Back to Basics: Why do Firms Invest in Scientific Research?*

Ashish Arora† Sharon Belenzon‡ Lia Sheer §

August 1, 2019

Abstract

Large firms appear to be shifting away from research and towards development. To understand this shift, we explore why firms invest in scientific research, especially in the presence of spillovers. Using data on 800,000 scientific publications of 4,090 American firms between 1980 and 2015, and patent citations to these articles, we document the relationship between corporate publications and their use in invention by the sponsoring firm and by rivals. Our analysis implies that corporate research can be both privately valuable and socially useful. Firms produce more publications when they are used internally, but fewer publications when they are used by rivals. Over our sample period spillovers have increased, while internal use has remained stable. Up to 45% of the decline in publication rate can be explained by an increase in the incidence and importance of spillovers to rivals.

*We thank Nick Bloom, Tim Bresnahan, Wesley Cohen, Dietmer Harhoff, David Hounshell, Richard Freeman, Alfonso Gambardella, Shane Greenstein, Brian Lucking, Petra Moser, David Mowery, Mark Schankerman, Scott Stern, John Van Reenen, and seminar participants at NBER summer institute, Harvard, Stanford, Berkeley, Maryland, Munich, and Bocconi for helpful comments. We thank Honggi Lee and JK Suh for excellent research assistance. All remaining errors are ours.

†Duke University, Fuqua School of Business and NBER

‡Duke University, Fuqua School of Business and NBER

§Duke University, Fuqua School of Business

1 Introduction

A defining characteristic of modern economic growth is the systematic application of science to technology (Kuznets, 1971). From synthetic fibers and plastics, modern materials, drugs, and computers and communications, few of the life-changing products invented in the twentieth century have not relied on science. In many cases, the underlying science was produced by firms, which accounted for a significant share of scientific research in the United States. Indeed, in their heyday, corporate labs, such as Bell Labs, produced Nobel prize winning research.

Private business investment in upstream research remains substantial. In 2015, the business sector invested $22 billion in basic research, accounting for 26% of all basic research in the United States.[1] However, over the last three decades, corporations have reduced their engagement in upstream research. NSF data show that the share of basic and applied research in corporate R&D in the United States has declined from about 30 percent in 1985 to less than 20 percent in 2015 (NSF 2017). The same data show that the real expenditure on research in industry, after increasing over the 1980s, barely grew over the 20 year period between 1990 to 2010.[2] Over the same period, corporate investment in downstream innovation activities (the "D" of R&D) has remained robust. Figure 1 presents trends in the annual number of publications ("R") and patent ("D") per firm for our sample firms. The pattern of declining publications and increasing patents suggests that the composition of corporate R&D is changing over time, with less "R" and more "D" (see also Arora et al. (2018); Arora et al. (2019)). The present paper investigates how the production of scientific research by American corporations is related to its use in invention by the focal firm and spillovers captured by rivals. We suggest that private returns to corporate research depend on the balance between two opposing forces: the benefits from the use of science in own downstream inventions, and costs of spillovers to rivals. Xeroxs Palo Alto Research Center (PARC) is a case in point. PARC was one of the most innovative corporate research

[1] The micro-economic literature has historically distinguished between scientific research and development activities (e.g., Nelson (1959)). More recently, macro-economic endogenous growth models have begun to distinguish between these two activities as well (e.g., Akcigit et al. (2017)).

[2] Impressionistic accounts also indicate that many leading American corporations began to withdraw from upstream scientific research in the 1980s (Mowery, 2009). Some corporate labs were shut down and others spun-off as independent entities. Bell Labs was separated from its parent company AT&T and placed under Lucent in 1996; Xerox PARC was spun off into a separate company in 2002. A more recent example is from DuPont. Research at DuPont started in 1903, and DuPont's research accomplishments soon came to rival those of the leading academic chemistry departments. In the 1960s, DuPont researchers published more articles in the *Journal of the American Chemical Society* than MIT and Caltech combined. However, in the 1990s, DuPont's attitude toward research changed, and in 2016 the company's management closed its Central Research and Development Lab.

lab in the 1970s. Failures by Xerox to commercialize PARC inventions are frequently cited as reasons for its ultimate demise. Critics frequently point to the large spillovers from PARC's research to companies such as Apple, Microsoft, and 3Com. Yet, the benefits Xerox obtained from PARC's research in areas that were closer to Xerox's core business, such as the laser printer, were substantial. These inventions, at least for a time, allowed the firm to recoup its investment in PARC, despite the spillovers it created.

Changes over time in the importance and magnitude of spillovers and internal use are potentially an important proximate cause of the decline in corporate production of scientific research. Figure 2 shows that whereas use of internal research in invention has remained relatively stable, spillovers have increased over time.[3] We explore how much of the shift away from research by corporations is related to changes in its use in internal inventions versus in inventions by rivals.

Our analysis includes all publicly traded American firms with at least one year of positive R&D expenditures and at least one patent over the period 1980-2015. The final sample consists of an unbalanced panel of 4,090 firms and 54,274 firm-year observations. We match our firm data to two innovation data sources: (i) scientific publications from Web of Science, and (ii) patents and citations from PatStat. We measure use of research in invention by citations in the firm's patents to its own scientific publications. Spillovers are measured by citations in the patents of others to the focal firm's publications. With these newly constructed data we present three main findings.

First, we show that the production of corporate publications per firm has fallen by 20-50% over the sample period, after accounting for changes in the scale of firms. This decline should be interpreted as a shift in the composition of firm R&D, away from upstream research and towards more downstream development activities. The decline is sharpest for large firms and it varies substantially across industries. In the life sciences, in particular, there is even an increase in

[3] These trends are pithily illustrated in the following quote from a former Bell Labs researcher (Odlyzko (1995), p.4):

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

publications up to 2005, followed by a gradual decline over the next ten years to early sample levels.

Second, we document a strong positive relationship between corporate research and its use in invention by the corporation itself. Firms produce more scientific publications when their past research is cited in their own patents. Although we do not claim our estimate of this relationship as causal, the relationship endures even after controlling for firm fixed effects, as well as a variety of time-varying firm characteristics. Moreover, the patterns of association are consistent with the notion that firms invest in scientific research if they have been able to use it for their downstream inventions. For instance, the relationship between own use of past research and the production of scientific publications is stronger for more recent research, for high-quality research, and for research that is cited by high-quality patents.

Third, we show that knowledge spillovers to product market rivals are associated with lower publication output. We find that the effect of spillovers on publications has strengthened over the sample years. Furthermore, the composition of spillovers has changed as well. While early in our sample fewer than 10% of corporate citations come from rivals in the same 4-digit SIC code as the focal firm, this percentage rises to 30% by the end of our sample period. Simply put, more knowledge is spilling over to close competitors and firms appear to be more responsive to such spillovers. The increase in the incidence and importance of spillovers to rivals accounts for up to 45% of the estimated decline in corporate publication rate.

In stressing the value of research as an input into invention we harken to early economic conceptualizations which draw on Vannevar Bush's "linear model," whereby technical progress rests upon scientific advance.[4] Griliches (1986) pioneered systematic empirical measurement of the private economic returns to scientific research ("R"), separately from development ("D"). Using data for approximately 1000 large manufacturing firms from 1957 through 1977, Griliches found that firms that spent a larger share of R&D on "R" were substantially more productive. While Griliches's work demonstrated the potential for significant private returns to scientific research, systematic empirical evidence on the mechanisms by which these returns are realized remained rare. With no direct evidence that firms are actually using their research in downstream inventions, the simplistic nature of the linear model (Kline and Rosenberg, 1986; David et al., 1992),

[4]In Vanevar Bush's own words (Bush (1945):241) "Basic research leads to new knowledge. ... 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."

and with the growing emphasis of the public good nature of scientific research (e.g., Dasgupta and David (1994), Arrow (1962), Nelson (1959)), economists began to model knowledge spillovers as drivers of economic growth and firm performance.[5] As spillovers took center stage, new explanations for why firms engage in research were advanced. These explanations include absorptive capacity (Cohen and Levinthal, 1989; Cockburn and Henderson, 1998; Griffith et al., 2004, 2006; Aghion and Jaravel, 2015), incentives for high-skilled scientist-inventors (Stern, 2004; Henderson and Cockburn, 1994; Cockburn and Henderson, 1998), and enhancing reputation to attract investors, prospective customers, or employees (Hicks, 1995; Audretsch and Stephan, 1996).

While these alternative explanations enrich our understanding of why firms invest in scientific research, the present paper aims to shift scholarly attention back to the view that firms invest in science to advance technology. Indeed, though signalling and reputation building may well be important motivations for investing in scientific research, our empirical results imply that they cannot be the entire story. For instance, if firms invest in science principally to signal to external stakeholders, then citations to their science by rivals ought to have a positive impact on innovation outcome. Similarly, if scientific publications are merely the price of attracting talented researchers, then firms and investors ought not to care if these publications are used, either by the firm itself or by its rivals. Yet, we find that citations by product market rivals reduce publication output. Furthermore, our market value regressions indicate that publications cited by rivals are less valuable than publications cited by other firms. By contrast, publications that are internally cited tend to be associated with greater value. Our findings, therefore, indicate that investment in scientific research is undertaken not merely for the indirect benefits of signalling to regulators and investors, or attracting talented researchers. Instead, our findings are consistent with the view that firms hope that their research will lead to more and better inventions, but that spillovers to rivals will not be large enough to wipe out the gains.

Given the importance of estimating the impact of spillovers, we use firm fixed effects, and control for a variety of time varying firm characteristics. To further guard against unobserved time-varying factors that could confound the underlying relationships of interest, we also present estimates where we instrument for citations by rivals using tax credits as instruments for patenting by rivals, following Bloom et al. (2013) and Lucking et al. (2018).

[5] See Grossman and Helpman (1991), Aghion and Howitt (1992), Romer (1990), and Jones and Williams (1998). See Griliches (1992) for an early review of spillovers in the micro literature.

Bloom et al. (2013) (hereafter, BSV) marks an important advance in our understanding of the complex nature of spillovers. Building on Jaffe (1986), who measures spillovers using aggregate R&D expenditures by rivals, BSV distinguishes between two types of spillovers: from product market rivals and from technology rivals. While the latter captures a learning effect, which should improve the focal firm's innovation outcomes and productivity, the former captures a rent-stealing effect due to an increase in the knowledge base of close competitors that should adversely affect the focal firm in the product market. There are three main differences between BSV (and its predecessors) and the present paper. First, we focus on knowledge spill-outs, as opposed to knowledge spill-ins. That is, we examine how the use of own knowledge by outsiders affects the focal firm, rather than how a focal firm is affected by knowledge produced by other firms.[6] Second, we introduce a direct measure of use of science in rivals' inventions. While previous work typically measures potential spillovers by the (weighted) sum of R&D performed by rivals, we measure spillovers directly as patent citations to science produced by a focal firm. Third, we study research, the upstream part of R&D, rather than R&D as a whole. To the best of our knowledge, this is the first large scale attempt to study spillovers from, and internal use of, science generated by corporations.

We contribute to the debate on why for-profit firms participate in the production of scientific knowledge. Our first contribution is to develop and validate a new measure of use of science in invention. While previous research using citations data for patents and publications was mostly done for selected industries and years, we cover a broad range of companies across many industries over a third of a century. We match publication records from Web of Science to front-page non-patent literature (NPL) references to firms on a large scale. We confirm that 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. Lastly, in matching patents and publications to Compustat firms, we improve and extend the NBER patent database, as described below.

Conceptually, we connect the Bush-Griliches view and the spillover perspective. Internal

[6] Our paper is closer to Belenzon (2012) which examines the relationship between private returns to R&D and the use of own research by the focal firm and by its rivals.

research can be both privately rewarding and socially valuable. As long as the private costs of spillovers are offset by the benefits of internal use, firms might have sufficient incentives to invest in scientific research. We show that corporate science has declined partly because spillovers have increased while internal use has remained stable. By distinguishing scientific research from downstream development and probing why for-profit firms appear to be reducing their investment in scientific research, our paper adds to the debate on the apparent decline in inventiveness (e.g., Bloom et al. (2017)) and the associated slow down in productivity growth. That is, the apparent decline in inventiveness may be related to the withdrawal of companies from research. Although university research has increased considerably, it may not be a perfect substitute for corporate research as an input into invention (Arora et al., 2019).

2 Data

We use scientific publications as our measure of corporate production of scientific knowledge and patents as our measure of inventive activity. We treat a citation by a patent to a corporate publication as an indicator that the patented invention used, or built upon, the knowledge in the publication. We combine data from five sources: (i) company and accounting information from S&P Compustat, (ii) scientific publications from Web of Science (WoS), (iii) patent and non-patent literature (NPL) citation from PatStat; (iv) subsidiary data from ORBIS, and (v) acquisition data from SDC platinum.

We build upon and extend the NBER 2006 patent data (Hall et al., 2001). We re-construct the NBER data from 1980 and extend it to 2015 while introducing several improvements. We also develop new data on corporate publications matched to NPL citations. The Data Appendix provides details on all of our data construction efforts. We discuss them briefly below.

Accounting panel data. We construct our estimation sample as following. We start with all North American Compustat records and select companies with active records and positive R&D expenses for at least one year during our sample period of 1980-2015. We exclude firms without patents and firms that are not headquartered in the United States. Our final estimation sample consists of an unbalanced panel of 4,090 firms and 54,274 firm-year observations.

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

The second major challenge comes from ownership changes. A parent company and a majority owned subsidiary may have different identification numbers and records in Compustat. Moreover, a single company may correspond to multiple firm identifiers due to changes in ownership (such as mergers, acquisitions and spinoffs). We identify ownership structures and ownership changes in three ways. First, we match our sample firms to ORBIS ownership files for the years 2002-2015 for annual subsidiary information.[7] Second, for firms that exit Compustat before 2002, we manually collect subsidiary names based on SEC filings. Third, we match our firms to M&A data from SDC Platinum to supplement information on ownership changes.

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

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

Patent citation to corporate publications. A major contribution of this paper is matching non-patent literature (NPL) citations to publications as our measure of the use of corporate science in invention. Using all patents granted in the period 1980-2015, we perform a many-to-many match between NPL citations and WoS publications (approximately 10 million citations

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

matched to 800 thousand corporate publications), allowing for more than one publication to be matched to each citation. For each possible match we construct a score that captures the degree of textual overlap between the free-text NPL format and the structured WoS record, which includes the following fields: article title, journal and authors. To exclude mismatches, we use a more detailed secondary matching algorithm that is based on different WoS fields: standardized authors names, number of authors, article title, journal name and year of publication. The matching algorithm accounts for misspelling, unstructured text, incomplete references, and other issues that may cause mismatches. We manually verify the accuracy of the matches. We obtain 272,387 patent citations to 63,668 corporate publications by 159,279 citing patents.

Note that papers and patents are matched "dynamically" such that if a sample firm divests a subsidiary, the patents and papers of that subsidiary are removed from the focal firm's stock of patents and publications from that point onward, but not before. This also applies to citations from patents to publications, and whether they are classified as internal citations or external ones.

2.1 Descriptive statistics

Our main sample and variables are at the firm-year level.[8] Table 1 presents descriptive statistics for our main variables over the sample period. Our sample includes a wide distribution of firm sizes: market value ranging from 7 million dollars (10th percentile) to 4.4 billion dollars (90th percentile) and sales ranging from 3 million dollars (10th percentile) to 3.7 billion dollars (90th percentile). About 70% of firms have at least one publication during the sample period (by construction, all firms have at least one patent). These firms produce on average 18 publications per year. The distribution of publications is highly skewed with the median firm producing one publication per year. We observe a similar pattern for patents, with an average of 24 patents per firm-year and a median of 2 patents.

Table 2 presents summary statistics for citation variables for firms with at least one publication that is cited by a patent. Of the 3,008 publishing firms, 1,172 firms receive at least one citation from a patent to their publications. Publications by these firms receive 10.5 citations from patents per year, comprising of 1.6 internal citations (i.e., from patents filed by the firm itself) and 8.9 external citations (i.e., from patents filed by others). Of the firms that publish,

[8] Appendix Table B1 summarizes the definition and data source for each variable.

561 make at least one citation to their own scientific publications in their patents. Publications by these firms receive 1.6 annual internal citations with one publication cited by one internal patent per year. Of the publishing firms, 1,114 receive at least one external citation to their publications. Publications by these firms receive 8.9 patent citations per year, with 7.4 patents citing 5.2 publications. For citations from corporate patents, we distinguish between citations from product market rivals and from from firms with overlapping technology. Citations from the latter are more common with 2.2 citations per year relative to 0.8 citations per year by product market rivals.

Table 3 presents mean comparison tests for differences in characteristics between publishing firms with high and low internal citations (conditional on receiving at least one patent citation). Firms with internal citations above the mean (defined as the ratio of internal citations received to the sum of internal and external citations received) produce more publications and patents per dollar of R&D, and are more R&D intensive.[9]

An important implication of our analysis is that as long as the private costs of spillovers are offset by the benefits of internal use, firms might have sufficient incentives to invest in scientific research. To this point, Table 4 examines whether publications that are used internally are also used externally by comparing the mean of external citations between publications that are internally cited and publications that are not. There is a strong positive relationship between external and internal citations. For publications with at least one internal citation, 30% are cited externally. For publications that are not cited internally, however, fewer than 3% are cited by an external patent. Number of external citations display a similar pattern, with 0.8 external citations to publications with at least one internal citation, as compared to 0.05 citations to publications that are not cited internally. Our finding that publications that are internally cited also generate most of the spillovers highlights the importance of our investigation of whether corporate research is sustainable in the presence of spillovers.

[Insert Tables 1-4 here]

[9]Table 13 explores the parametric relationship between R&D productivity with publications and internal and external citations and confirms the positive relationship between publications and internal use with R&D productivity.

2.2 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). In the survey, lab directors in R&D performing firms were asked, among other things, about the extent to which their R&D projects used scientific knowledge from various sources.[10] Of the firms in our sample, 772 firms were also covered in the CMS, with patents granted between 1991 and 1999 (a total of 29,318 patents). Figures 3A-3C show a strong correlation between citations to science per patent by the surveyed firms' patents and the share of the firm's R&D projects that used scientific research.

Figures 3A and 3B present mean comparisons for the number of patent citations to publications per patent, by firms with above the median (high) and below the median (low) reported use of public science in R&D projects.[11] The figures show that firms with high self-reported use of public science also cite more public science (publications by all universities and research institutions (Figure 3A) and publications from top 200 American universities in the Shanghai Ranking (Figure 3B)).[12] Figure 3C 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 fields the firm reported as having influenced its R&D projects and the research areas cited by its patents.

Table 5 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 and complete sets of four-digit industry SIC codes and year dummies. In sum, firms whose patents cite scientific publications also reported that science contributed to their R&D projects. Furthermore, the fields of science that contributed the most to a firm are also those whose publications the firm's patents

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

[11] Reported use is based on responses to the questions "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?" (CMS 1994 Q.18)

[12] We also find that citations to science are positively strongly related to share of PhDs or MD 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 excludes firms that indicated their main field in Q22 as 'others'. Publications were classified into main fields based on key related words under the WoS journal subject category field. For example, for "Organometallics Journal", the related subject category is "Chemistry, Inorganic & Nuclear; Chemistry, Organic" and accordingly it is classified under main field of chemistry.

cite, and firms who draw on public science also tend to cite public science in their patents.[14] To our knowledge, this is the first direct validation of patent citations to scientific articles as a measure of the use of science in invention.

[Insert Table 5 here]

3 Econometric analysis

3.1 Time trend in corporate publications

We begin our econometric analysis by providing evidence of a declining production of corporate publications.[15] We estimate time trends in the rate of corporate publications as follows[16]:

$$\ln(Publications_{it}) = \alpha_0 + \alpha_1 Trend + \boldsymbol{Z}'_{it-2}\boldsymbol{\gamma} + \boldsymbol{\eta_i} + \epsilon_{it} \tag{1}$$

$Publications_{it}$ is the number of publications by firm i in year t, $Trend$ is the time trend computed as year t minus 1980 and is presented in decennial units (i.e., per decade). $\boldsymbol{Z}_{it-2}$ is a vector of two-year lagged firm-year controls, including patent stock, R&D stock, and sales.[17] $\boldsymbol{\eta_i}$ is a complete set of firm and year dummies, respectively. ϵ_{it} is an iid error term. Standard errors are clustered at the firm level.

