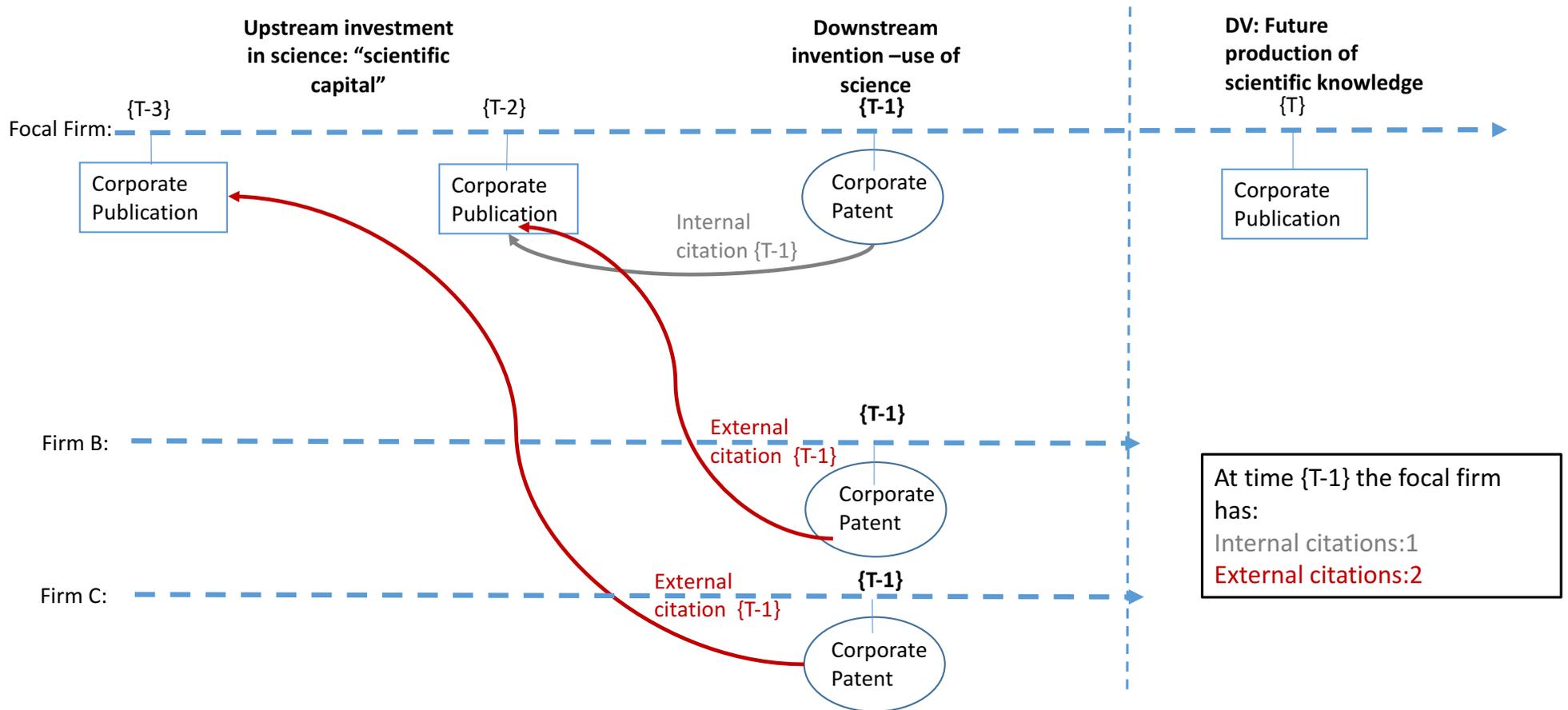
Appendix B: Supplement Tables and figures

Table B1. Challenges in assigning a unique company id over time

Name	GVKEY	CUSIP	begyr	endyr	comment
GENZYME CORP-CONSOLIDATED	119053	63799A936	1988	2002	Old CUSIP
GENZYME CORP	12233	372917104	1996	2006	New CUSIP
GENZYME TISSUE REPAIR	118653	372917401	1996	1999	Related subsidiary CUSIP
GENZYME SURGICAL PRODUCTS	121742	372917609	1997	1999	Related subsidiary CUSIP
GENZYME MOLECULAR ONCOLOGY	117298	372917500	1997	2002	Related subsidiary CUSIP
GENZYME BIOSURGERY	143176	372917708	1999	2002	Related subsidiary CUSIP