Table 6 presents the estimation results. Consistent with Figure 1, we expect a falling publication rate, $\hat{\alpha_1} < 0$. This is confirmed by our estimates. Column 1 presents results from a pooled specification with a complete set of 4-digit SIC industry dummies, without firm fixed-effects. The estimates indicate that, over our sample period covering 3.5 decades, publication rate has fallen by more than 50%. Column 2 adds firm fixed effects, which lowers the time trend estimate considerably, implying an overall decline of about 20%. A possible explanation for the large fall in

[14]An earlier version of this paper, using a smaller sample, also showed that firms whose scientists also filed patents were also more likely to have its patents cite its papers. That is, patents were more likely to cite internal science when scientists and inventors overlap.

[15]Unless stated otherwise, one is added to number of publications and all specifications include a dummy variable for firm-year observations with zero publications.

[16]Bold indicates a vector representation.

[17]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 $Patentstock_t = Pat_t + (1-\delta)Patentstock_{t-1}$ where Pat_t is the number of patents in year t.

the time trend estimate is a substantial entry of low-publishing firms over the sample years. That is, firms entering the public equity markets over time are less likely to publish scientific research. Columns 3 presents similar estimates when weighting publications by the number of citations they receive from other publications and normalizing each citation by the average number of citations received by all other WoS publications published in the same year to account for truncation.

Because the number of journals is rising over time, comparing early to late publication rates might underestimate the fall in corporate publications when not accounting for the rise in available journal space. Column 4 presents time trend estimates when excluding new journals (journals established post-1990). As expected, the time trend effect rises in absolute value when holding journal space constant throughout the later sample period. Publications fall by about 40%.[18].

Columns 5 and 6 split the sample by firm size using median sales value. Large firms account for ten times the publications, and they exhibit the greatest shift in R&D composition, as reflected in the much sharper decline in publication rates (a total fall of 50% in publications by large firms relative to a 16% fall in publications by small firms).

Columns 7 breaks up the time trends into eight periods and includes separate dummies for each period (the base period is 1980-1985) to account for non-linear time effects. The decline in publication rate over the sample period is 27%, higher than the 19% estimated in the equivalent within-firm specification that imposes linear year effects (Column 2).

Columns 8-10 examine the robustness of our results to having zeros in our dependent variables. We present three specifications. Column 8 excludes firm-year observations with zero publications. The trend effect increases substantially and indicates a fall in publications of about 60% over our sample period. To reduce the prevalence of observations with zero publications, Column 9 restructures our panel to firm-5-year cohorts (instead of firm-1-year) using firm-year averages (hence, instead of having 35 periods, this specification includes only 7 periods. $Trend$ is defined accordingly, with the value of 0 for period 1 and the value of 7 for the last period). The trend effect increases as well, indicating a 35% decline in publications over our sample period.

Column 10 estimates our original panel using a negative binomial specification with firm fixed effects accounted for using pre-sample means (Blundell et al., 1995). For each firm in our sample, we calculate the 4-year average value of publications and exclude these years from our

[18] We also experimented with removing only the low quality new journals (journals with impact factor below unity) The decline in publications is about 30% over the complete sample period

sample. We refer to these average values as pre-sample means– our firm fixed effects control in the regression. The implied total decline in publication rate is 28%, higher than the within-firm estimates obtained from OLS (19%, Column 2).[19]

It is possible that some of the decline in publication output may reflect greater secrecy about scientific research rather than a decline in scientific research itself. However, a variety of direct and indirect evidence suggests that this is unlikely to be the entire story. For one, there are well documented cases of firms such as Xerox, HP, IBM, AT&T and DuPont reducing their scientific research. Second, aggregated data from the NSF Science and Engineering surveys show that privately performed basic and applied research, as a share of total private R&D has fallen. To examine further the secrecy explanation, we propose the following test. If firms are persisting in research but merely keeping it secret instead of publishing, we would expect a larger fall in publication rate for firms that operate in states that extend greater protection to trade secrets. We follow Klasa et al. (2018) and exploit variation in the adoption of Inevitable Disclosure Doctrine (IDD) by U.S. state courts. IDD is a legal doctrine that restricts worker mobility from one organization to another in cases where they might be inevitably disclosed 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.

We create an IDD dummy variable that receives the value of one if IDD is in effect in the focal firm's state in a given year, and zero otherwise. We add an interaction term between IDD and trend. If secrecy drives the drop in publication rate, we expect a negative and significant interaction effect. That is, the drop in publication rate should be larger for firms operating in states with stronger trade secret protection. Yet, as shown in Column 12, this is not the case, as the coefficient estimate on the trend-IDD interaction is positive rather than negative, but small and statistically indistinguishable from zero.

From Table 6, we conclude that firms are withdrawing from research, and that later entrants to the sample are less engaged in research. Further, these results are not artifacts of how the

[19]Corporate scientists coauthor with researchers from other institutions. Because our definition of a corporate publication includes publications with at least one corporate author, an important part of our sample includes co-authorship. We explore alternative definitions of corporate publications and single out coauthorship as follows. Publications are classified as non-coauthored if (i) all authors belong to a single firm, or (ii) the lead author of a publication is from of our corporate sample firms (based on "reprint" address). Publications are classified as co-authored if there is at least one non-corporate author and the leading author is not a corporate author. We find that non-collaborative and collaborative publications exhibit opposite trends over time. Our (unreported) estimates indicate that non-coauthored publications are falling over time at a rate of about 35% over our sample period, relative to an average decline of 19% when using the complete sample of corporate publications.

dependent variable is defined, nor of the estimation method.

[Insert Table 6 here]

Table 7 presents time trend estimation results across industries. We focus on three main industry categories[20]: (i) life sciences, (ii) IT & Software Communication, and (iii) Chemicals & Energy. There is substantial heterogeneity in the behavior of corporate publications over time by industry. While there is a decline in the publication rate in ICT and electronics and in physical sciences, the pattern for life science is less clear. For life sciences, there is an increase in publication rate from 1980 to about 2000 (Columns 1-3 indicate that publication rate in 2000 are 30-40% higher than the rate in 1980-1985), followed by a gradual decline until the end of our sample, where publication rate is statistically the same as at the beginning of our sample period.

Several factors may be responsible for the different time trend in life sciences. Insofar as patents are more effective in protecting innovations in life sciences relative to other industries, the returns from investments in research may be higher in the pharmaceutical sector than in other sectors, making spillovers less harmful. The commercial applicability of upstream research is also much more apparent in the pharmaceuticals industry (Li et al. (2017)). Consistent with this, we find that publications by life-science firms receive 2.5 more citations from patents than publications by non-life-science firms do (based on own calculations), underscoring the higher relevance of research to invention in the sector. Finally, corporate research in life sciences may have benefited from biomedical research funded by the National Institutes of Health, which increased dramatically, from US$2.5 billion in 1980 to US$29 billion in 2015.

[Insert Table 7 here]

3.2 Internal citation and publication output

We estimate the relationship between internal use of past research, measured by patent citations to internally produced science, and investment in research, measured by number of publications with a corporate author, as follows:

$$\ln(Publications_{it}) = \beta_0 + \beta_1 \ln(Internal\ citation_{it-2}) + \boldsymbol{Z}'_{it-2}\boldsymbol{\gamma} + \boldsymbol{\eta_i} + \boldsymbol{\tau_t} + \epsilon_{it} \qquad (2)$$

[20] Appendix Table B2 includes a list of all four-digit SIC codes that fall in each industry.

$Internal\ citation_{it-2}$ is the two-year lagged one plus the number of patent citations made by firm i's patents granted up to year t-2 (inclusive) to its own scientific publications. $\boldsymbol{Z}_{it-2}$ is a vector of two-year lagged firm-year controls. Our coefficient of interest is β_1 and we expect $\hat{\beta}_1 > 0$. One possible concern is that firms with a higher number of publications or patents are more likely to randomly cite one of their publications, which would bias $\hat{\beta}_1$ upward. To mitigate this concern, all 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).[21]

Table 8 presents the estimation results.[22] Column 1 presents the estimation results from a pooled specification with four-digit SIC dummies. 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 firm cross-section by averaging variables at the firm level. In Column 3, which adds firm fixed effects to the specification in Column 1, $\hat{\beta}_1$ falls sharply, indicating that the relationship between internal citation and publications reflects a substantial degree of heterogeneity across firms. Yet, $\hat{\beta}_1$ remains substantively and statistically significant: one additional internal citation is associated with an increase of 0.9 publications per year.

Columns 4-6 confirm that the results are robust to a variety of changes in sample and measurement. Column 4 restricts the sample to firms with at least one publication during the sample period and Column 5 controls for internal patent citations to own patents ("self-citations") to mitigate the concern that $\hat{\beta}_1$ captures a patent "self-citation" effect (Hall et al., 2005; Belenzon, 2012). Column 6 presents an alternative measure of internal citation, ratio of internal citations of all patent citations to science. The same pattern of results continues to hold.[23]

Not all citations to science are equally relevant. We expect internal citations to be more

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

[22] *Internal citation* has a mean of 1.8 with a standard deviation of 9.2.

[23] In unreported robustness checks, we exclude references to articles related to clinical trial phase in the pharmaceutical and biotech industry, which are considered to be very applied research and might even be considered part of "*D*." We identify clinical trial publications by related phrases in the title and abstract of each publication record (e.g., clinical trial, clinical study, pre-clinical trial, subjects). We locate fewer than 100 internally cited clinical trial publications and exclude them from the analysis, with very little change to the results. Additional unreported robustness checks also show that our results are not affected if we exclude citations added by patent examiners, or if we exclude citations where patent inventors cite their own publications.

relevant to a firm's decision to invest in research when the cited publication (i) is of high scientific impact, (ii) is more recent (and thus less likely to be merely a background reference), and (iii) is cited by the firm's valuable patents. These predictions are confirmed in Columns 7-9. Column 7 distinguishes between citations to old and 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 (0.172), the estimated coefficient of internal citations to old science is close to zero and statistically insignificant. The two estimates are statistically different from each other with a p-value<0.01.

Column 8 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. The coefficient estimate on citation to high quality internal publications is positive and statistically significant (0.088), while the coefficient estimate on internal citation to low quality publications is less than half of that and is statistically insignificant, albeit that the two estimates are not statistically different from each other at the 5% significance level.

Column 9 distinguishes between citations made 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 with publications is stronger for high quality citing patents compared to low quality citing patents (0.071 vs. 0.037; estimates are not statistically different from each other).

In summary, Columns 7-9 show that internal patent citations to science that matter for the production of future science are citations that come from high quality patents of the sponsoring firm to recent, high-quality, publications. These results are consistent with the view that scientific output is an input into downstream inventive activity, and that to justify further investment in research, managers need to demonstrate that their recent scientific work is useful for the downstream inventive activity of the sponsoring firm.

[Insert Table 8 here]

3.3 Knowledge spillovers

This section investigates how investment in research is related to external citations: citations made by patents filed by others to the research published by the focal firm. Not all citations represent profit-reducing spillovers. If a firm invests in research to signal quality to regulators and customers, or to attract talented researchers, citations of its publications by others would validate its claims to quality and reinforce the signal. That is, external citations, rather than representing profit-reducing spillovers, would increase profits. At a minimum, citations by external patents would not be associated with lower profits or lower investment in scientific research by the focal firm. On the other hand, if firms invest in research because it is an input into internal inventive activity, the use of this research by rivals (but not by non-rivals) would lower the return to such investments (Nelson, 1959; Arrow, 1962; Bloom et al., 2013). This suggests distinguishing between citations from rivals from those from other firms.

We build on Bloom et al. (2013) to construct $RIVAL$ and $TECH$ as our measures of citations from firms that are close to the focal firm in the product market and technology space, respectively. Arguably, firms suffer more intensely from spillovers if their sales distributions across different line-of-business segments are more similar. In the same vein, firms are closer in the technology space if their distributions of patents across technology classes are more similar. We follow BSV and measure product market proximity as the cosine similarity of vectors representing the shares of industry segment sales for each pair of firms (labeled as SEG). Industry segments are from Compustat's operating segments database (a total of up to 70 segments per firm). Citations received by firm i from firm j are weighted by SEG_{ij}, the "distance" of citing firm j from the cited focal firm i in the product market. The variable $RIVAL_i$ captures citations from product market rivals as the weighted sum of all citations received. Technology space proximity is measured as the cosine similarity of vectors representing the shares of patents across 4-digit IPC classes for each pair of firms (labeled as TEC). The corresponding measure $TECH_i$ uses TEC_{ij}, the distance in technology space, as weights, for citations received by firm i from patents of firm j.[24]

[24]For this analysis, we include only citations by corporate patents. SEG proximity for each cited-citing

Insofar as $RIVAL$ represent profit-reducing spillovers, we expect these to be negatively related to publications. $TECH$ citations may proxy for the quality of publications, and hence could have the opposite sign.

We estimate the relationship between publications and spillovers, measured as patent citations by outsiders to the focal firm publications, as following:

$$\begin{aligned} \ln(Publications_{it}) = \beta_0 &+ \beta_1 \ln(Internal\ citation_{it-2}) \\ &+ \beta_2 \ln(RIVAL_{it-2}) + \beta_3 \ln(TECH_{it-2}) \\ &+ \boldsymbol{Z}'_{it-2}\boldsymbol{\gamma} + \boldsymbol{\eta_i} + \boldsymbol{\tau_t} + \epsilon_{it} \end{aligned} \tag{3}$$

$RIVAL_{it-2}$ is one plus the number of citations the focal firm receives from outsiders, weighted by the proximity of the citing firm to the focal firm in the product market. $TECH_{it-2}$ is the equivalent measure for citations by outsiders weighted by proximity to the focal firm in the technology space.[25] Our interest is in the coefficient β_2. If spillovers lower private returns to scientific research, we expect $\hat{\beta}_2 < 0$.

Table 9 presents the estimation results.[26] Column 1 includes the aggregate number of external citations made 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, the coefficient estimate on external citation is negative and statistically indistinguishable from zero. Column 2 distinguishes between $RIVAL$ and $TECH$ citations. The coefficient estimate on $RIVAL$, $\hat{\beta}_2$, is negative and statistically significant, while the coefficient estimate on $TECH$ has the opposite sign. 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. This positive relationship may capture an unobserved firm-specific publication quality effect. Column 3 includes only publishing firms with no major change in the estimates.

firm pair is the absolute un-centered correlation between their sales segment share vectors and is calculated as $|S_iS_j/\sqrt{S_i \times S_j}|$, where s_i is business segment sales shares vector for firm i. The measure ranges from 0 (least correlated) to 1. Similarly, TEC proximity is computed based on firm's patent share distribution across technology fields (4-digit IPC).

[25] $RIVAL_{it} = \sum_{k=1}^{N} D_{ik} \times SEG_{ik}$, where n_{ik} is the number of citations from patents of firm k to publications by firm i up to year t-2 (inclusive), and SEG_{ik} is the product market proximity between the two firms. We compute $TECH$ similarly as $TECH_{it} = \sum_{k=1}^{N} n_{ik} \times TEC_{ik}$, where TEC_{ik} is the proximity in the technology space.

[26] $External\ citation$ has a mean of 5.4 with a standard deviation of 45.3. $RIVAL$ has a mean of 1.4 with a standard deviation of 4.2 and $TECH$ has a mean of 2.1 with a standard deviation of 12.2.

Based on the estimates from Column 2 and evaluated at the sample mean, one additional citation from a product market rival is associated with a 0.9 decline in publications[27]. This decline is equal in magnitude to the increase in publications due to an additional internal citation (Column 3 of Table 8). An interpretation of this result is that a single citation by a rival would not lead to a decline in scientific research if the focal publication is also cited internally.

Column 4 presents estimation results using a Mahalanobis-based extension of the SEG weights to construct $RIVAL$. As with the earlier measure of SEG, cosine similarity is calculated between two vectors of shares of industry segment sales for each pair of firms, but the vectors are further weighted by each firm's share of sales within each of the industry segments.[28] For example, if Firm 1 and Firm 2 have similar sales shares across operating segments, the proximity score of the firms would be high. At the same time, if Firm 1 and Firm 2's sales both account for a large share of total industry sales in a specific segment, say Segment A, then Segment A would be given a high weight in determining the proximity score between Firm 1 and Firm 2. Column 5 presents estimation results using the Mahalanobis extension of $TECH$.[29] The results in Columns 4 and 5 are similar to those in Column 2. 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.

Finally, Column 6 presents estimates of how the responsiveness of publication output to spillovers to rivals has changed over time. It shows that β_2 has become larger in absolute magnitude over time. That is, publication output appears to have become more sensitive to spillovers to rivals. Greater sensitivity to spillovers would suggest that they have become more harmful.

[27]Calculated as: $(-0.144 \times 0.45 \times 19)/1.4 = 0.9$, where -0.144 is the coefficient on $ln(RIVAL_{it-2})$, 19 is average $Publication$ plus one, 0.45 is average SEG value and 1.4 is average $RIVAL$.

[28]Formally, we define $B = [B'_1, B'_2, ..., B'_N]$ as a matrix with each column consisting of shares of segment-specific sales within each firm. These shares are weighted by the share of the total segment sales that each firm's sales account for, resulting in $W = [W'_1, W'_2, ..., W'_N]$, where W'_1 is weighed shares of segment-specific sales for Firm 1. Given these vectors of weighted sales shares for each firm, proximity between two firms, say Firm 1 and Firm 2, is the cosine similarity of the vectors: $SEG_{12} = \dfrac{W'_1 \cdot W'_2}{||W'_1||||W'_2||}$.

[29]Formally, we define $T = [T'_1, T'_2, ..., T'_N]$ as a matrix with each column consisting of patent shares across 4-digit IPCs within each firm. These shares are weighted by the share of the total patents within each 4-digit IPC that a firm's patents account for, resulting in $V = [V'_1, V'_2, ..., V'_N]$, where V'_1 is weighted shares of patents across 4-digit IPCs for Firm 1. Given these vectors of weighted patent shares for each firm, proximity between two firms, say Firm 1 and Firm 2, is the cosine similarity of the vectors: $TEC_{12} = \dfrac{V'_1 \cdot V'_2}{||V'_1||||V'_2||}$.

In turn, this might reflect more intense product market competition. It might also be that each firm faces more technically-competent rivals so that a given piece of knowledge disclosed to the outside world is more likely to be seized upon and used to reduce the potential profit from that knowledge. As we discuss below, and as Figure 2 shows, this interpretation is also consistent with a growth in the volume of spillovers themselves.

[Insert Table 9 here]

3.4 Stock market value

If internal citations increase 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 firm 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 (we also present a Tobin's Q version of this specification):

$$\begin{aligned}\ln(Market\ value_{it}) = {} & \phi_0 + \phi_1 \ln(Internal\ citation_{it-2}) \\ & + \phi_2 \ln(RIVAL_{it-2}) + \phi_3 \ln(TECH_{it-2}) \\ & + \phi_4 \ln(Assets_{it-1}) + \boldsymbol{Z}'_{it-2}\boldsymbol{\gamma} + \boldsymbol{\eta_i} + \boldsymbol{\tau_t} + \epsilon_{it}\end{aligned} \tag{4}$$

Internal citation here is the cumulative number of citations made by the focal firm's patents to its own publications up to year t-2, $RIVAL$ is the cumulative number of citations from outsiders weighted by SEG up to year t-2, and $TECH$ is the equivalent construct where external citations are weighted by $TECH$. *Assets* is the book value of physical capital[31] and $\boldsymbol{Z}$ is a vector of controls including two-year lagged sales, and perpetual stocks of R&D, publications and patents. 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 coefficients ϕ_1 and ϕ_2. We expect $\hat{\phi}_1 > 0$ and $\hat{\phi}_2 < 0$.

Table 10 presents the estimation results. Column 1 shows a positive relationship between internal citation and market value (elasticity of 0.07). Internal citations are positively related to value, consistent with scientific research being an input into invention. There is no relationship

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

[31] Assets is the sum of net plant, property and equipment, inventories, investments in unconsolidated subsidiaries, and intangibles other than R&D.

between external citations (the sum of $RIVAL$ and $TECH$) and value, consistent with the view that not all citations are the same. That is, some proxy for quality whereas others represent profit-reducing spillovers. Accordingly, column 2 distinguishes between external citations received from rivals in product markets ($RIVAL$) and those from other firms operating in similar technical domains ($TECH$). This decomposition of external citations results in a negative and statistically significant coefficient estimate on $RIVAL$. The coefficient estimate on $TECH$ is positive, similar to our findings from Table 9. Columns 3 and 4 show that the results are robust to the Mahalanobis version of product market weights and to restricting the sample only to publishing firms.