Table B2. Dynamic match example

UO Company	GVKEY1	Company1	begyr1	endyr1	GVKEY2	Company2	begyr2	endyr2	GVKEY3	Company3	begyr3	endyr3	GVKEY4	Company4	begyr4	endyr4
LORAL CORP	6807	LORAL CORP	1960	1993	62640	LORAL SPACE & COMMUN	1994	2006								
CELANESE CORP	2827	CELANESE CORP-OLD	1950	1985	13934	HOECHST CELANESE CORP	1987	1996	125434	CELANESE AG	1998	2002	162254	CELANESE CORP	2003	2006

Table B3. Most frequent abbreviated words

ADV	AEROSP	AGR	AMER	ANAL	ANALYT	ANIM	APPL	APPLICAT
ASSOC	AUTOMAT	BIOL	BIOMED	BIOPHARM	BIOSCI	BIOSURG	BIOSYS	BIOTECH
BIOTHERAPEUT	CHEM	CLIN	COMMUN	COMP	CORP	CTR	DEV	DIAGNOST
DYNAM	EDUC	ELECTR	ENGN	ENVIRONM	FAVORS	GEN	GENET	GRAPH
GRP	HLDG	HLTHCR	HOSP	INC	IND	INFO	INNOVAT	INST
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

Table B4. Matching Citations to Publications - Examples

Citation	Publication info			Comment
	Title	Authors	Journal information	
LIN, KUN SHAN, ET AL., SOFTWARE RULES ***GIVES*** PERSONAL COMPUTER REAL WORD POWER , **INTERNATIONAL ELECTRONICS**, VOL. **53,** NO. 3, FEB. 10, **1981**, PP. **122 125**.	"SOFTWARE RULES ***GIVE*** PERSONAL-COMPUTER REAL WORD POWER"	LIN KS, FRANTZ GA, GOUDIE K	**ELECTRONICS 54** (3): 122-125 1981	Typo in title and journal Vol.; initials vs. full name
U. WACHSMANN, R. F. H. FISCHER AND J.B. HUBER, MULTILEVEL CODES: THEORETICAL CONCEPTS AND PRACTICAL DESIGN RULES, **IEEE TRANS INFORM. THEORY**, VOL. 45, NO. 5, PP. **1361-1391**, JUL. **1999**.	"MULTILEVEL CODES: THEORETICAL CONCEPTS AND PRACTICAL DESIGN RULES"	WACHSMANN U, FISCHER RFH, HUBER JB	**IEEE TRANSACTIONS ON INFORMATION THEORY** 45 (5): 1361-1391 JUL 1999	Several names listed; variation in journal name
DESIGN CHARACTERISTICS OF GAS JET GENERATORS, **BORISOV**, **1979**, PP. 21 25.	"DESIGN CHARACTERISTICS OF GAS-JET GENERATORS"	BORISOV YY	SOVIET PHYSICS ACOUSTICS-USSR 26 (1): 21-25 **1980**	Typo in year; diff in location of title within the citation
KERNS, SHERRA E., THE DESIGN OF RADIATION HARDENED ICS FOR SPACE: A COMPENDIUM OF APPROACHES, PROCEEDINGS OF THE IEEE, NOV. 1988, PP. 1470 1509.	"THE DESIGN OF RADIATION-HARDENED ICS FOR SPACE - A COMPENDIUM OF APPROACHES"	**KERNS SE**, SHAFER BD, ROCKETT LR, PRIDMORE JS, BERNDT DF, VANVONNO N, BARBER FE	PROCEEDINGS OF THE IEEE 76 (11): 1470-1509 NOV 1988	Several authors w/o "et al."
GENESTIER ET AL (BLOOD, 1997, VOL. 90, PP. 3629-3639).	"FAS-INDEPENDENT APOPTOSIS OF ACTIVATED T CELLS INDUCED BY ANTIBODIES TO THE HLA CLASS I ALPHA 1 DOMAIN"	GENESTIER L, PAILLOT R, BONNEFOYBERARD N, MEFFRE G, FLACHER M, FEVRE D, LIU YJ, LEBOUTEILLER P, WALDMANN H, ENGELHARD VH, BANCHEREAU J, REVILLARD JP	BLOOD 90 (9): 3629-3639 NOV 1 1997	No title within citation- however, perfect match in all other features
STEPHEN M. BEBGE, LYLE D. BIGHLEY AND DONALD C. MONKHOUSE PHARMACEUTICAL SALTS JOURNAL OF PHARMACEUTICAL SCIENCES, 1977, 66, 1-19.	"PHARMACEUTICAL SALTS"	**BERGE SM, BIGHLEY LD, MONKHOUSE DC**	JOURNAL OF PHARMACEUTICAL SCIENCES 66 (1): 1-19 1977	Several names listed; variation of names
L. YOUNG AND D. SHEENA, **METHODS & DESIGNS:** SURVEY OF EYE MOVEMENT RECORDING METHODS, BEHAV. RES. METHODS INSTRUM., VOL. 5, PP. 397-429, 1975.	"SURVEY OF EYE-MOVEMENT RECORDING METHODS"	YOUNG LR, SHEENA D	BEHAVIOR RESEARCH METHODS &INSTRUMENTATION 7 (5): 397-429 1975	diff in title
MICROWAVE JOURNAL, VOL. 22, NO. 2, FEB. 1979, DEDAHAM US PP. 51 52, **H. C. CHAPPELL.**	"DESIGNING IMPEDANCE MATCHED IN-PHASE POWER DIVIDERS"	CHAPPELL HC	MICROWAVE JOURNAL 22 (2): 51-52 1979	no title - however, perfect match in all other features; diff position of author's name within citation