The estimates from Column 2 indicate that one additional internal citation mitigates the negative market value effect of approximately 1.1 $RIVAL$ citations.[32]

Column 5 presents an important result. We explore how the spillover-value relationship varies over time by interacting citations by rivals with a time trend. The coefficient estimate on this interaction is negative and statistically significant. That is, our estimates indicate that the relationship between value and spillovers to rivals becomes increasingly negative. This finding complements our findings from Column 6 in Table 9, which document a similar pattern for the relationship between spillovers and publications, and is consistent with the view that spillovers have become more harmful over time.

Columns 6-8 divide the sample into three main sectors. Overall, the patterns are similar. Internal use is associated with higher value (although the coefficient estimate for chemicals and materials is negative, but imprecisely estimated) whereas use by product market rivals is associated with lower value. There are, however, some interesting differences. As Table 7 also suggested, the life science sector is an outlier, in that the estimates suggest that spillovers are not harmful for private value. One possible explanation is that in the life science sector, citations by rivals do not reflect spillovers. Instead, they reflect potential opportunities to license to the citing firm. That is, strong intellectual property rights and an active market for technology allow sponsoring firms to monetize knowledge leakage to rivals.

Columns 9 and 10 present estimation results for Tobin's Q specifications. The negative value-spillover relationship remains robust. However, the relationship between value and internal

[32] The marginal effect of an additional citation, evaluated at the sample mean, is $(-0.076 \times 3519/3.5) \times 0.45 = -34.4$ (3.5. is the sample average of $RIVAL$ citations; average SEG value $= 0.45$ – the average value an additional citation receives in constructing $RIVAL$). The same calculation for internal use is: $(0.069 \times 3519/6.4) = 37.9$ (6.4 is one plus average stock of internal citations). Hence, to offset one rival citation, 0.9 internal citations are needed.

use appears to be small and imprecisely estimated.

Taken together, the evidence from Table 10 supports the view that private returns 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]

3.5 Instrumental variable estimation

An important concern is that investment in research and citations made in patents can be driven by common unobserved or miss-measured time-varying effects, such as technological opportunity shocks. Unobserved technological opportunities would generate a positive correlation between the number of inventions by rivals, and hence citations to focal publications, and focal publications, leading to an upward bias in the OLS estimate of the spillover coefficient. To address this concern, we rely on a source of exogenous variation in rival patenting, unrelated to research opportunities.

We follow Bloom et al. (2013) and Lucking et al. (2018) and exploit state variation in tax credits as an instrument for rivals' patenting.[33] For each firm in our sample, we calculate its cost of R&D and regress the number of patents against this cost. The predicted number of patents from this regression is used as our input into calculating a focal firm-specific aggregate number of predicted patents by its rivals, where the aggregation is based on the weighting procedures discussed in Section 3.3. The aggregate rival patents are used as our instrument for $RIVAL$ in the publication and stock market value regressions. We implement the IV approach by first projecting our logged patent count variable on both state and federal tax credit components of R&D user cost. Table B3 presents the results from this estimation (joint F-statistics of both

[33] As in Bloom et al. (2013), we look at Hall-Jorgenson's user cost of capital for firm i in state s at time t (Hall and Jorgenson, 1967): $\rho_{it} = \dfrac{(1-D_{it})}{(1-\tau_{st})}\left[I_t + \delta - \dfrac{\Delta p_t}{p_{t-1}}\right]$, where D_{it} is the discounted tax credits and depreciation allowance and τ_{st} is firm's tax rate. $\left[I_t + \delta - \dfrac{\Delta p_t}{p_{t-1}}\right]$ is constant across firms and is therefore ignored, and only the component, $\rho^P_{it} = \dfrac{(1-D_{it})}{(1-\tau_{st})}$ is considered. ρ^P_{it} is further decomposed into federal and state components, where the state-level instrument is constructed using the estimates of state-specific R&D tax rates from Wilson (2009) and each firm's distribution of patent inventors across states. Formally, the R&D tax rate for firm i in year t based on state tax credit is: $\rho^S_{it} = \sum_s \theta_{ist}\rho^S_{st}$, where θ_{ist} is 10-year moving average of the fraction of firm i's patent inventors in state s in year t and ρ^S_{st} is the tax rate for state s in year t. (The notation is borrowed directly from Bloom et al. (2013).) As in Bloom et al. (2013) and Hall (1993), the federal tax credit component is calculated by multiplying the difference between firm specific base R&D expenditure and actual R&D expenditure with the appropriate credit rate. The definition of base R&D expenditure has changed in 1990 from a maximum of prior 3-year rolling average of R&D expenditures (or 50% of current year's expenditure) to R&D to sales ratios between 1984 and 1988 times current year's sales (up to a ratio of 0.16).

tax variables is above 14.5). Next, we calculate the predicted value of logged patent count using the regression estimates, $\hat{\pi}_{it}$. For each firm i, we compute $RIV\hat{A}LPAT_{it} = \sum_i SEG_{ij}\hat{\pi}_{jt}$, where SEG_{ij} is the distance in product space between firm j and focal firm i, using cosine distance (and, in a separate specification, the Mahalanobis distance) for computing $RIVAL$ described earlier. Finally, we use $RIV\hat{A}LPAT_{it-2}$ as an instrument for $RIVAL_{it-2}$.

Conceptually, our IV strategy is to use the variation in R&D, and hence, in patenting, of a firm due to R&D cost shifters (i.e., tax credits) to purge $RIVAL$ of the variation arising from unobserved shocks to scientific and technical opportunities. In plain language, if a particular line of scientific inquiry becomes economically promising, firms might produce more patents and papers in that field, thereby confounding the interpretation of the estimated relationship between citations received from patents and future publications. R&D tax credits affect the marginal cost of R&D, but not the benefit. Therefore, they offer a source of variation in R&D that is independent of the confounding variation.

Table 11 presents the estimation results for instrumenting $RIVAL$ with $RIVAL\hat{P}AT_{it-2}$. Columns 1-3 present the results for publications. Column 1 presents the first stage results of regressing $RIVAL$ against $RIV\hat{A}LPAT_{it-2}$. As expected, higher R&D by rivals leads to more citations by these rivals to the focal firm's publications. Column 2 presents the second stage results of regressing publications on the predicted $RIVAL$ from stage one using the 2SLS procedure. While there is no change in the coefficient estimate on internal citations, the estimate on $RIVAL$ rises in absolute magnitude, indicating a larger negative effect of rival citations on focal publications. The IV estimate is about three times larger than the OLS estimate, indicating that to offset the negative effect of an additional rival citation, 2.4 internal citations are required (as compared to one internal citation using OLS estimates).[34] Column 3 shows the second stage results when we use the Mahalanobis distance to construct $RIVAL$. The estimated coefficient barely changes.

Columns 4-6 present the results for stock market value. Column 4 shows the first stage where $RIVAL$ is regressed against $RIVAL\hat{P}AT_{it-2}$. Column 5 presents the second stage of the 2SLS estimates. Once again, the coefficient estimate on internal citation is very similar to the OLS estimates in Table 10. However, the coefficient estimate on $RIVAL$ is significantly larger

[34] An additional $RIVAL$ citation lowers publications by $(-0.305 \times 0.45 \times 19)/1.8$ (0.45 is average SEG value– the contribution of an additional citation to $RIVAL$, 1.8 is average $RIVAL$, and 19 is one plus average annual publications).

in magnitude. Using the Mahalanobis distance to construct $RIVAL$ results in somewhat smaller coefficient estimates as well as somewhat smaller standard errors.

The estimates in Table 11 are consistent with the view that spillovers to rivals are privately harmful, and as a result cause firms suffering from the spillover to reduce investment in research.

[Insert Table 11 here]

3.6 Spillovers over time

Figure 2, presented earlier, shows that citations that corporate publications receive from both internal and external patents have been rising over time, but citations from external patents have been rising faster. Table 12 presents more systematic evidence on these trends. Column 1 shows that a given publication receives more citations from external patents (corporate and non-corporate). Column 2 shows that citations received from patents filed by other Compustat firms have increased as well. Column 3 confirms what Figure 2 suggested, namely that although internal citations have increased, the increase is much smaller that the increase in external citations. The point estimates suggest that citations from other companies have increased five times as quickly as internal citations. Columns 4 and 5 show that the composition of spillovers have changed as well, with more coming from product market rivals (citing firms that operate in the same 4-digit SIC code as the focal firm). Thus, whereas Table 9 suggested that companies are becoming more sensitive to a given level of spillovers, Columns 1-5 indicate that spillovers themselves are becoming more extensive. These results indicate that growing spillovers, and the growing sensitivity to spillovers, are both at work in explaining the decline in publication output.

Columns 6-9 present additional evidence on the evolving nature of spillovers. They show an increase in the concentration of citing firms. The average size of a citing firm rises substantially (Column 6), and the number of citations these firms make per publication increases. That is, more patents by the same citing firms cite a focal publication (Column 7). At the same time, the number of new citing firms per cited publication has fallen considerably (Column 8). The evidence from Columns 6-8 on increased concentration in the number of citing firms and rising citation intensity by these firms are consistent with a rise in the technical overlap among rivals, which can potentially explain why spillovers have become more harmful over time.

To quantify how spillovers affect the decline in publication rate, consider a simplified version of our estimation equation:

$$Publications_{it} = \alpha + \beta_t RIVAL_{it} + \boldsymbol{Z}'_{it}\boldsymbol{\gamma} \tag{5}$$

where $Z(it)$ includes all other variables that condition publication output. Note that the publication response to spillovers, $\beta(t)$ varies over time. Then, holding $Z(it)$ constant, the proportionate change in publications due to changes in spillovers can be written as

$$\Delta Publications_{it} = \beta_t \Delta RIVAL_{it} + RIVAL_{it}\Delta\beta_t \tag{6}$$

From Table 12, column 1, number of external citations *per publication* increases at 0.7% per year, or 25% over the complete sample period. On average our sample firms hold 1.2 publications that have received at least one external corporate patent citation. Thus, Δ $RIVAL_{it}$= 0.3 (0.25×1.2).[35] From Table 11, column 2, we can calculate β_t as: $-0.31 \times 13.6 = -4.2$.[36] Finally, Table 9, column 6 implies that $\Delta\beta_t = -3.3$ ($-0.07 \times 3.5 \times 13.6$). Thus, the first term in 6, which measures the decline due to an increase in the incidence of spillovers, is -1.3 (0.3×-4.2). The second term, which measures the effect of a heightened responsiveness to spillovers, is -4.6 ($= -3.3 \times 1.4$). Based on the estimates from Table 6, Column 5, up to 45% of the decline in publication rate can be explained by an increase in the incidence and importance of spillovers to rivals.

[Insert Table 12 here]

3.7 R&D productivity

In this section we relate the productivity of the firm's R&D program to the production and use of scientific research. Table 13 shows that scientific publication stock (adjusted for quality) is associated with a higher number of citation-weighted patents per R&D dollar. Further, holding R&D and publications fixed, the higher the internal use of corporate science, the more productive is the firm's R&D program.

[35] A change in number of publications that can be cited will also contribute to an increase in citations from rivals, which we remove for this calculation.

[36] Evaluated at the sample mean of 19 publications, 1.4 $RIVAL$ citations and an elasticity of publications with respect to $RIVAL$ of -0.31.

Table 13 presents the estimation results of the following specification:

$$\begin{aligned}\ln(Patents_{it}) = \lambda_0 + \lambda_1 Internal\ citation_{it-2} + \lambda_2 \ln(R\&D\ stock)_{it-2} \\ + \lambda_3 \ln(Publication\ stock)_{it-2} + \boldsymbol{\eta_i} + \boldsymbol{\tau_t} + \epsilon_{it}\end{aligned} \tag{7}$$

Our measure of R&D output, $Patents_{it-2}$, is the the annual flow of patents weighted by the number of citations each patent receives divided by average number of citations received by all other patents granted in the same year. *Internal citation* is number of patent citations to own publications per publication (lagged by two years). Other controls include stock of R&D, lagged by two years, and the the stock of publications, lagged by two years. We expect that firms with more scientific R&D programs to be more productive ($\hat{\lambda_3} > 0$) and that own science should boost productivity more when it is useful for invention ($\hat{\lambda_1} > 0$).

As shown in Table 13, both predictions are confirmed in our data. Column 1 presents the pooled estimates, with industry effects. Controlling for R&D and publication stocks, a 10% increase in internal citation raises patenting output by slightly more than 2.5%. This estimate halves in value in Column 2, which controls for firm fixed effects. Column 3 allows the coefficient estimate on R&D stock to depend upon publication stock as well as the stock of internal citations. This implies interactions between R&D stock and publication stock, as well as between R&D stock and internal citations. The estimates indicate that firms that publish produce more patents per R&D dollar, and especially if they are able to use their research in invention. That is, both interaction terms are positive and statistically significant.

The results are robust to alternative specifications. Column 4 normalizes internal citations by scale of research. Similarly, Column 5 uses the log of citation-weighted patents to R&D as the dependent variable. Results continue to hold for both specifications. Allowing the coefficient on R&D to depend on publication stock and share of internally cited publications once again indicates that firms that publish have more productive R&D programs, especially when the inventive activity of the firm benefits from the research produced.

[Insert Table 13 here]

4 Conclusion and discussion

Our paper contributes to the growing discussion of why American firms are withdrawing from investment in scientific research. Our findings support the view that firms invest in research because its scientific output feeds into downstream technology development. But while firms benefit from this research in their internal invention, they are hurt when the knowledge spills over to rivals. 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. Over time, internal research has remained useful for corporate invention, but spillovers have tended to become more important, contributing to the decline of corporate participation rate.

Put differently, we find that even as firms make greater use of the scientific knowledge produced by others, they themselves are less willing to produce such knowledge, preferring to focus attention and resources elsewhere: from upstream research to downstream development, from "R" to "D". This shift, though likely privately profitable, is not without social costs. The results in Table 13 point to one such possible cost. The declining corporate engagement in research may be contributing to the reported decline in R&D productivity and the associated decline in productivity growth (e.g., Bloom et al. (2017)). Our findings suggest that a more careful investigation of the link between a possible decline of R&D productivity and the decline in scientific research represents a useful line of further inquiry.

The more obvious cost is that firms contribute less to the pool of knowledge available to advance innovation. That firms benefit from a public good, but are unwilling to contribute to it may be ironic, but certainly not surprising to economists. Our findings raise a different question: Why have spillovers become more significant over time, both in size and in impact upon research investments? The obvious candidate answer, namely product market competition (Aghion et al. (2005)), has to contend with recent research pointing to growing market concentration and the rise of superstar firms (David et al. (2017)). It may well be that though firms enjoy market power, their stay at the top is more short lived due to greater competition for the market (Segal and Whinston, 2007). Another driver of change in spillovers may have to do with changes in information technology and the speed and efficiency with which knowledge diffuses. A third important factor conditioning spillovers is intellectual property protection, which might affect spillovers (Galasso and Schankerman (2014)) and the direction of technical change (Moser (2005)).

Here, once again, there are contradictory impulses. On the one hand, it is widely acknowledged that intellectual property protection in the United States was strengthened in the 1980s. On the other hand, the last decade has seen a push back, with several court cases weakening patent protection. Moreover, there are significant differences across industries as well. As noted earlier, life sciences, where intellectual property protection is the strongest (Williams (2013)), have seen the smallest decline, perhaps because profit-reducing spillovers are the least widespread there, compared to sectors such as materials, chemicals, and information technology. These speculations indicate that an important direction for future research is to understand what lies behind the increases in the volume and importance of knowledge spillovers across firms.

References

Aghion, P., Bloom, N., Blundell, R., Griffith, R., and Howitt, P. (2005). Competition and innovation: An inverted-U relationship. The Quarterly Journal of Economics, 120(2):701–728.

Aghion, P. and Howitt, P. (1992). A Model of Growth through Creative Destruction. Econometrica, 60(2):323–351.

Aghion, P. and Jaravel, X. (2015). Knowledge spillovers, innovation and growth. Economic Journal, 125(583):533–573.

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

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

Arora, A., Belenzon, S., Patacconi, A., and Suh, J. (2019). The Changing Structure of American Innovation: Some Cautionary Remarks for Economic Growth. NBER Working Papers: No. 25893.

Arrow, K. (1962). Economic welfare and the allocation of resources for invention. In The rate and direction of inventive activity: Economic and social factors, pages 609–626.

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

Belenzon, S. (2012). Cumulative innovation and market value: evidence from patent citations. Economic Journal, 122(559):265–285.

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

Bloom, N., Schankerman, M., and Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. Econometrica, 81(4):1347–1393.

Blundell, R., Griffith, R., and Reenen, J. V. (1995). Dynamic count data models of technological innovation. Economic Journal, 105(429):333–344.

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

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):157–182.

Cohen, W. M. and Levinthal, D. A. (1989). Innovation and learning: the two faces of R & D. Economic Journal, 99(397):569–596.

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). NBER Working Papers: No. 7552.

Dasgupta, P. and David, P. A. (1994). Toward a new economics of science. Research policy, 23(5):487–521.

David, H., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2017). The fall of the labor share and the rise of superstar firms. NBER Working Papers: No. 23396.

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):73–90.

Galasso, A. and Schankerman, M. (2014). Patents and cumulative innovation: Causal evidence from the courts. The Quarterly Journal of Economics, 130(1):317–369.

Griffith, R., Harrison, R., and Van Reenen, J. (2006). How special is the special relationship? using the impact of us R&D spillovers on UK firms as a test of technology sourcing. American Economic Review, 96(5):1859–1875.

Griffith, R., Redding, S., and Reenen, J. V. (2004). Mapping the two faces of R&D: Productivity growth in a panel of oecd industries. Review of Economics and Statistics, 86(4):883–895.

Griliches, Z. (1986). Productivity, R and D, and basic research at the firm level in the 1970's. American Economic Review, 76(1):141–154.

Griliches, Z. (1992). The search for R&D spillovers. Scandinavian Journal of Economics, 94(0):S29–47.

Grossman, G. M. and Helpman, E. (1991). Innovation and growth in the global economy. MIT press.

Hall, B. H. (1993). R&D tax policy during the 1980s: success or failure? Tax Policy and the Economy, 7:1–35.

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

Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2001). The NBER patent citation data file: Lessons, insights and methodological tools. NBER Working Papers: No. 8498.

Hall, R. E. and Jorgenson, D. W. (1967). Tax policy and investment behavior. American Economic Review, 57(3):391–414.

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

Hicks, D. (1995). Published papers, tacit competencies and corporate management of the public/private character of knowledge. Industrial and Corporate Change, 4(2):401–424.

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

Jones, C. I. and Williams, J. C. (1998). Measuring the social return to R&D. The Quarterly Journal of Economics, 113(4):1119–1135.

Klasa, S., Ortiz-Molina, H., Serfling, M., and Srinivasan, S. (2018). Protection of trade secrets and capital structure decisions. Journal of Financial Economics, 128(2):266–286.

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

Kuznets, S. (1971). Nobel Prize Lecture: Modern Economic Growth: Findings and Reflections. Nobelprize.org.

Li, D., Azoulay, P., and Sampat, B. N. (2017). The applied value of public investments in biomedical research. Science, 356(6333):78–81.

Lucking, B., Bloom, N., and Van Reenen, J. (2018). Have R&D spillovers changed? NBER Working Papers: No. 24622.

Moser, P. (2005). How do patent laws influence innovation? evidence from nineteenth-century world's fairs. American Economic Review, 95(4):1214–1236.

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

Nelson, R. R. (1959). The simple economics of basic scientific research. Journal of Political Economy, 67(3):297–306.

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

Romer, P. M. (1990). Endogenous technological change. Journal of Political Economy, 98(5, Part 2):S71–S102.

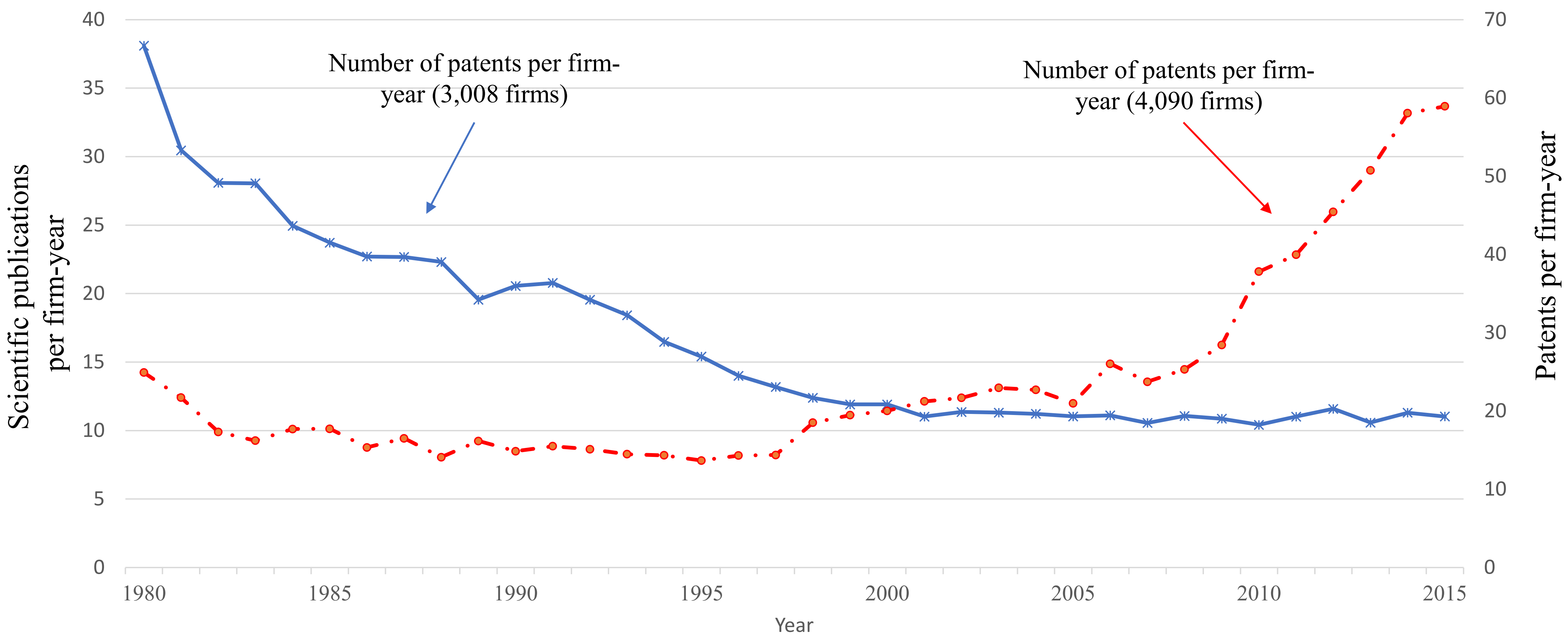
Segal, I. and Whinston, M. D. (2007). Antitrust in innovative industries. American Economic Review, 97(5):1703–1730.

Stern, S. (2004). Do scientists pay to be scientists? Management Science, 50(6):835–853.

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

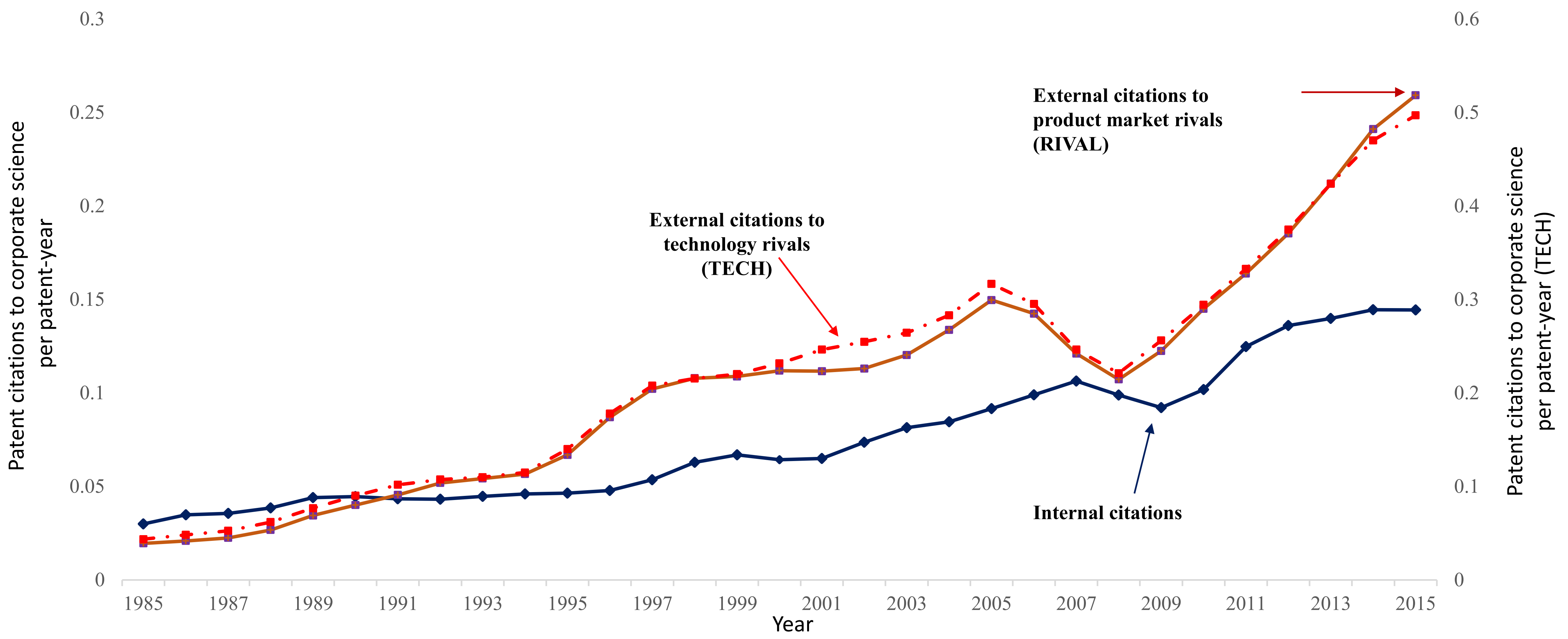
Wilson, D. J. (2009). Beggar thy neighbor? the in-state, out-of-state, and aggregate effects of R&D tax credits. Review of Economics and Statistics, 91(2):431–436.

Figure 1. Trend in Corporate Scientific Publications and Patents, 1980-2015



Note: The figure presents per-firm-year publications and patents over time for our sample of corporate firms. The sample is at the firm-year level and includes an unbalanced panel of 4,090 U.S. headquarter publicly traded ultimate owner parent companies (of which 3,008 are publishing companies) over the sample period 1980-2015. These firms have at least one year with positive R&D expenditures, at least one patent and at least 3 consecutive years of active records in Compustat during our sample period. Annual publication is conditional on at least one publication stock and excludes scientific publications from new journals post 1990.

Figure 2. Trends in Use of Corporate Science by Corporate Patents, 1985-2015



Note: The figures presents trends over time in citations to corporate science per corporate patents for our sample of corporate firms. The sample is conditional on firm-years with at least one granted patent. Corporate science includes WoS scientific publications with at least one author employed by our sample of corporate firms. Internal citations per patent include citations to the publications with at least one author employed by the focal firm. RIVAL and TECH measure the product market proximity and the technology market proximity, respectively.

Figure 3. Patent citations to science are related to importance of research

Figure A. Patent citation to WoS articles, all publications

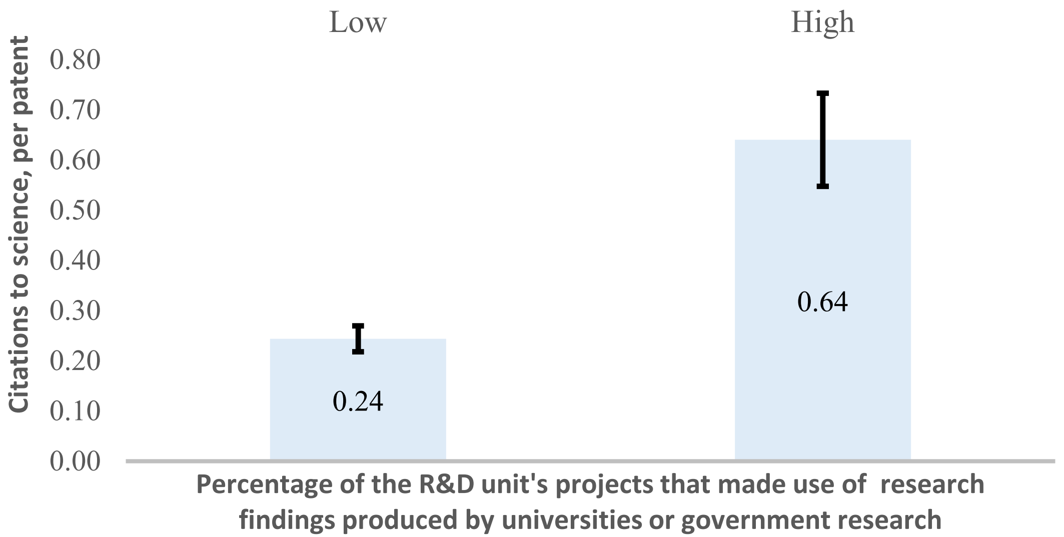


Figure B. Patent citation to WoS articles by top 200 U.S. universities

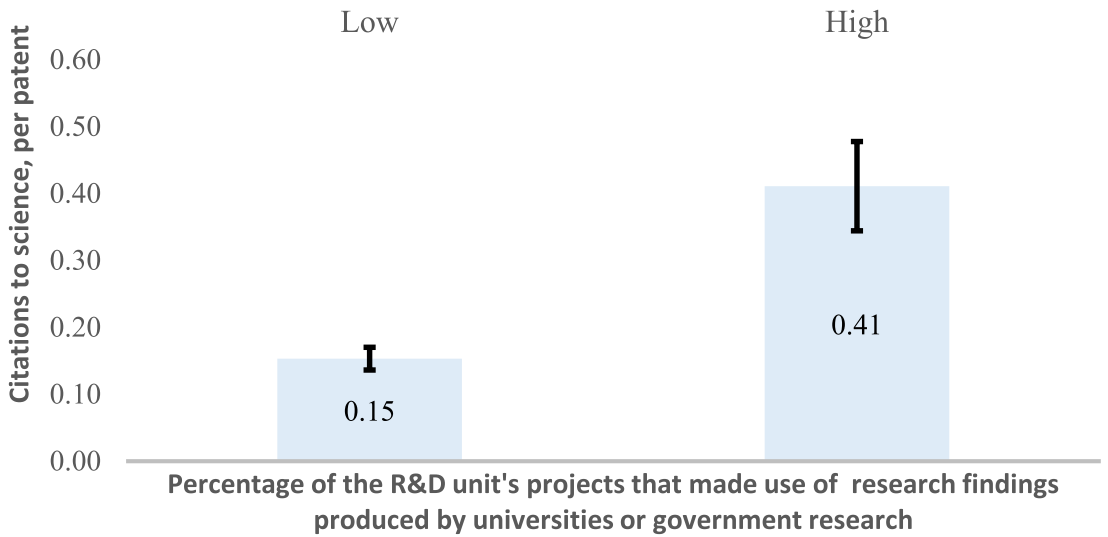
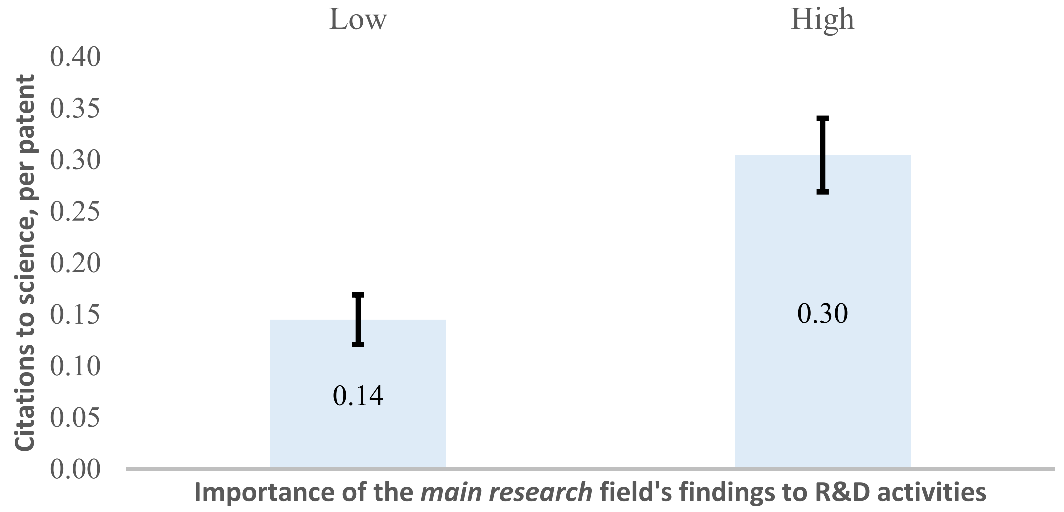


Figure C. Patent citation to WoS articles in main research field



Note: Figures 1A-1C present the relationship between patent citations to science and Carnegie Mellon survey (CMS) answers (Cohen et al., 2000) pertaining to the role of science in corporate R&D. Figures A & B present mean comparisons for citations to science, per patent 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 CMS Q18 (see main text for exact questions). Classification to high versus low is based on median value survey response: Low -636 firms; High-110 firms. Figure C presents mean comparisons for citations to science, per patent in main research field by low and high importance of the main research field's findings to the firm's R&D activities. Based on CMS Q.22. The sample is restricted to firms that indicated their main field in Q22 (excluding 'Others' category). Publications were classified to main fields based on WoS journal 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 vertical 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.

Table 1. Summary Statistics for Main Variables

VARIABLE	# Obs.	# Firms	Mean	Std. Dev.	Distribution 10th	50th	90th
Scientific publications count	42,994	3,008	18	100	0	1	22
Scientific publications stock	42,994	3,008	108	627	0	5	110
Patents count	54,274	4,090	24	138	0	2	34
Patents stock	54,274	4,090	123	667	1	7	179
R&D expenditures ($mm)	54,274	4,090	100	512	0.46	9	136
R&D stock ($mm)	54,274	4,090	446	2,411	1.4	37	567
Market value ($mm)	54,274	4,090	3,519	21,492	7	144	4,405
Tobin's Q	54,274	4,090	5	6	0	2	20
Sales ($mm)	54,274	4,090	2,265	11,414	3	124	3,726
Assets ($mm)	54,274	4,090	1,702	9,980	2	61	2,434

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 4,090 U.S. headquarter publicly traded ultimate owner parent companies (of which 3,008 are publishing companies) over the sample period 1980-2015. These firms have at least one year with positive R&D expenditures, at least one patent and 3 consecutive years of active records in Compustat during our sample period.

Table 2. Summary Statistics for Patent Citations to Corporate Publications

	(1)	(2)	(3)	(4)
VARIABLE	Number of firms with positive values	Citations per firm-year	Number of citing patents per firm-year	Number of cited publications per firm-year
Patent citations, all	1,172	10.5	8.4	6.0
Internal patent citations	561	1.6	1.1	1.1
External patent citations	1,114	8.9	7.4	5.2
External patent citations, corporate	764	3.5	2.9	2.1
External patent citations, RIVAL	721	0.8	0.7	0.5
External patent citations, TECH	758	2.2	1.8	1.5

Notes: This table provides summary statistics for patent citations to scientific publications by our sample firms. The sample is at the firm-year level and includes only firms with at least one publications that is cited by a patent. RIVAL and TECH measure citations from product market rivals and technology space, respectively.

Table 3. Mean Comparison Tests for Firms with High and Low Internal Patent Citations to Publications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		High Share of internal citations (above mean)			Low Share of internal citations (below mean)		
VARIABLE	(3) minus (6)	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
Publications flow/R&D expenditures($mm)	0.118**	376	0.539	0.678	796	0.422	0.682
Patents stock/R&D expenditures($mm)	0.038	376	0.458	0.488	796	0.420	0.481
R&D expenditures($mm)/Sales($mm)	1.097**	376	3.155	7.184	796	2.058	5.887

Notes: This table presents mean comparison tests for firms with high and low share of internal citations (defined as the ratio of internal citations and total patent citations received). The sample includes firms with publications that receive at least one patent citation. The unit of analysis is a firm. Yearly values are averaged over the period 1980-2015. * and ** indicate that the difference in means is significant at the 5% and 1% level, respectively.

Table 4. The relationship between internal and external patent citations to corporate science

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Sample: All publications (5 year citation lag)					
		Publications that are internally cited			Publications that are not internally cited		
VARIABLE	(3) minus (6)	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
Share cited by an external patent	0.272**	6,750	0.299	0.458	655,015	0.027	0.161
Number of external citations	0.742**	6,750	0.792	2.427	655,015	0.050	0.560

Notes: This table presents mean comparison tests for corporate publications that are internally cited *vs.* publications that are not internally cited. It examines the extent to which publications that are internally cited are also cited by outsiders. Unit of analysis is a corporate publication. The sample includes all publications published from 1980 through 2010. Patent citations are from patents granted up to 5 years post publication. ** denotes that the difference in means is significant at the 1 percent level.

Table 5. 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 Web of Science 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 6. Corporate Publication Over Time

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent variable	ln(1+Number of publications)						ln(1+Number of publications)		ln(Number of publications)		Number of publications	ln(1+Number of publications)
	Pooled	Within Firms	Cite-weighted	Excluding new journals	Large firm	Small firms	Period dummies	Excluding new journals	Exclude zeros	5-year cohorts	Negative Binomial	Trade secret protection
Time trend	-0.153**	-0.054**	-0.052**	-0.116**	-0.146**	-0.046**			-0.178**	-0.087**	-0.089**	-0.061**
	(0.005)	(0.013)	(0.013)	(0.013)	(0.022)	(0.017)			(0.031)	(0.016)	(0.028)	(0.014)
Time dummy for												
$2011 \leq \text{Year} \leq 2015$							-0.274**	-0.233**				
							(0.037)	(0.052)				
$2006 \leq Year \leq 2010$							-0.241**	-0.226**				
							(0.034)	(0.044)				
$2001 \leq Year \leq 2005$							-0.203**	-0.197**				
							(0.031)	(0.038)				
$1996 \leq Year \leq 2000$							-0.158**	-0.118**				
							(0.027)	(0.029)				
$1991 \leq Year \leq 1995$							-0.153**	-0.133**				
							(0.022)	(0.022)				
$1986 \leq Year \leq 1990$							-0.115**	-0.122**				
							(0.015)	(0.015)				
$1980 \leq Year \leq 1985$							Base	Base				
Time trend × IDD_{t-2}												0.005
												(0.009)
ln(*R&D stock*)$_{t-2}$	0.225**	0.128**	0.108**	0.093**	0.126**	0.075**	0.111**	0.121**	0.272**	0.244**	0.347**	0.077**
	(0.003)	(0.014)	(0.014)	(0.012)	(0.019)	(0.019)	(0.013)	(0.021)	(0.034)	(0.035)	(0.029)	(0.008)
ln(*Sales*)$_{t-2}$	0.077**	0.055**	0.045**	0.044**	0.243**	0.019**	0.051**	0.107**	0.147**	0.179**	0.183**	0.057**
	(0.002)	(0.006)	(0.006)	(0.005)	(0.018)	(0.007)	(0.005)	(0.015)	(0.017)	(0.025)	(0.016)	(0.006)
Pre-sample FE											0.730**	
											(0.030)	
Firm fixed-effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Dependent variable average:	15.9	15.9	18.8	11.1	25.4	2.1	14.1	29.2	35.2	22.5	18.4	15.9
Number of firms	4,090	4,090	4,090	4,090	2,048	2,042	4,090	1,172	2,696	2,833	3,161	4,090
Observations	46,094	46,094	46,094	46,094	27,187	18,907	46,094	19,678	20,767	7,110	34,838	46,094
R-squared	0.71	0.93	0.89	0.94	0.90	0.80	0.94	0.94	0.86	0.91	0.25	0.71

Notes: This table examines time trends in corporate scientific publication for the period 1980-2015. In Columns 1-6, 9 and 11 Trend is divided by 10 (i.e., presented in decennial units). In Column 10 Trend is per 5-year cohort. In Column 3, publications are weighted by citations received from other publications. Columns 5 and 6 restrict the sample to above and below median sales, respectively. Column 11 presents Negative Binomial estimates with firm fixed effects accounted for using pre-sample means of publications. Columns 4 and 8 exclude scientific publications from new journals post 1990. In Column 12 IDD 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). Columns 1-8 & 12 include a dummy variable that equals one for years with zero publications. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