Table B5. Inevitable Disclosure Doctrine (IDD)- Effective Dates by State for IV Estimation

State	Firm-year obs. in sample	Year IDD became effective in state	State	Firm-year obs. in sample	Year IDD became effective in state
AL	248	-	MT	39	-
AR	122	1997	NC	946	1976
AZ	629	-	ND	6	-
CA	13,037	-	NE	145	-
CO	1,391	-	NH	418	-
CT	1,646	1996	NJ	3,489	1987
DC	70	-	NM	66	-
DE	167	1964	NV	326	-
FL	1,783	1960-2001	NY	4,347	1919
GA	1,071	1998	OH	2,151	2000
HI	53	-	OK	236	-
IA	282	1996	OR	583	-
ID	109	-	PA	2,410	1982
IL	2,585	1989	RI	202	-
IN	918	1995	SC	275	-
KS	241	2006	SD	46	-
KY	163	-	TN	315	-
LA	199	-	TX	3,139	1993-2003
MA	4,246	1994	UT	724	1998
MD	1,099	-	VA	983	-
ME	102	-	VT	46	-
MI	1,424	1966-2002	WA	1,058	1997
MN	2,292	1986	WI	1,018	-
MO	745	2000	WV	85	-
MS	87	-	WY	3	-

Note: The firm's state for each year is based on the publication address.