Table 7. Corporate Publication Time Trend - by Industry

	Dependent variable: ln(1+Number of publications)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Industry:	Life science			IT, Software & Communication			Chemicals and Energy		
	Within Firms	Excluding new journals	Cite-weighted	Within Firms	Excluding new journals	Cite-weighted	Within Firms	Excluding new journals	Cite-weighted
Cohort dummies for:									
$2011 \leq Year \leq 2015$	0.172	0.066	0.226	-0.274*	-0.407**	-0.332**	-0.492**	-0.708**	-0.488**
	(0.125)	(0.117)	(0.140)	(0.114)	(0.121)	(0.114)	(0.117)	(0.126)	(0.127)
$2006 \leq Year \leq 2010$	0.215	0.127	0.222	-0.247*	-0.379**	-0.320**	-0.413**	-0.599**	-0.415**
	(0.119)	(0.114)	(0.135)	(0.105)	(0.111)	(0.105)	(0.097)	(0.108)	(0.115)
$2001 \leq Year \leq 2005$	0.318**	0.242*	0.327*	-0.249*	-0.353**	-0.289**	-0.334**	-0.480**	-0.348**
	(0.110)	(0.106)	(0.130)	(0.105)	(0.108)	(0.104)	(0.088)	(0.096)	(0.104)
$1996 \leq Year \leq 2000$	0.432**	0.356**	0.466**	-0.170	-0.289**	-0.242*	-0.187**	-0.367**	-0.157
	(0.100)	(0.097)	(0.119)	(0.095)	(0.103)	(0.099)	(0.067)	(0.076)	(0.083)
$1991 \leq Year \leq 1995$	0.475**	0.423**	0.487**	-0.188*	-0.247**	-0.238**	-0.131*	-0.238**	-0.167*
	(0.093)	(0.092)	(0.109)	(0.080)	(0.080)	(0.082)	(0.053)	(0.060)	(0.073)
$1986 \leq Year \leq 1990$	0.298**	0.280**	0.275**	-0.163**	-0.151**	-0.189**	-0.164**	-0.180**	-0.206**
	(0.068)	(0.067)	(0.087)	(0.058)	(0.057)	(0.064)	(0.042)	(0.044)	(0.060)
$1980 \leq Year \leq 1985$	Base	Base	Base	Base	Base	Base	Base	Base	Base
$\ln(R\&D\ stock)_{t-2}$	0.270**	0.206**	0.311**	0.138**	0.079*	0.088	0.171**	0.149**	0.160**
	(0.032)	(0.031)	(0.037)	(0.051)	(0.038)	(0.046)	(0.039)	(0.040)	(0.042)
$\ln(Sales)_{t-2}$	0.045**	0.036**	0.041**	0.041**	0.027*	0.035**	0.081**	0.071**	0.056*
	(0.011)	(0.010)	(0.012)	(0.014)	(0.011)	(0.013)	(0.022)	(0.024)	(0.026)
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable average:	40.859	30.877	57.859	22.123	14.369	24.222	21.786	17.910	24.920
Number of firms	668	668	668	706	706	706	282	282	282
Observations	6,522	6,522	6,522	6,173	6,173	6,173	4,187	4,187	4,187
R-squared	0.93	0.94	0.89	0.93	0.94	0.89	0.94	0.94	0.88

Notes: This table examines time trends in scientific publication for main industries. Industry classification is based on SIC codes (see Table B2 for detailed list). Columns 2 ,5 and 9 exclude scientific publications from new journals post 1990. All columns include a dummy variable that equals one for years with zero publications. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

Table 8. Internal Use and Publication Output

	Dependent variable: ln(1+number of publications)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Pooled	Between-firms	Within-firms	Publishing firms only	Controlling for patent self-citations	Share of internal citations	New vs. Old science	High vs. Low quality Publications	High vs. Low quality citing patents
ln(*Internal citation to publications*)$_{t-2}$	0.605**	1.389**	0.087**	0.081**	0.071**				
	(0.034)	(0.138)	(0.027)	(0.026)	(0.027)				
ln(*Patent self-citations*)$_{t-2}$					0.030**				
					(0.008)				
Internal citations /*Total NPL citations received* $_{t-2}$						0.116**			
						(0.044)			
ln(*Internal citation to publications*, NEW)$_{t-2}$							0.172**		
							(0.028)		
ln(*Internal citation to publications*, OLD)$_{t-2}$							-0.026		
							(0.027)		
ln(*Internal citation to publications*, High Quality)$_{t-2}$								0.088**	0.071**
								(0.026)	(0.027)
ln(*Internal citation to publications*, Low Quality)$_{t-2}$								0.018	0.037
								(0.042)	(0.026)
ln(*R&D stock*)$_{t-2}$	0.190**	0.166**	0.139**	0.154**	0.137**	0.142**	0.139**	0.140**	0.140**
	(0.011)	(0.025)	(0.017)	(0.019)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
ln(*Patent stock*)$_{t-2}$	0.103**	0.148**	0.044**	0.052**	0.030**	0.047**	0.044**	0.044**	0.044**
	(0.009)	(0.016)	(0.009)	(0.011)	(0.009)	(0.010)	(0.009)	(0.009)	(0.009)
ln(*Sales*)$_{t-2}$	0.068**	0.034**	0.072**	0.091**	0.071**	0.072**	0.072**	0.072**	0.072**
	(0.006)	(0.008)	(0.007)	(0.009)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Firm fixed-effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies (4-digits)	Yes	Yes	-	-	-	-	-	-	-
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable average:	15.9	8.7	15.9	19.8	15.9	15.9	15.9	15.9	15.9
Number of firms	4,090	4,090	4,090	3,008	4,090	4,090	4,090	4,090	4,090
Observations	46,076	4090	46,076	36,966	46,076	46,076	46,076	46,076	46,076
R-squared	0.68	0.69	0.89	0.88	0.89	0.89	0.89	0.89	0.89

Notes: This table presents OLS estimation results for the relationship between past internal patent citation to own publications and annual publications. Internal citations include patent citations up to year t-2 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, share of internal citations is defined as ratio of citations the firm's publications receive from own patents to total NPL citations received by the focal firm's publications. All specifications include a dummy variable that receives the value of one for firms that never published up to the focal year; a dummy variable that receives the value of one for firms without yearly granted patents; and a dummy variable that receives the value of one for firms without annual patent citations to own publications. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

Table 9. Knowledge Spillovers: External Patent Citation to Science and Publications

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: ln(1+Number of publications)					
	External citations	RIVAL & TECH	Only publishing firms	Mahalanobis RIVAL	Mahalanobis RIVAL, TECH	Time trend
ln(*Internal citation to publications*)$_{t-2}$	0.102**	0.087**	0.083**	0.087**	0.082**	0.089**
	(0.025)	(0.024)	(0.024)	(0.024)	(0.023)	(0.024)
ln(*External citation to publications* , RIVAL)$_{t-2}$		-0.144**	-0.147**	-0.104*	-0.147**	0.051
		(0.055)	(0.054)	(0.051)	(0.050)	(0.087)
ln(*External citation to publications* , TECH)$_{t-2}$		0.101*	0.098*	0.063	0.163**	0.106*
		(0.047)	(0.046)	(0.043)	(0.053)	(0.047)
ln(*External citation to publications* , RIVAL)$_{t-2}$ × Trend						-0.069*
						(0.032)
ln(*External citation to publications*)$_{t-2}$	-0.026					
	(0.024)					
ln(*R&D stock*)$_{t-2}$	0.140**	0.140**	0.154**	0.140**	0.139**	0.139**
	(0.017)	(0.017)	(0.019)	(0.017)	(0.017)	(0.017)
ln(*Patent stock*)$_{t-2}$	0.045**	0.045**	0.053**	0.045**	0.045**	0.044**
	(0.009)	(0.009)	(0.011)	(0.009)	(0.009)	(0.009)
ln(*Sales*)$_{t-2}$	0.072**	0.071**	0.091**	0.072**	0.071**	0.072**
	(0.007)	(0.007)	(0.009)	(0.007)	(0.007)	(0.007)
Trend						-0.073**
						(0.016)
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	No
Dependent variable sample average	15.9	15.9	19.8	15.9	15.9	15.9
Number of firms	4,090	4,090	3,008	4,090	4,090	4,090
Observations	46,076	46,076	36,966	46,076	46,076	46,076
R-squared	0.89	0.89	0.88	0.89	0.89	0.89

Notes: This table presents OLS estimation results for the relationship between external and internal patent citation to own publication and annual publications. All the citation variables include patent citations up to year t-2 to publications published up to the same year. In Column 1, external citation includes patent citations by any patent. Columns 2-6 include only external citation by corporate patents. RIVAL and TECH capture product market and the technology space proximities between a focal firm and its citing firms, respectively. All specifications include a dummy variable that receives the value of one for firms that never published up to the focal year; a dummy variable that receives the value of one for firms without yearly granted patents; and a dummy variable that receives the value of one for firms without annual patent citations to own publications. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

Table 10. Stock Market Value and Patent Citations to Corporate Science

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable:	ln(Market value)								ln(Tobin's Q)	
	Base	RIVAL and TECH	Mahalanobis RIVAL	Publishing firms only	Time trend	Life science	IT Software & Communication	Chemicals & Energy	RIVAL and TECH	Time trend
ln(*Internal citation stock*)$_{t-2}$	0.069**	0.069**	0.070**	0.065**	0.077**	0.161**	0.177**	-0.057	0.003	0.005
	(0.017)	(0.018)	(0.018)	(0.017)	(0.019)	(0.036)	(0.062)	(0.035)	(0.014)	(0.015)
ln(*External citation stock*)$_{t-2}$	0.009									
	(0.014)									
ln(*External citation stock*, RIVAL)$_{t-2}$ × Trend					-0.143**					-0.114**
					(0.014)					(0.011)
ln(*External citation stock*, RIVAL)$_{t-2}$		-0.076*	-0.065*	-0.078*	0.354**	0.141	-0.489**	-0.220**	-0.079**	0.253**
		(0.037)	(0.030)	(0.037)	(0.058)	(0.074)	(0.168)	(0.056)	(0.030)	(0.047)
ln(*External citation stock*, TECH)$_{t-2}$		0.086*	0.071**	0.072*	0.087*	-0.073	0.278	0.278**	0.075**	0.082**
		(0.033)	(0.026)	(0.033)	(0.035)	(0.070)	(0.162)	(0.053)	(0.027)	(0.028)
ln(*Patent stock*)$_{t-2}$	-0.014	-0.012	-0.012	-0.019*	-0.020*	-0.123**	0.060*	-0.019		
	(0.009)	(0.009)	(0.009)	(0.010)	(0.009)	(0.026)	(0.025)	(0.028)		
ln(*Publication stock*)$_{t-2}$	0.066**	0.067**	0.067**	0.072**	0.045**	0.059	0.044	0.051		
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.037)	(0.046)	(0.038)		
ln(*R&D stock*)$_{t-2}$	0.082**	0.082**	0.083**	0.074**	0.091**	0.178**	0.082*	0.083*		
	(0.011)	(0.011)	(0.011)	(0.012)	(0.011)	(0.035)	(0.039)	(0.032)		
ln(*Sales*)$_{t-2}$	0.224**	0.222**	0.223**	0.226**	0.216**	0.083**	0.162**	0.390**		
	(0.009)	(0.009)	(0.009)	(0.010)	(0.009)	(0.015)	(0.024)	(0.035)		
ln(*Assets*)$_{t-2}$	-0.001	-0.001	-0.000	-0.001*	-0.001	-0.001	0.001	0.001		
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)		
Patent stock$_{t-2}$/*Assets*$_{t-2}$									0.013**	0.013**
									(0.002)	(0.002)
Publication stock$_{t-2}$/*Assets*$_{t-2}$									0.003	0.002
									(0.002)	(0.002)
R&D stock$_{t-2}$/*Assets*$_{t-2}$									-0.001	-0.001
									(0.001)	(0.001)
Sales$_{t-2}$/Assets$_{t-2}$									0.037**	0.039**
									(0.003)	(0.003)
Trend					0.502**				0.037**	0.098**
					(0.015)				(0.003)	(0.010)
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Dependent variable sample average	3,979	3,979	3,979	4,796	3,979	4,003	5,992	6,913	4.381	4.381
Number of firms	3,987	3,987	3,987	2,959	3,987	663	683	281	3,970	3,970
Observations	41,217	41,217	41,217	33,686	41,217	6,097	5,447	3,923	40,784	40,784
R-squared	0.87	0.87	0.87	0.87	0.86	0.84	0.88	0.92	0.68	0.66

Notes: This table presents OLS estimation results for the relationship between citation stock and stock market value. Tobin's-Q is the ratio of market value to assets. All specifications include a dummy variable that receives the value of one for firms that never published up to the focal yea, and a dummy variable that receives the value of one for firms without citations to own publications up to the focal year. Robust standard errors (in brackets).

Table 11. Instrumental Variable Estimation: Federal and State R&D Tax Credit

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	ln(External citation, RIVAL)$_{t-2}$	ln(1+Number of publications)	Mahalanobis RIVAL	ln(External citation, RIVAL)$_{t-2}$	ln(Market value)	Mahalanobis RIVAL
	First Stage	Second stage	Second stage	First Stage	Second stage	Second stage
Predicted RIVAL patents$_{t-2}$	0.007**			0.001**		
	(0.000)			(0.000)		
ln(Internal citation to own publications)$_{t-2}$	-0.004	0.087**	0.086**	0.019**	0.082**	0.074**
	(0.006)	(0.015)	(0.015)	(0.007)	(0.022)	(0.021)
ln(External citation to own publications, RIVAL)$_{t-2}$		-0.305**	-0.154**		-0.533**	-0.232**
		(0.078)	(0.057)		(0.129)	(0.079)
ln(External citation to own publications, TECH)$_{t-2}$	0.525**	0.211**	0.093**	0.701**	0.435**	0.185**
	(0.011)	(0.053)	(0.035)	(0.009)	(0.101)	(0.056)
ln(Patent stock)$_{t-2}$	0.005**	0.046**	0.046**	0.006**	-0.010	-0.009
	(0.001)	(0.005)	(0.005)	(0.001)	(0.010)	(0.010)
ln(R&D stock)$_{t-2}$	0.000	0.140**	0.140**	0.000	0.083**	0.083**
	(0.001)	(0.008)	(0.007)	(0.002)	(0.013)	(0.013)
ln(Sales)$_{t-2}$	-0.001	0.071**	0.072**	0.001	0.222**	0.223**
	(0.001)	(0.004)	(0.004)	(0.001)	(0.011)	(0.011)
ln(Publication stock)$_{t-2}$				0.002	0.066**	0.068**
				(0.002)	(0.016)	(0.016)
ln(Assets)$_{t-2}$				0.001	-0.001	-0.001
				(0.001)	(0.001)	(0.001)
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Weak identification(Kleibergen-Paap)	F=404		F=338	F=319		F=432
Dependent variable sample average	0.30	15.9	15.9	2.5	3,979	3,979
Number of firms	3,734	3,734	3,734	3,603	3,603	3,603
Observations	45,720	45,720	45,720	40,833	40,833	40,833
R-squared	0.93	0.09	0.09	0.97	0.23	0.23

Notes: This table presents instrumental variable estimation results for the effect of RIVAL citations on publication and market value. Data on Federal and State R&D tax credit is based on Bloom et al. (2017). All specifications include a dummy variable that receives the value of one for firms that never published up to the focal year, and a dummy variable that receives the value of one for firms without citations to own publications up to the focal year. Robust standard errors (in brackets).

Table 12. Time Trend in External Patent Citations to Science

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	ln(# External citations per publication)	ln(# External corporate citations per publication)	ln(# Internal citations per publication)	ln(# same 4-digit SIC citations)	Share same-SIC citing firms	Citing firm ln(sales)	ln(# citations per citing firm)	ln(Number of new citing firms)
Time trend	0.070**	0.025**	0.005	0.056**	0.067**	0.259*	0.063**	-0.099**
	(0.014)	(0.009)	(0.010)	(0.019)	(0.020)	(0.117)	(0.011)	(0.013)
ln(*Sales*)	-0.008	-0.001	0.001	0.002	-0.008	0.295**	0.006	-0.024**
	(0.010)	(0.007)	(0.005)	(0.012)	(0.011)	(0.110)	(0.012)	(0.008)
ln(*Citations received from publications*)	0.028**	0.011**	0.012**	-0.020**	-0.004	0.054*	-0.055**	0.001
	(0.003)	(0.002)	(0.002)	(0.005)	(0.003)	(0.023)	(0.006)	(0.002)
ln(*External citation from patents*)				0.093**	0.001	-0.056	0.224**	0.081**
				(0.018)	(0.004)	(0.034)	(0.014)	(0.007)
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal subject category dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable sample average	4.181	0.274	0.11	0.740	0.368	23,465	1.610	0.45
Number of firms	1,338	1,338	1,338	612	612	612	612	976
Observations	58,982	58,982	58,982	8,641	8,641	8,641	8,641	26,655
R-squared	0.06	0.04	0.05	0.47	0.56	0.24	0.28	0.30

Notes: This table examines time trends in patent citations per publication. Unit of analysis is a publication-year for the sample period 1980-2010. The value of one is added to all count dependent variables and all their values are based on use up to 5 years from publication date. Columns 1-3 are restricted to publications with at least 1 external citation by any patent granted between 1980-2015. Columns 4-8 are restricted to publications with at least 1 external corporate citation (granted between 1980-2015 and no later than 5 years from the publication of the cited article). Trend is defined at the journal publication year and is divided by 10 (i.e., presented in decennial units). Total Forward citations by publications is the stock of forward citations from other publications during the sample period. Total external citation by patents is the stock of external patent citations per publication during the sample period. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

Table 13. Scientific Publications and R&D Productivity

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	ln(1+Number of citation-weighted patents)				ln(citation-weighted patents/xrd)
	Pooled	Within firms	Interaction Within firms	Interaction Within firms	Interaction Within firms
ln(R&D stock)$_{t-2}$	0.438**	0.304**	0.232**	0.237**	-0.120**
	(0.004)	(0.026)	(0.024)	(0.034)	(0.032)
ln(Citation-weighted publication stock)$_{t-2}$	0.154**	0.084**	-0.073	-0.068	-0.038
	(0.005)	(0.018)	(0.040)	(0.038)	(0.033)
ln(Internal citation stock)$_{t-2}$	0.276**	0.143**	-0.272*		
	(0.015)	(0.039)	(0.112)		
ln(External citation stock)$_{t-2}$	0.045**	-0.008	-0.102**		
	(0.012)	(0.029)	(0.032)		
ln(R&D stock)$_{t-2}$ × ln(Citation-weighted publication stock)$_{t-2}$			0.035**	0.035**	0.022**
			(0.008)	(0.007)	(0.007)
ln(R&D stock)$_{t-2}$ × ln(Internal citation stock)$_{t-2}$			0.055**		
			(0.016)		
Internal citation stock$_{t-2}$/publication stock$_{t-2}$				-3.348**	-3.591**
				(0.955)	(0.985)
External citation stock$_{t-2}$/publication stock$_{t-2}$				0.001**	0.001**
				(0.000)	(0.000)
ln(R&D stock)$_{t-2}$ × (Internal citation stock/publication stock)$_{t-2}$				0.580**	0.666**
				(0.171)	(0.187)
Firm fixed-effects	No	Yes	Yes	Yes	Yes
Industry dummies	Yes	-	-	-	-
Year dummies	Yes	Yes	Yes	Yes	Yes
Dependent variable sample average	34.675	34.675	34.675	48.074	0.662
Number of firms	4,090	4,090	4,090	2,868	2,775
Observations	46,094	46,094	46,094	32,185	29,638
R-squared	0.59	0.82	0.82	0.84	0.64

Notes: This table presents results OLS estimation results of a patent equation examining the relationship between R&D productivity and corporate science. Patents are weighted by number of citations received from other patents and publications are weighted by number of citations received from other publications. Columns 4-5 restrict the sample to firms with at least one publication during the sample period. All specifications include a dummy variable that receives the value of one for firms that never published up to the focal year; and a dummy variable that receives the value of one for firms without citations to own publications up to the focal year. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms.

Table B1. 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", 1980-2015
Publication stock	Publication stock per firm-year is calculated using a perpetual inventory method with a 0.15% growth rate and 15 percent depreciation rate (Hall et al., 2005), such that the Publication stock in year t for firm i is calculated by: $Publication\_stock_t=Pub_t+(1-\delta)Publications\_stock_{t-1}$, where Pub_t is the focal firm's publication count in year t. $\delta=0.15$.	Web of Science articles, covered in "Science Citation Index" and "Conference Proceedings Citation Index-Science", 1980-2016
Patent count	Patent count in year t for firm i	United States Patent and Trademark Office (USPTO) patents granted 1980-2015 from PatStat database
Patent Stock	Patent stock per firm-year is calculated using a perpetual inventory method with a 0.15% growth rate and 15 percent depreciation rate (Hall et al., 2005), such that the Patent stock in year t for firm i is calculated by: $Patent\_stock_t=Patent_t+(1-\delta)Patent\_stock_{t-1}$, where $Patent_t$ is the focal firm's patent count in year t. $\delta=0.15$.	United States Patent and Trademark Office (USPTO) patents granted 1980-2015 from PatStat database
Internal citations 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 by corporate sample firm patents 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. For Mahalanobis based measure firm-specific shares are weighted by the size of the segment-specific share they represent.	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 by corporate sample firm patents 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.
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 0.15% growth rate and 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 B2. SIC Classification by Main Industries

Category	Description	Related 4-digit SIC codes for our firm sample
Life science (Medical, Pharmaceuticals and Biotechnology)	Drugs, pharmaceuticals, biotech and medical devices- Manufacture, Sale & Services	2833 2834 2835 2836 5122 5912 8060 8071 8082 8090 8093 8731 8734
IT & Software & Communication	IT & Software - Development, Provider, Sale & Services; Telecom, Communication- system, equipment, services;	3661 3663 3669 4812 4813 4822 4832 4833 4841 4888 4899 5040 5045 7370 7371 7372 7373 7374
Chemicals & Energy	Chemicals- Manufacture & Sale. Energy: Electricity, Oil, Gas, Power station- including: utility, exploration, equipment, services, etc.	1000 1040 1220 1311 1381 1382 1389 1400 2800 2810 2820 2821 2840 2842 2844 2851 2860 2870 2890 2891 2911 2950 2990 3320 3330 3334 3341 3350 3357 3360 3390 3460 3470 4923 5051 5160

Table B3. Predicting Patents and R&D using Federal and State R&D Tax Credit

	(1)	(2)
Dependent variable:	ln(1+Number of patents)	ln(R&D)
ln(Federal tax credit component of R&D user cost)	-2.218**	-4.589**
	(0.469)	(0.346)
ln(State tax credit component of R&D user cost)	-0.341*	-0.334**
	(0.135)	(0.104)
Firm fixed-effects	Yes	Yes
Year dummies	Yes	Yes
Joint F-test of the tax credits	F=14.58	F=93.85
Dependent variable sample average	30.14	116.44
Number of firms	3,156	3,156
Observations	39,663	39,663
R-squared	0.82	0.92

Notes: Data on Federal and State R&D tax credit is based on Lucking, Bloom, Van Reenen (2018). Robust standard errors (in brackets).

DATA APPENDIX

This appendix describes the methodology used to construct our database of publicly listed American firms matched to assignees of patents from the United States Patent and Trademark Office (USPTO) and scientific publications from the Web of Science for the period 1980-2015.

We combine data from five main sources: (i) company and accounting information from U.S. Compustat 2018, (ii) scientific publications from Web of Science, (iii) patents and their non-patent literature (NPL) citations from PatStat; (iv) subsidiary data from ORBIS, and (v) mergers and acquisition data from SDC platinum.

We match: (i) corporate subsidiaries to Compustat ultimate owner (UO) firms; (ii) acquisition data to Compustat companies and their related subsidiaries; (iii) patent data to Compustat companies and their related subsidiaries; (iv) scientific publications to Compustat companies and their related subsidiaries; and (v) patent citations to scientific articles. We discuss the details of our methodology below.

A. ACCOUNTING DATA PANEL

Our methodology builds on and extends the NBER 2006 patent database (Hall et al., 2001) by extending the time period by a decade (now from 1980 to 2015) and implementing several methodological improvements for the complete sample period.

We start with all North American Compustat records obtained through WRDS in August 2018 and select companies with active records and positive R&D expenses for at least one year during our sample period, 1980-2015[1]. We exclude firms that are not headquartered in the United States based on their current headquarter location. After matching the remaining firms to patent assignees from the USPTO, we further restrict our sample to ultimate owner (UO) Compustat firms with at least one patent during our sample period. A UO firm enters the sample once it is publicly traded and remains in our data until the end of the sample period, unless it is acquired, dissolved, or taken private. All UO firms in our final sample have at least 3 consecutive years of active records in Compustat. Our UO firm level data are supplemented with subsidiary level data for each of our sample UO firms. Our final estimation sample consists of an unbalanced panel of 4,090 UO firms and 54,274 firm-year observations. Our unique UO firm identifier in the accounting data panel is labeled as PERMNO_ADJ_LONG. The process of defining an UO firm and its related subsidiaries is explained below.

We face several challenges when working with Compustat data, as following.

1) **Unique company identifier over time.** Compustat uses GVKEY to track companies over time[2]. However, a single company may correspond to multiple GVKEYs within the Compustat database due to changes in ownership and other accounting changes over the sample period (e.g., the pet food company Ralston Purina is listed under 2 different GVKEYs: (i) 1980-1993 under "RALSTON PURINA-CONSOLIDATED" (GVKEY 008935) and (ii) 1993-2000 under "RALSTON PURINA CO" (GVKEY 028701)). The Compustat database does not link related company identifiers making it difficult to track companies over time only based on GVKEY.

2) **Name changes.** While scientific publication and patent records contain the owner name at the time of their publication, companies appear in the Compustat file under their most current name with no records of previous names. Company names may change over the course of our sample period due to general name changes[3] and

[1] We define an active record as a year with positive common shares traded (CSHTR_F). We do this to avoid including years with data based on prospectus submitted by the focal company as part of the filing process before the firm became publicly traded.
[2] GVKEY code remains the same, regardless of changes in TICKER, CUSIP, and firm names and thus is preferred on the later as a firm identifier for Compustat records. Compustat database only provides the most recent TICKER, CUSIP and name for each security with no historical info available.
[3] e.g., name abbreviations (for example, "APPLIED MOLECULAR GENETICS INC" changed its name in 1983 to "AMGEN INC"),

M&As[4], including reverse takeovers[5]. A company with a name change (which might have been accompanied with an ownership change) without a corresponding change in its GVKEY in Compustat may cause us to assign the record incorrectly to its most recent holder for all sample years. Without historical information on the record's ownership, we cannot correctly link patents and scientific publications to their relevant financial records.

3) **Ownership structure.** A parent company and a majority owned subsidiary may have different identification numbers and records within Compustat. While innovative activities typically take place inside numerous subsidiaries, we aggregate the data to the UO level. Since Compustat database does not link parent companies and majority owned publicly traded subsidiaries, extant manual checks are required.[6]

4) **Changes in ownership.** Ownership of a firm can change throughout the sample period due to mergers, acquisitions and spinoffs[7]. While firms typically stop being traded independently after an M&A, their existing stock of publications and patents must be reassigned to the new owner. Moreover, in many cases, the acquiring entities continue to file patents and produce scientific publications post-acquisition. Compustat data do not provide information on ownership changes, thus we rely on SDC Platinum's M&A data to adjust our matched patent and publication data.

We implement the following procedures to manage these challenges.

I. NAME CHANGES

One of our key contributions is identifying name changes of Compustat firms over the sample years 1980-2015. To the best of our knowledge, this has not been done consistently for a broad range of companies across many industries over a third of a century. Past research mainly considers the name that appears for each record in the most recent Compustat file (CONM variable) as the relevant name for the complete period the security was traded. The variable CONM, however, is the current name of the Compustat record as of the date the file was downloaded with no historical name information provided by Compustat. As shown above, company name changes may not be accompanied by changes in the original GVKEY firm identifier on Compustat, leading to assigning a record to its most recent holder for the complete sample period. Instead of building on the recent Compustat name, we link our Compustat records to WRDS's "CRSP Monthly Stock" file, which records historical names for each month the security was traded and perform extensive manual checks using SEC filings to validate all related names for our sample period. We find that in our sample 30 percent of Compustat records have more than one related name[8]. Accounting for all historical names significantly improves the accuracy and scope of the matches we perform across various databases as well as the linkage to relevant financial data. We elaborate on our name change methodology below, using several examples.

Example 1: SEALED POWER and GENERAL SIGNAL

Based on Compustat records there were two different companies up to 1998:

1) GVKEY 9556, name: 1) SEALED POWER CORP (1962-1988); 2) SPX CORP (1988-1998) -name change

2) GVKEY 5087, name: GENERAL SIGNAL CORP (1950-1997)

[4] e.g., "WESTINGHOUSE ELECTRIC CORP" (GVKEY 011436) purchased "CBS INC" in 1995 and changed its own name to "CBS CORPORATION" in 1997 keeping the same GVKEY Compustat firm identifier.

[5] e.g., in 1993 the private company Dentsply International Inc acquired the public company GENDEX CORPORATION (GVKEY 013700) in a reverse takeover and became publicly traded under the "DENTSPLY INTERNATIONAL INC" name and the original GVKEY.

[6] e.g., Thermo Electron's publicly traded majority owned spun-out subsidiaries (all of which returned to be privately owned after 1999) need to be accounted under the parent company THERMO ELECTRON CORP (GVKEY 010530) for the complete period.

[7] e.g., "AT&T CORP" (GVKEY 001581) stopped being traded independently in 2005 after it was acquired by "SBC COMMUNICATIONS INC" (GVKEY 009899) which in turn changed its own name to "AT&T INC".

[8] This is comparable to the findings of Wu (2009), who finds that during 1925-2000 over 30% of CRSP-listed firms changed their names at some point after going public. For name changes occurring between 1980-2000 the paper finds that the top 3 reason for name changes are: (i) M&As & restructure activity (36%); (ii) Change in focus of operation (17%); (iii) brand or subsidiary name adoption (12%)

In 1998 SPX Corp acquired General Signal Corp in a reverse merger transaction and General's GVKEY (5087) became the new security of SPX traded under the new name “SPX CORP”. At the same time, the original SPX records are renamed retroactively in Compustat as “SPX CORP-OLD” and stopped being traded.

3) GVKEY 5087, name: SPX CORP (1998-2017)

Our approach is to treat these GVKEYs as two separate companies up to 1997 accounting for all relevant names (SEALED POWER CORP, SPX CORP for GVKEY 9556 and GENERAL SIGNAL CORP for GVKEY 5087) in our matches and to connect the SPX CORP name to General's original GVKEY (5087) only from 1998.

When we examine NBER 2006 patent dataset we find that the two companies are collapsed under the same company (same PDPCO id) and that for the purpose of Compustat accounting information General’s original GVKEY (5087) is used for the complete period while the original SPX GVKEY (9556) is disregarded:

Table A1. Data for SPX Corp in NBER 2006

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

Practically, this means that all the patents of SPX CORP are matched to General's financial data up to 1998. To verify we tracked the NBER files and confirmed that indeed SPX patents pre-1998 are matched to General's GVKEY. Moreover, patents related to “GENERAL SIGNAL CORP” (757 patents without considering related subsidiaries) as well as “SEALED POWER CORP” (36 patents without considering related subsidiaries) are located in the 2006 NBER raw patent match but are not assigned to any Compustat record.

We assume NBER did not track ownership of GVKEYs historically as there was no available data to do so at the time. However, as shown in this example using the current Compustat name can be misleading. Availability of data on historical name changes enables us to have a better understanding of the firms included in our sample and their origin. We are able to improve the accuracy of their match to the different databases (by using the complete history of firm names) and their linkage to relevant financial data.

Compiling historical names

To locate historical names, we use the WRDS’s “CRSP Monthly Stock” file, which includes historical monthly information on names for each security alongside its historical CUSIP code and a unique permanent security identification number assigned by CRSP, the PERMNO code, which is kept constant throughout the trading period regardless of changes in name or capital structure.[9] We compute for each name the starting and end year based on their trading dates in the “CRSP Monthly Stock” file.

Using WRDS “CRSP/Compustat Merged Database - Linking Table” we link each PERMNO to Compustat GVKEY code. As shown above a PERMNO can have multiple GVKEYs related to it, however CRSP also has cases where under the same GVKEY we find several PERMNO codes mainly due to significant M&As that occurred during the years or periods the firm was unlisted. Our main firm identifier PERMNO_ADJ builds on the original CRSP PERMNO id with several adjustments. (i) In cases where under the same GVKEY we find several PERMNO codes we replace it with one main PERMNO code[10] – for example, OWENS Corning GVKEY (008214) was split to two PERMNO codes 24811 and 91531 due to it being unlisted between 2003-2005, however we keep PERMNO_ADJ

[9] For example, while SPHERIX INC is related to 2 different GVKEYs (002237 for 1980-2013 and 018738 for 2013-current) it has a unique PERMNO code for the entire period (18148). Similarly, Google Inc PERMNO code is 90319 and it remains the same after the company reorganized as ALPHABET INC in 2015.

[10] In the final accounting data panel, we split firms based on big jumps in sales, patents or publications. For example, we split PERMNO_ADJ 66093 to the period before and after SBC Communications Inc acquired AT&T Corp and became AT&T Inc. PERMNO_ADJ_LONG is the final UO identifier in the accounting data panel after the split.

the same for the complete period (24811). (ii) We manually add a PERMNO_ADJ code for firms in our Compustat sample that did not appear in the "CRSP Monthly Stock" file[11].

We perform extensive manual checks on the name list, including (i) identifying companies with similar names[12]; (ii) adding from SEC fillings additional names that are missing from "CRSP Monthly Stock" file; and (iii) adjusting names manually – the "CRSP Monthly Stock" tends to abbreviate long words in the company name that it provides. We located those cases and manually corrected them to avoid mismatches.[13]

Standardizing firm names

Prior to matching, we standardize firm names to reconcile company names that may be spelled differently across the databases. We compose a standardization code used on both the source and the target names to increase the number of exact matches.

Each company name was first standardized by converting all strings to uppercase characters and cleaning all non-alphabetic characters as well as Compustat related indicators (e.g., -OLD, -NEW, -CL A) and other common words (e.g., THE).

Additionally, an important step in standardizing the company names is standardizing abbreviations. We formed a list that includes over 80 abbreviated words matched to their various original words. For example, LABORATORIES, LABORATORY, LABS, LABO, LABORATORIE, LABORATARI, LABORATARIO, LABORATARIA, LABORATORIET, LABORATORYS and LABORATORIUM were all abbreviated to "LAB". The list was compiled from the most frequently abbreviated words in WOS affiliation field (accordingly, the list is targeted to our sample). This list is presented in Table A2.

Table A2. 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

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

[11] There are differences between CRSP and Compustat coverage- for example, CRSP only includes firms listed in USA major exchanges and specifically excludes regional exchanges, while Compustat includes all 10-K filer firms in North America. Moreover, CRSP coverage for major exchanges has expanded gradually over the years (e.g., ARCA was only added from 2006).

[12] For instance, RACKABLE SYSTEMS INC (GVKEY 162907) changed its name to SILICON GRAPHICS INTL CORP after it acquired the public company SILICON GRAPHICS INC (GVKEY 012679) in 2009 – we need to make sure that we count SILICON GRAPHICS related publications and patents under Rackable's GVKEY only from 2009. Similarly, we need to distinguish between the original BIOGEN INC (GVKEY 002226) and the new BIOGEN INC (GVKEY 024468) that was formed only after the merger with IDEC PHARMACEUTICALS CORP in 2003.

[13] It is also worth mentioning that the "CRSP Monthly Stock" file reports acronym firm names with extra space between the initial letters (e.g., E G & G INC and not EG&G INC). This has to be taken into consideration when performing matches to other databases that do not use this format.

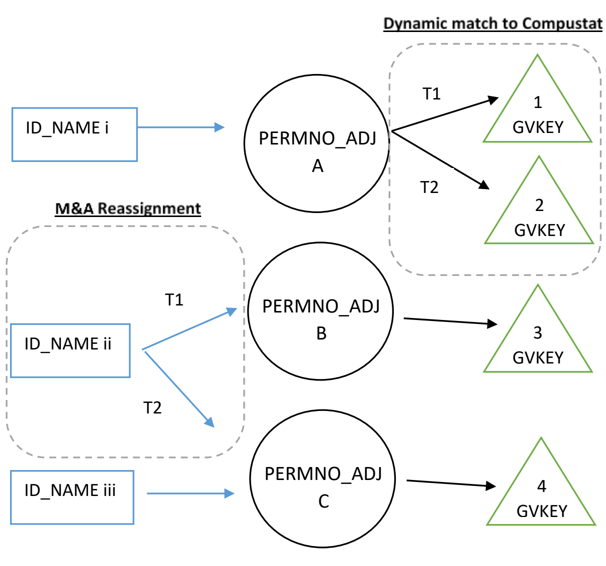
The last step in name standardization is to locate abbreviations that are commonly used by companies instead of their official name. For example, “INTERNATIONAL BUSINESS MACHINES CORP”, will also appear under its common abbreviation “IBM” and ‘’GENERAL ELECTRIC CO” under “GE”. We also add the names of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

Constructing the UO name list

All our matching is done at the firm name level. We assign each firm name a unique identifier ID_NAME and indicate the first and last year the name is relevant for a PERMNO_ADJ. We then perform dynamic matching of names to PERMNO_ADJ based on SDC’s M&A data. M&A reassignment includes up to five reassignments per name over the sample period (explained in further details below). Our UO historical standardized name list “B2B_UO_name_list.dta” including the dynamic reassignment will become publicly available for researches to match to their database of interest. Main variable of file “B2B_UO_name_list.dta” are described below:

Variable name	
NAME_STD	Historical standardized UO firm names (1980-2015) for firms that were included in our initial Compustat sample[14]
ID_NAME	Name ID unique at name_std-permno_adj1
PERMNO_ADJ$_{0\text{-}5}$	Owner firm id: up to 5 owners + "0" is usually the pre-IPO owner if applicable.
NAME_ACQ$_{0\text{-}5}$	Owner name
FYEAR$_{0\text{-}5}$	First year assigned to owner
NYEAR$_{0\text{-}5}$	Last year assigned to owner

Figure A1. Description of dynamic changes



Note: This figure illustrates the dynamic structure of the data. The dynamic reassignment accounts for: (i) changes in Compustat identification numbers (GVKEY), and (ii) M&A reassignment. Each name (ID_NAME) can be assigned throughout the sample period to more than one firm (PERMNO_ADJ) and each firm can be linked to more than one Compustat record (GVKEY).

[14] The data includes only names of UO parent firms included in our initial Compustat sample. Exceptional are names of top laboratories and names of majority owned publicly traded subsidiaries that appeared in our initial Compustat sample and were collapsed under the UO parent firm. In order to include subsidiary data, the user should match the name list to subsidiary databases such as ORBIS. The standardization code that was used to standardize the names will be available under NAME_STD.do. Standardized names include legal entity and other common words - in cases where users want to match to a cleaner version of the name, they should apply their own script to further clean the names. When matching the name list to other databases users should include extensive manual inspection to matched result.

II. DYNAMIC REASSIGNMENT

We build on the strategy used by NBER patent match (2006) to perform a dynamic reassignment for our subset of UO Compustat firms (see Figure A1). The dynamic reassignment accounts for: (i) changes in Compustat identification numbers (challenge 1 above) - dynamically matching Compustat accounting information for firms that are related to more than one GVKEY record, and (ii) M&A reassignment based on SDC data and construction of a complete name history for the period 1980-2015 (Challenge 4 above). For M&A reassignment, we include up to five ownership reassignments for each firm name that appears in our initial Compustat subsample and was acquired by another firm in our sample. Unless a name is reassigned to another PERMNO_ADJ it stays with the focal firm until the end of sample (or the firm's trading period). We dynamically reassign related patents and scientific publications of the acquired UO firm and its related subsidiaries to acquirer firms accordingly (will be discussed in more detail below).

Each PERMNO_ADJ is then linked to Compustat GVKEYs. For cases where there are changes in Compustat identification numbers over the sample period, we dynamically match PERMNO_ADJ to GVKEYs. In the final accounting data panel, we further split firms based on big jumps in sales, patents or publications. PERMNO_ADJ_LONG is the final UO identifier in the accounting data panel after the split.