Figure B1- Frequency distribution for inventor-author overlap
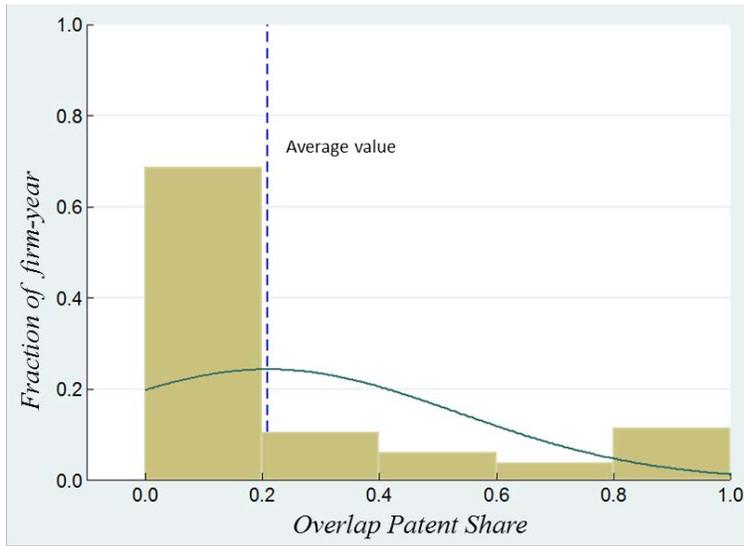
Figure B2- Frequency distribution for inventor-author overlap, by industry
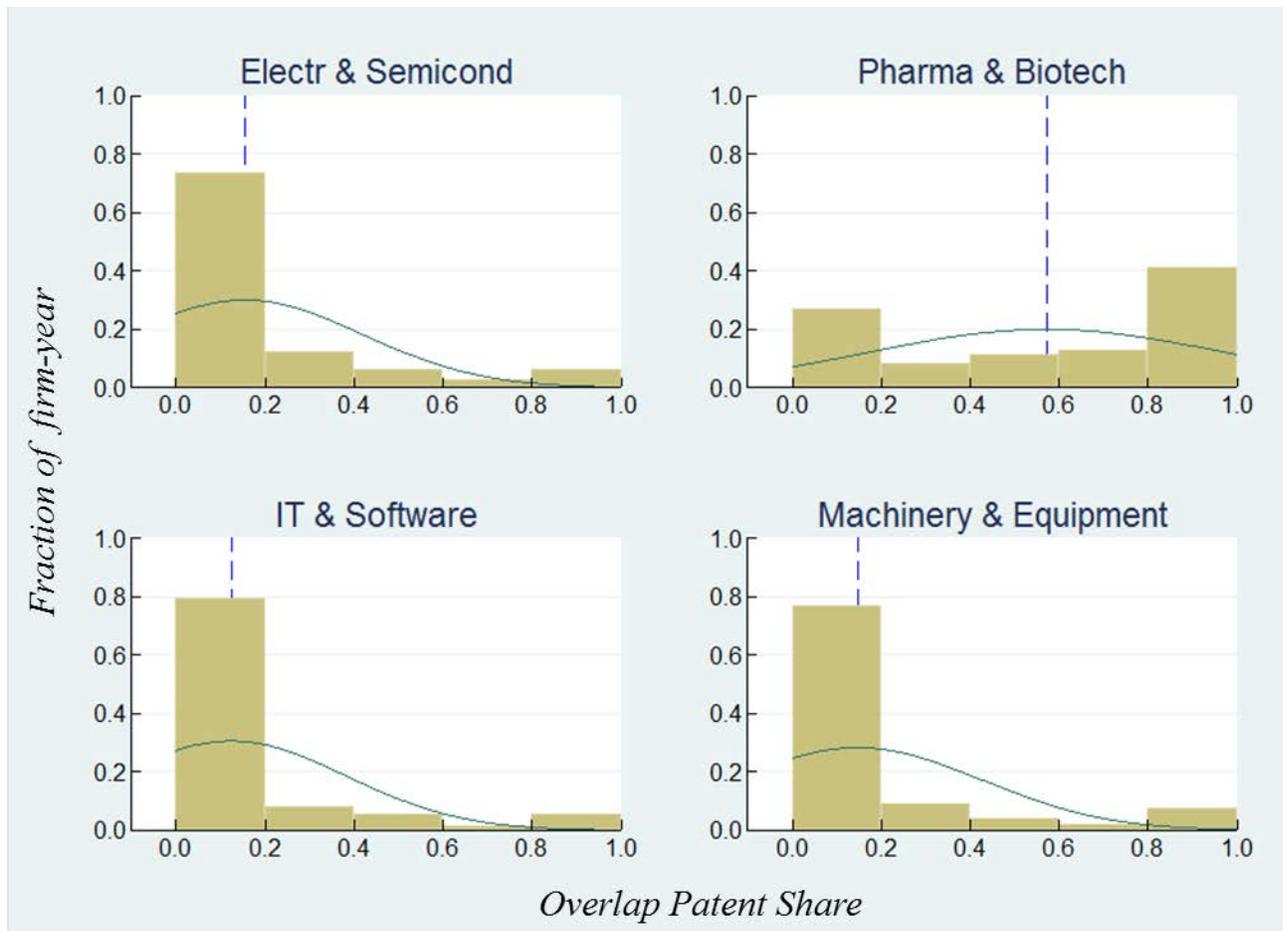
Note: Classification to industries is based on four-digit main SIC code.