Example 2: TIME-WARNER and AMERICAN ONLINE

Warner Communication and its subsidiaries were independent and publicly traded companies until their merger with Time Inc in 1989 when Time-Warner Inc was formed. In the second half of 2000 Time-Warner was merged with American Online to form AOL Time Warner. In 2003 the company dropped the "AOL" from its name and was renamed Time-Warner Inc. AOL remained a subsidiary until it was spun-out in 2009.

The NBER 2006 patent match reveals:

1) Warner Communication and its subsidiary related patents are correctly matched to WARNER COMMUNICATIONS INC (GVKEY 11284) up to the merger with Time Inc. However, they are not dynamically assigned after 1988 to Time Warner or any other company, implying that the patent stock and patent flow of Time-Warner (and later AOL Time-Warner) from patents related to Warner communication and its subsidiaries (e.g., Warner Bros, WEA Manufacturing (before it was acquired) – above 60 patents up to 2006) are below the true value after the acquisition in1989.

2) Time-Warner related patents from 1991 to 2000 (before the merger with American Online Inc in late 2000) are matched incorrectly to GVKEY 25056, which during those years was solely American-Online in original Compustat financial records. The current name of GVKEY 25056, TIME WARNER INC, which is likely to have misled NBER to link the Time Warner patents to it, was only adopted in 2003 when the "AOL" was dropped from the official name. Moreover, AMERICAN ONLINE INC and AOL related patents (152 patents up to 2006 based on NBER raw patent match) are not linked to any Compustat record. AOL-TIME WARNER related patents, on the other hand, are matched to a "Pro-Form" Compustat record that is active for only 2 years 1999-2000: AOL TIME WARNER INC-PRO FORM (GVKEY 142022). All of which imply that AOL Time Warner's flow of patents is below the true level throughout the period.

Having a complete history of names enables us to correctly identify each Compustat record and its origin and dynamically match each firm name in our sample to the correct financial records accordingly:

1) AMER ONLINE INC (and later AOL) is matched from 1980 until its spinout in 2009 to GVKEY 25056 and after to AOL INC (GVKEY 183920).

2) Warner Communication is matched up to the merger with Time Inc to WARNER COMMUNICATIONS INC (GVKEY 11284) and later dynamically transferred ending up in AOL -Time Warner GVKEY (25056) starting 2001.

3) AOL -Time Warner is matched to AOL -Time Warner GVKEY (25056) starting 2001 after the merger was approved.

4) As a side note- Time Inc is not included as an UO in our sample as it did not have R&D expenses, but it is included as a subsidiary name under the Time-Warner UO company.

Case study 3: PHARMACIA & UPJOHN and MONSANTO

In 1995 original Pharmacia merged with Upjohn to form Pharmacia & Upjohn. In 2000, original Monsanto merged with Pharmacia & Upjohn to form Pharmacia Corporation (New Pharmacia). Between 2000-2002 the new Pharmacia gradually spun off its agricultural operations to a new created subsidiary, Monsanto Company (New Monsanto). In 2003 the new Pharmacia was acquired by Pfizer and is now a wholly owned subsidiary of Pfizer.

Having a complete history of names enables us to correctly identify each Compustat record's historical ownership and dynamically match each firm name in our sample to its relevant financial records in each period. It also allows us to correctly compute patent stock and flow (Table A3). For instance, linking each patent to its correct financial record can be a concern for papers that link patents to market value, specifically those distinguishing different types of patents (e.g., high vs. low cited patents).[15]

[15] The following are additional examples: (I) Patents of Honeywell before the merger with Allied Signal (3,112 patents) are incorrectly linked to Allied Signal's GVKEY (001300) up to 1999, while the financial records of the original Honeywell Inc are disregarded (GVKEY 5693). (II) Patents of TELEDYNE INC (GVKEY 10405) pre-merger with the publicly traded ALLEGHENY LUDLUM CORP in 1996 (to form ALLEGHENY TELEDYNE INC, which in 1999 was renamed ALLEGHENY TECHNOLOGIES INC after TELEDYNE was spun-off as free-standing public company) are not linked GVKEY 10405 (634 patents up to 1999, of which 597 patents are pre-1996 merger). In addition, ALLEGHENY LUDLUM CORP's (GVKEY 13708) patents (254 patents, of which 240 patents pre-1996 merger) were not dynamically moved to TELEDYNE INC post-merger. This means that in 1996 (post-merger) the patent stock of GVKEY 10405 is missing at least 789 patents (not including related subsidiary patents). (III) For the new Biogen Inc (GVKEY 24468) NBER does not include patents of IDEC pharmaceuticals, who was the owner of the security before Biogen and IDEC merged in 2003 (40 patents)

Table A3. PHARMACIA & UPJOHN and MONSANTO dynamic match

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

III. AGGREGATING DATA TO THE UO FIRM LEVEL

To merge parent Compustat companies and their independent majority-owned publicly traded Compustat subsidiaries (Challenge 3 above), we locate related firms in our initial Compustat subsample based on name similarity as well as by matching the firm names to ORBIS subsidiary data. Where needed, we perform manual checks to confirm majority ownership using SEC 10-K filings. We aggregate the data to the UO parent-company level, accordingly.[16] We further link private subsidiaries to their UO firm based on ORBIS data (will be explained separately below). Accordingly, 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.

B. OWNERSHIP STRUCTURE

We rely on two main sources to construct ownership data: (i) SDC Platinum and (ii) ORBIS.

I. SDC M&A MATCH

Ownership changes of the UO Compustat firms in our sample are tracked through the SDC Platinum database with each firm name dynamically matched to up to five PERMNO_ADJ between years 1980 and 2015. Based on M&A deals available in SDC Platinum from 1980 to 2015, we downloaded detailed information on the acquirer and target firm names, acquirer and target firm CUSIPs, types of deals, execution dates, and percentage of shares owned after each transaction. We exclude deals that we identify as asset or business unit acquisitions.

We restrict the sample to deals involving a change in ownership that resulted in majority ownership (more than 50% of shares) for the acquirer. Execution dates are used to define the years a target firm begins or ends (in case of several acquisitions during the sample period) being owned by an acquirer. We then standardized both target and acquirer names similar to the standardization done for Compustat firm names. We match each deal's target and acquirer firm to our list of Compustat firms using both CUSIP numbers and all standardized historical names. It is important to use historical data as the information is recorded on SDC at the time of acquisition. We retain deals where both acquirer and target firms are matched to a Compustat firm in our sample. We track up to five ownership changes for each target firm name after it enters Compustat and one additional reassignment before it became publicly traded if relevant (i.e., if it was a subsidiary of another Compustat firm in our sample prior to its own IPO)[17]. We Assume that if a firm is acquired all of its patents and publications are transferred to the acquirer firm.

Example 4: NABISCO

During our sample period Nabisco has changed ownership four times. In 1981 Nabisco merged with the publicly traded company Standard Brands to form Nabisco Brands. Then, in 1985 R.J. Reynolds merged with Nabisco Brands to create RJR Nabisco, which eventually became Nabisco Group holding after the tabaco business was spun out in 1999. In 2000, Nabisco was acquired by Phillip Morris, which combined Nabisco with its Kraft brand. Finally, in 2001 Kraft (together with Nabisco) was spun out as a publicly traded company that later on became Mondelez International Inc. In our dataset all Nabisco related patents and publications are dynamically transferred between Compustat records and UO firms based on its ownership throughout the years:

[16] For example, GENZYME CORP (GVKEY 12233) - after verifying ownership on SEC filings: GENZYME MOLECULAR ONCOLOGY (GVKEY 117298), GENZYME TISSUE REPAIR (GVKEY 118653), GENZYME SURGICAL PRODUCTS (GVKEY 121742) and GENZYME BIOSURGERY (GVKEY 143176) are all accounted under their parent company GENZYME CORP (GVKEY 12233). While, GENZYME TRANSGENICS CORP (a.k.a. GTC BIOTHERAPEUTICS, GVKEY 028563) is a standalone alone company in our data as it was not majority owned by GENZYME CORP after it spun-off.

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

Table A4. Nabisco dynamic match

Years	relevant PERMNO_ADJ	related GVKEY	Original Compustat name for period	Current Compustat name	Comments
1981-1985	14533	7674	NABISCO BRANDS INC	NABISCO BRANDS INC	**1981**: Standard Brands company merged with Nabisco Inc (both publicly traded at the time) to form Nabisco Brands Inc.
1986-1999	14218	9113	R J R NABISCO INC	NABISCO GROUP HOLDINGS CORP	**1985**: R.J. Reynolds Industries merged with Nabisco Brands to form R J R Nabisco Inc (which eventually became Nabisco Group Holding when the tabaco business spun-out of RJR Nabisco in 1999)
2000	13901	8543	PHILIP MORRIS COS INC	ALTRIA GROUP INC	**2000:** Nabisco was acquired by Phillip Morris (today known as Altria Group Inc), which combined Nabisco with its Kraft brand.
2001-2015	89006	142953	KRAFT FOODS INC	MONDELEZ INTERNATIONAL INC	**2001:** Kraft together with its subsidiary Nabisco split from Phillip Morris as a new publicly traded company, which later on becomes Mondelez International Inc.

Examining NBER 2006, we find that for the purpose of Compustat accounting information all Nabisco related patents are linked to GVKEY 9113 from 1950 to 1999. Though the current name related to GVKEY 9113 is "Nabisco Group Holding Corp", based on the historical name information, we know that up to the merger of R.J. Reynolds with Nabisco it belonged solely to R.J. Reynolds. Reynold's patents, on the other hand (Over 419 patents for the period before it spun-out of RJR Nabisco and not including patents of acquired companies such as Heublein Inc), are not assigned by NBER to GVKEY 9113 and they are only being linked to Compustat records after the tabaco business spun-out of RJR Nabisco and became independently traded again under GVKEY 120877 (eventually merging with U.S. operations of British American Tobacco to form Reynolds American Inc). As a result, in 1998 the patent stock in NBER for GVKEY 9113 ("Nabisco Group Holding Corp") is 495 (consisting solely of Nabisco matched patents) whereas it should be 914 if it included R.J. Reynolds related patents. Furthermore, NBER does not dynamically move Nabisco's patent stock or account for its patent flow after 1999 when it was bought by Philip Morris and eventually became part of Kraft (a total of 529 Nabisco related patents up to 2006).

Table A5. Data for Nabisco in NBER 2006

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

Example 5: CHEMTURA CORPORATION

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

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

1) CROMPTON & KNOWLES CORP starting 1980
2) CK WITCO CORP starting 1999
3) CROMPTON CORP starting 2000

4) CHEMTURA CORP starting 2005

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

1) Uniroyal Chemical Corporation (acquired 1996)
2) Witco Corp (acquired 1999)
3) Great Lakes Chemical (acquired 2005)

When we examine NBER 2006 patent dataset we find that the only name that was matched to CHEMTURA CORP (GVKEY 3607) is "CHEMTURA CORP" (PDPASS 13245038). As the Chemtura name was adopted in 2005, only 1 patent was matched for that name. In addition, none of the acquired publicly traded companies were dynamically transferred to CHEMTURA CORP post acquisition. It is likely that a lack of information on historical names led NBER to rely on post-acquisition name (Chemtura) and thus prevented it from accounting for the M&A activities.

By considering all the previous names (without their subsidiaries and the acquired companies) related to GVKEY 3607: (i) Crompton & Knowles Corp; (ii) CK Witco Corp and (iii) Crompton Corp - based on the NBER raw patent match, we locate 220 additional patents up to 2006 that were not linked to any Compustat record that should be assigned to *Crompton & Knowles* (77 patents), *CK Witco* (26 patents)), and *Crompton* (117 patents). In addition, the acquired *Uniroyal Chemical Corp* has a patent stock of 379 patents in 2006 (out of which 185 patents are post acquisition) and the acquired Witco company has a patent stock of 405 in 2006 (out of which 62 patents are from post-acquisition period) and *Great Lake Chemicals* has a patent stock of 183 in 2006 (out of which 3 patents are in 2006, the year after the company was acquired).

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

II. <u>ORBIS SUBSIDIARY MATCH</u>

Due to the complexity of measuring large firms' innovative activities, which typically take place inside numerous subsidiaries, we aggregate the data to the ultimate-owner-parent-company level based on majority ownership. There are several challenges in keeping track of subsidiaries owned by UO Compustat firms, which may publish and patent in their own name. First, many of these subsidiaries are private and manual checks are sometimes required to verify which of the several similarly named companies was acquired by the firm. Furthermore, subsidiary ownership may change over the years. Companies may spin out their subsidiaries, some of which might go public or sold to other firms, where they are maintained as stand-alone subsidiaries and continue to patent or publish. Tracking subsidiary ownership is a main challenge we deal with and is explained below.

For firms with at least 50 patents over the sample years at the PERMNO_ADJ UO level, we collect all related subsidiary names using ORBIS and SEC filings as explained below.

We obtained ORBIS files for years 2002 to 2015, which provide us with snapshots of ownership structures for each of the years. One caveat is that the coverage of subsidiaries in the first few years of data files is not complete.

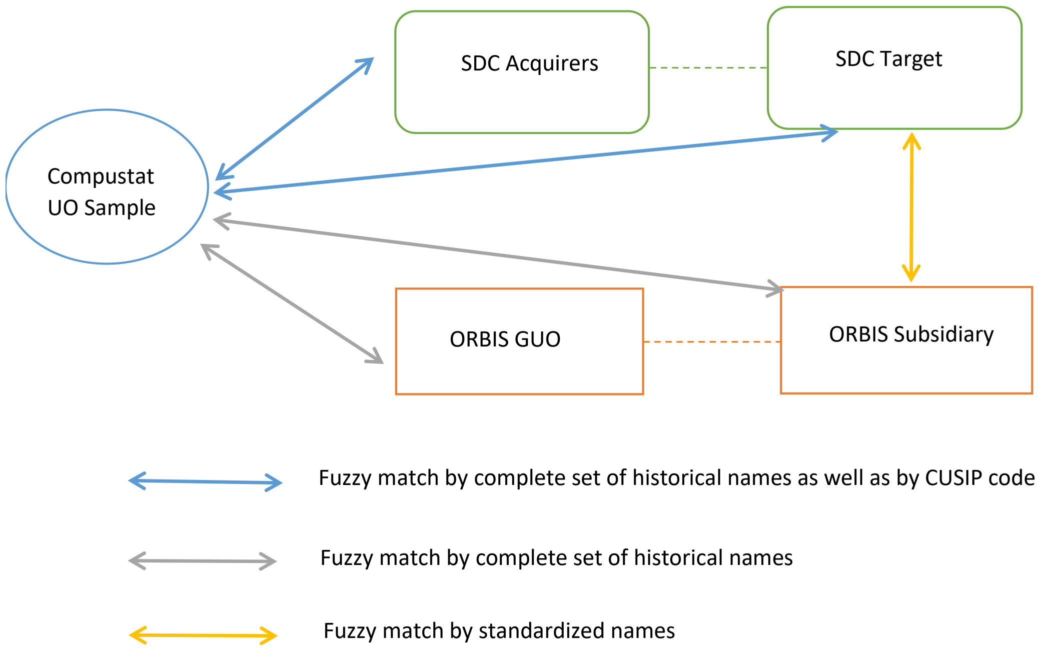
We start by standardizing the names of "Global Ultimate Owner" (GUO) firms and match the names to standardized historical Compustat names of firms with more than 50 patents at the UO level. Once again, it is important to use historical names for this match as the names that appear in each of our ORBIS files are as of the year the file was recorded.

Next, we link the subsidiaries of the successfully matched ORBIS owners to the PERMENO_ADJ of the corresponding parent firms. We restrict our sample to subsidiaries that are majority owned by the parent firm. After standardizing each subsidiary name similar to the standardization done for Compustat names, we obtain the first and last year it appears under a PERMENO_ADJ during 2002-2015. To avoid duplicated matching efforts, we drop subsidiaries that have the same organic name as the parent UO firm as they were already matched at the UO Compustat level. Some subsidiary names appear under more than one PERMNO_ADJ due to acquisitions throughout the years. Because we use yearly snapshots of ownership structure from ORBIS, we are able to account for name changes of subsidiaries over the period.

For firms that exit Compustat before 2002, we manually collect subsidiary names based on their latest available 10-K SEC filing.[18]

Since our sample starts from 1980 and the ORBIS files are only from 2002, we try our best to account for ownership changes of the subsidiaries for the years preceding 2002 using SDC and Compustat databases. We elaborate on our approach below.

Figure A2. Subsidiary matching Description



1) Fuzzy match between standardized subsidiary names and standardized SDC target name. For the matched result we locate:

a) Cases where the acquirer firm is a UO Compustat firm in our sample, which include:

(i) cases where the acquirer firm has the same PERMNO_ADJ as the parent firm of the subsidiaries. These cases confirm the direct acquisition of the subsidiary by the parent firm and provide us with the start date of the subsidiary (the year of acquisition) under the parent firm.

(ii) cases where the acquirer firm is a UO Compustat firm in our sample that was acquired by the parent firm of the subsidiary (i.e. the PERMNO_ADJ of acquirer and the PERMNO_ADJ of the parent of the subsidiary are related through acquisition). These cases confirm an indirect acquisition of the subsidiary by the parent firm and provide us with the start year of the subsidiary under the parent firm – i.e., year the ORBIS parent firm acquired the Compustat acquirer firm or the year of acquisition of the subsidiary (the latest).

b) cases where the acquirer firm is not a UO Compustat firm in our sample:

(i) if the CUSIP code of the UO parent firm related to the target firm (as indicated in SDC file) is the same as a CUSIP code related to the PERMNO_ADJ of the ORBIS parent of the subsidiary, it indicates that the subsidiary was acquired from the parent firm by the acquirer and provides us with the end date for the subsidiary under the parent firm – the year of acquisition.

[18] We do so for top 100 firms based on R&D spending.

(ii) for each acquirer firm's direct CUSIP code we search the complete SDC file for a deal where it was acquired by a firm with a CUSIP code related to the PERMNO_ADJ of the ORBIS parent of the subsidiary. These cases indicate indirect acquisitions, in which the subsidiary was acquired by a non Compustat sample firm that was itself acquired by the subsidiary's ORBIS parent firm. Such cases provide us the start year of the subsidiary under the parent firm –i.e., year the ORBIS parent firm acquired the non-Compustat acquirer firm or the date of acquisition of the subsidiary (the latest).

2) Fuzzy Match of cleaned subsidiary names

As the subsidiary name list includes closely related firm names with different legal entity, we use a clean version of the names that omits legal entity and other common words and we fuzzy match it to both clean Compustat names and the list of clean subsidiary names we found relevant acquisitions for in (1) above.

The fuzzy match to Compustat enables us to link each matched subsidiary name to the dynamic year sequence we constructed for UO Compustat firms. For the fuzzy match to the list of acquired subsidiaries, we adopt the relevant start &/end year we located in (1) above to all related subsidiaries.

3) As an additional check, we manually go over subsidiaries that did not match under 1) or 2) above and appear under more than one parent firm in our ORBIS sample or have more than 100 matched publications or patents. For these cases we check online sources and manually adjust their start and end date. Finally, for subsidiaries we were not able to identify start or end year, we make the assumption that they belong to the UO firm from its start date until end date. However, if the UO firm appeared in Orbis files for more than 3 years before the subsidiary was first linked to it, we adopt the first year the subsidiary is connected to the parent ORBIS firm as the start date of the subsidiary, under the assumption that it was acquired during that year by the parent firm.

All subsidiaries are assumed to move with their parent firm in cases where the parent firm is acquired, unless a subsidiary has a different end date from its parent firm or it is related to a the Compustat dynamic year sequence.

C. MATCHING

We perform several matches to construct our data including: (1) matching patent data to Compustat companies and their related subsidiaries; (2) matching scientific publications to Compustat companies and their related subsidiaries; (3) mapping patent citations to publications. We discuss each of these procedures below.

I. MATCHING PATENT DATA TO COMPUSTAT COMPANIES AND THEIR SUBSIDIARIES

We procced to matching our firm names to the assignees of the patents granted by USPTO using PatStat, which includes approximately 5.3 million patents for years 1980 through 2015.

We first remove published patent applications (i.e. publication numbers longer than 7 characters), non-utility patents, including Design, Reissue, Plant and T documents, and reexamination certificates. Next, we remove patents assigned to individuals or government entities (for example, an assignee that includes the string "DECEASED" or "U.S. DEPARTMENT"). We are then left with 4.97 million granted utility patents.

In order to compare assignee names to the standardized firm names in our sample, we standardize assignee names similar to the firm name standardization explained above. Assignee name standardization includes converting names to upper case, removing excess spaces, cleaning non-alphanumeric characters and replacing legal entity endings including commonly abbreviated terms (for example, "CORPORATION" is replaced with "CORP"; "LABORATORIES" and "LABS" with "LAB"). At the end of this process, we are left with 897K unique standardized assignee names.

The matching strategy includes several distinct steps. We begin by matching firm names to assignees using exact match. We then perform several fuzzy matching techniques to account for names that are slightly different but are in fact the same entities. Extensive manual checks at the assignee name and patent level were performed to ensure the quality of the matches.

(1) Exact Matching

Exact matching was conducted by comparing standardized assignee names to standardized firm names. Exact matching was carried out twice, once without and once with legal entities. The latter step was performed to account for firms whose names differ only by the legal entity. Extensive manual checks are performed to verify the matches. Special care was taken in cases where firm or assignee names are generic, when several different firms share a common portion of a name, or when firm names contain a common given or family name. To resolve ambiguities, we performed web searches and examined the actual patent documents.

(2) Fuzzy Matching

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

Fuzzy matching was performed by converting assignee and firm names into a term frequency–inverse document frequency (TF-IDF) matrix and calculating a cosine similarity score for each of assignee and firm name pair. To reduce the size of the task, the results were limited to assignees with at least 30 related patents. Given potential matches, final matches were selected manually by examining the top scoring assignee-firm pairs for each assignee.

To overcome the challenge of matching assignee names that include an additional description (such as division or location), we perform a cruder fuzzy matching between each standardized assignee name and the fully standardized firm names to locate cases where a firm name is a substring of an assignee name (for example, "EMERSON" is matched to "EMERSON CLIMATE TECH", a division within the firm). To reduce the size of the task, we perform this matching for the top 300 patenting firm names (based on previous matching procedures) and limit the search to assignee names with at least 5 related patents. Given potential matches, final matches are selected manually.

As an additional robustness check, we employed RAs to manually search the data for firm names with more than 100 patents (based on previous matching procedures) and find matching assignee names that might have been missed during exact and fuzzy matching.

Overall, this process yields 1.29 million patents mapped to the Compustat firms and their subsidiaries in our sample via patent number and NAME_ID. When a patent has several assignees, we match the patent to multiple firms and assign fractional patent ownership to each assignee (i.e., 1/number of assignees).

II. MATCHING SCIENTIFIC PUBLICATIONS TO COMPUSTAT FIRMS AND THEIR SUBSIDIARIES

After obtaining our initial subsample of firms and the various firm names 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," while excluding social sciences, arts and humanities articles.

Each publication record contains detailed information including title of the publication, authors, journal 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, "*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".

We apply a many-to-many fuzzy matching algorithm between each standardized name and the affiliation field for each publication (approximately 47 million publications, 8 million conference proceedings and 60 thousand names), while allowing for more than one firm to be matched to each publication (to allow for collaborative publications).

We first standardize the affiliation string of each Web of Science publication similar to the name standardization process explained above. The standardization removes special characters such as ampersands and words that indicate legal entity such as “INC” or “CORP”. It also ensures that common words such as “technology” and “chemicals” that frequently appear in company names are abbreviated in the same manner[19].

Second, we perform exact matching on company names and publication affiliation string using regular expressions. In addition, we calculate Levenshtein edit distances between company name-publication affiliation pairs. This is necessary because misspellings are common (e.g. BRISTOL MYERS SQUIBB misspelled as BRISTOL MEYERS SQUIBB). Since the company name in a publication affiliation is typically embedded in a longer string which includes buildings, street names, cities, zip-codes and country names, even correct matches will incur large distances. Therefore, we use a “partial” Levenshtein distance which calculates the edit distance between the shortest common segment between two strings. That is, our “partial” edit distance for company name “IBM” and affiliation “IBM Corp, SSD, San Jose, CA 951953 USA” will be zero, whereas a raw Levenshtein distance would be 35.

Third, we conduct manual checks on fuzzy-matched company name-publication affiliation pairs. In particular, we exclude matches from company names to eponymous buildings (e.g. Gillette Hall), schools (e.g. Heinz College), hospitals (e.g. Du Pont Children’s Hospital), charitable foundations, and endowed chairs. We also conduct manual checks on company-publication pairs with zero edit distances (exact matches) if the company names overlap with a common last name (e.g. ABBOTT), a geographic/historical location (e.g. BABYLON, BRISTOL), or branch of science & engineering such as “APPLIED MATERIALS” or “SEMICONDUCTOR”, as these are especially prone to being false positive matches. We also ensure that similar but distinct company names do not match to the same affiliation field (e.g. NORTHROP and GRUMMAN before their merger in 1994 are treated as distinct companies and will not match to NORTHROP GRUMANN). In cases where company names are the same, we 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.

At the end of this procedure, we obtain a match between a WOS record ID and our NAME_ID. We find 800 thousand unique articles from more than 14 thousand different journals that were published from 1980 through 2015, with at least one author employed by our sample of Compustat firms and their subsidiaries.

III. MATCHING NPL PATENT CITATIONS TO WEB OF SCIENCE ARTICLES

Patent citations to science are obtained from the Non-Patent Literature (NPL) citations section located at the front page of patents taken from the PatStat database. An example of a front-page patent citation to non-patent literature is provided in Figure A3. We obtain all NPLs related to patents granted in the period 1980-2015 (including corporate sample firm patents and non-corporate patents). We first remove NPL citations that we identify as non-publication references (e.g., reference that includes the string “PATENT ABSTRACT”, “U.S. APPLICATION NO.”, “US COURT”, “PRODUCT INFORMATION”, “DATA SHEET”, “WHITEPAPER”). We then proceed to match NPLs to corporate publications from Web of Science (approximately 10M citations and 800K corporate publications). This step presents a significant challenge due to differences in structure between NPL and publication string text-NPL patent citations to publications are highly non-standardized (see Table A6 for examples). We begin with a many-to-many match, allowing more than one publication to be matched to each NPL. 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 algorithm accounts for misspelling, unstructured text, incomplete references, and other issues that may cause mismatches.

[19] For instance, the word “technology” in a company name can be plural (“technologies”) or abbreviated (“technol”, “tech”). These special cases are abbreviated to “TECH” in our standardization code.

We will use the example below to illustrate the complication of the match and the algorithm we applied to detect a match.[20]

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 and (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 matching algorithm. After we standardize and clean the different stings, 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 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) name 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 impose 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, publication year, page numbers and volume, while accounting for typos, abbreviations (e.g., "INTERNATIONAL ELECTRONICS" vs. "ELECTRONICS") and differences in format of the string between the datasets (e.g., "VOL. 53, NO. 3" vs. "53(3)").

Finally, we use different combinations of the match results for the different features (title, authors and journal information) according to their relative importance, in order to determine a final match. We perform extensive manual checks to confirm matches[21]. At the end of this procedure, we obtain unique identification numbers for the citation, the citing patent and the cited publication.

We further differentiate between internal citations (patent citation by the focal firm's patent to its own publication) and external citations (patent citation to the focal firm's publication by other corporate and non-corporate patents). For external citations from corporate sample firms we construct technology and segment proximity measures between the cited and the citing firms as explained below. For the purpose of classifying internal or external citation, we rely on the original UO firm the publication was affiliated with at its publication year[22].

---

[20] The following example (first line in Table A6) 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.

[21] There are several cases where the NPL reference is a citation to a working paper and we are able to match it to the final published paper that appears on WOS data base – we consider those as matches.

[22] i.e., if Company B acquires Company A (let's assume A is a Compustat firm in our sample pre-acquisition): Citations by B's patents post-acquisitions to A's publications that were published pre-acquisition are classified as external citations. However, citation from B's patents to A's publications published post-acquisition are classified as internal citations. Moreover, as opposed to publication and patent stock variables, citations do not move dynamically between firms in case of acquisition.

Following the above procedures, we obtain 272,387 unique patent citations to 63,668 unique corporate publications (8 percent of corporate publications), by 159,279 citing patents (including corporate sample firm patents and non-corporate patents). Of the cited publications, 80 percent receive only external citations and the remaining receive at least one internal citation. The temporal structure of citations and publications are illustrated in Figure A6.

REFERENCES

[1] Hall, B.H., Jaffe, A.B. and Trajtenberg, M., 2001. The NBER patent citation data file: Lessons, insights and methodological tools (No. w8498). National Bureau of Economic Research.
[2] Wu, Y., 2010. What's in a name? What leads a firm to change its name and what the new name foreshadows. Journal of Banking & Finance, 34(6), pp.1344-1359.

Table A6. Matching Citations to Scientific Publications - Examples

Citation	Publication info: Title	Publication info: Authors	Publication info: Journal information	Comment
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

Figure A3. 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

(12) **United States Patent**
Liu et al.

(10) **Patent No.:** **US 6,468,596 B1**
(45) **Date of Patent:** **Oct. 22, 2002**

(54) **LASER-ASSISTED IN-SITU FRACTIONATED LUBRICANT AND A NEW PROCESS FOR SURFACE OF MAGNETIC RECORDING MEDIA**

(75) Inventors: **Youming Liu**, Palo Alto; **Jialuo Jack Xuan**, Milpitas; **Xiaohua Shel Yang**, Fremont; **Chung-Yuang Shih**, Cupertino; **Vidya K. Gubbi**, Milpitas, all of CA (US)

(73) Assignee: **Seagate Technology LLC**, Scotts Valley, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/577,674**

(22) Filed: **May 25, 2000**

Related U.S. Application Data

(60) Provisional application No. 60/144,357, filed on Jul. 15, 1999.

(51) **Int. Cl.**[7] **C08F 2/48**; C08J 7/18; C23C 14/30

(52) **U.S. Cl.** **427/508**; 427/554; 427/596

(58) **Field of Search** 427/510, 554, 427/555, 556, 597, 127, 226, 258, 261, 264, 270, 271, 402, 508

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,674,340 A 7/1972 Jacob et al. 350/157
3,764,218 A 10/1973 Schedewie 356/118
3,938,878 A 2/1976 Fox 350/150

(List continued on next page.)

FOREIGN PATENT DOCUMENTS

OTHER PUBLICATIONS

P. Baumgart et al., "A New Laser Texturing Technique For High Performance Magnetic Disk Drives" IBM storage Systems Division and IBM Almadon Research Center, San Jose, CA.
D. Kuo et al., "Laser Zone Texturing on Glass and Glass–Ceramic Substrates" Seagate Recording Media, Fremont, CA.
P. Baumgart et al., "Safe Landings: Laser Texturing of High–Density Magnetic Disks" IBM Corp., *Data Storage* 1996.
A. Tam et al., "Laser Cleaning Techniques for Removal of Surface Particulates" IBM Research Division, San Jose, *Journal of Applied Physics* 71 (7), Apr. 1, 1992, pp. 3515–3523.
K. Johnson et al., "In–Plane Anisotropy in Thin–Film Physical Origins of Orientation Ratio (Invited)" IBM Storage Systems Division, San Jose, CA, *IEEE Transactions on Magnetics* vol. 31, No. 6, Nov. 1995, pp. 2721–2727.
J. Miles et al., "Micromagnetic Simulation of Textured Induced Orientation in Thin Film Media" the University of Manchester, Manchester, M13 9PL, U.K., *IEEE Transactions on Magnetics* vol. 31, No. 6, Nov. 1995, pp. 2770–2772.
C. Kissinger et al., "Fiber Optic Probe Measures Runout of Stacked Disks" B.W. Brennan Associates, *Data Storage* Jul./Aug. 1997.

Primary Examiner—Shrive P. Beck
Assistant Examiner—Eric B. Fuller
(74) *Attorney, Agent, or Firm*—McDermott, Will & Emery

(57) **ABSTRACT**

A magnetic recording medium is formed with enhanced tribological performance by applying a raw, unfractionated lubricant having a wide molecular weight distribution over a disk surface and treating the deposited lubricant with a laser light beam to effect in-situ fractionation of the lubricant to a very narrow molecular weight distribution. Embodiments of the present invention also include laser treating a deposited lubricant to increase the thickness of the bonded lube layer.

IEEE TRANSACTIONS ON MAGNETICS, VOL. 31, NO. 6, NOVEMBER 1995 2721

In-Plane Anisotropy in Thin-Film Media: Physical Origins of Orientation Ratio (Invited)

Kenneth E. Johnson, Mohammad Mirzamaani, and Mary F. Doerner
IBM Storage Systems Division, San Jose, CA 95193

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

(12) **United States Patent**
Cabral, Jr. et al.

(10) **Patent No.:** **US 7,193,323 B2**
(45) **Date of Patent:** **Mar. 20, 2007**

(54) **ELECTROPLATED COWP COMPOSITE STRUCTURES AS COPPER BARRIER LAYERS**

(75) Inventors: **Cyril Cabral, Jr.**, Ossining, NY (US); **Stefanie R. Chiras**, Peekskill, NY (US); **Emanuel Cooper**, Scarsdale, NY (US); **Hariklia Deligianni**, Tenafly, NY (US); **Andrew J. Kellock**, Sunnyvale, CA (US); **Judith M. Rubino**, Ossining, NY (US); **Roger Y. Tsai**, Yorkstown Heights, NY (US)

(73) Assignee: **International Business Machines Corporation**, Armonk, NY (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **10/714,966**

(22) Filed: **Nov. 18, 2003**

(65) **Prior Publication Data**

US 2005/0104216 A1 May 19, 2005

(51) **Int. Cl.**
H01L 23/48 (2006.01)
H01L 23/52 (2006.01)

(52) **U.S. Cl.** **257/751**; 257/752; 257/762

(58) **Field of Classification Search** 257/751–753, 257/758, 759, 761–763

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,695,810 A 12/1997 Dubin et al.
6,168,991 B1 * 1/2001 Choi et al. 438/254
6,323,128 B1 11/2001 Sambucetti et al.
6,342,733 B1 1/2002 Hu et al.
6,528,409 B1 3/2003 Lopatin et al.
6,573,606 B2 * 6/2003 Sambucetti et al. 257/762
2003/0010645 A1 1/2003 Ting et al.
2003/0075808 A1 * 4/2003 Inoue et al. 257/774

OTHER PUBLICATIONS

A. Kohn, et al., "Characterization of electroless deposited Co(W,P) thin films for encapsulation of copper metallization" Materials Science and Engineering A302 (2001) pp. 18-25.
C.-K. Hu, et al., "Reduced electromigration of Cu wires by surface coating" IBM T.J. Watson Research Center, Yorktown Heights, New York, 2002.

(Continued)

Primary Examiner—Hung Vu
(74) *Attorney, Agent, or Firm*—Connolly Bove Lodge & Hutz, LLP; Robert M. Trepp

(57) **ABSTRACT**

A composite material comprising a layer containing copper, and an electrodeposited CoWP film on the copper layer. The CoWP film contains from 11 atom percent to 25 atom percent phosphorus and has a thickness from 5 nm to 200 nm. The invention is also directed to a method of making an interconnect structure comprising: providing a trench or via within a dielectric material, and a conducting metal containing copper within the trench or the via; and forming a CoWP film by electrodeposition on the copper layer. The CoWP film contains from 10 atom percent to 25 atom percent phosphorus and has a thickness from 5 nm to 200 nm. The invention is also directed to a interconnect structure comprising a dielectric layer in contact with a metal layer; an electrodeposited CoWP film on the metal layer, and a copper layer on the CoWP film.

18 Claims, 6 Drawing Sheets

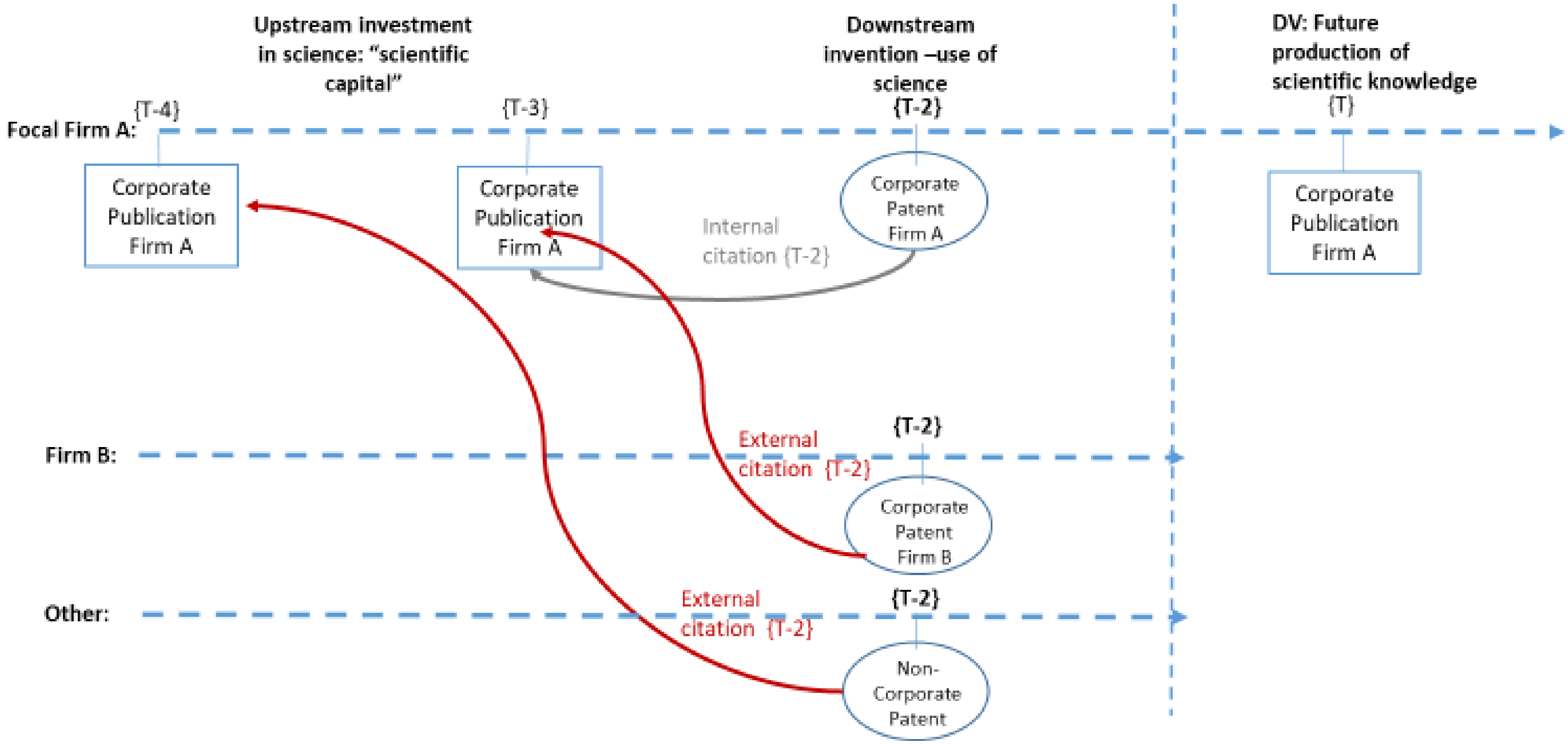
APPLIED PHYSICS LETTERS VOLUME 81, NUMBER 10 2 SEPTEMBER 2002

Reduced electromigration of Cu wires by surface coating

C.-K. Hu,[a] L. Gignac, R. Rosenberg, E. Liniger, J. Rubino, and C. Sambucetti
IBM T. J. Watson Research Center, Yorktown Heights, New York 10598
A. Domenicucci and X. Chen
IBM Microelectronics Division, Hopewell Junction, New York 12533
A. K. Stamper
IBM Microelectronics Division, Essex Junction, Vermont 05452
(Received 28 May 2002; accepted for publication 11 July 2002)

Note: this figure presents examples of front-page patent reference to non-patent literature. Below each patent reference is the related scientific publication that is being cited. Example (i) is an external patent citation to IBM's publication and example (ii) is an internal patent citation to IBM's publication.

Figure A4. Timeline- Production and Use of Research



Note: this figure illustrates the temporal structure of citations and publications. At time T-2 the focal firm (Firm A) has: (i) one Internal citation and (ii) two external citation- one from a patent filed by a sample Compustat firm (Firm B) and another from non-Compustat assignee (other).