NBER WORKING PAPER SERIES

# THE RISE OF SCIENTIFIC RESEARCH IN CORPORATE AMERICA

Ashish Arora
Sharon Belenzon
Konstantin Kosenko
Jungkyu Suh
Yishay Yafeh

Working Paper 29260
http://www.nber.org/papers/w29260

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2021

We thank seminar participants at Boston University, the ASSA 2021 annual meeting, Strategy Science Conference and Wharton Technology and Innovation Conference, as well as Tim Simcoe, Bhaven Sampat and Tarun Khanna for helpful comments and suggestions. We thank Gioia Blayer for excellent research assistance. All remaining errors are our own. Arora, Belenzon and Suh acknowledge support from the Fuqua School of Business, Duke University. Belenzon and Yafeh gratefully acknowledge financial support from the Israel Science Foundation, Grant No. 963-2020. Kosenko is grateful for support from the Bank of Israel. Yafeh also acknowledges financial support from the Krueger Center at the School of Business Administration of the Hebrew Univeresity. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research or the Bank of Israel.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w29260.ack

The Rise of Scientific Research in Corporate America
Ashish Arora, Sharon Belenzon, Konstantin Kosenko, Jungkyu Suh, and Yishay Yafeh
NBER Working Paper No. 29260
September 2021
JEL No. N8,N82,O32

**ABSTRACT**

Corporate science in America emerged in the interwar period, as some companies set up state-of-the-art corporate laboratories, hired trained scientists, and embarked upon basic research of the kind we would associate today with academic institutions. Using a newly assembled dataset on U.S. companies between 1926 and 1940 combining information on corporate ownership, organization, research and innovation, we attempt to explain the rise of corporate research. We argue that it was driven by companies trying to take advantage of opportunities for innovation made possible by scientific advances and an underdeveloped academic research system in the United States. Measuring field-specific scientific backwardness in several different ways, we find that large firms, business group affiliated firms, and firms close to the technological frontier were more likely to initiate scientific research. We also find that companies in monopolistic or concentrated industries were more likely to engage in basic research. Corporate research was positively correlated with novel and valuable patents, and with market-to-book ratios. For companies choosing to do so, investment in corporate research seems to have paid off. The results shed light on the link between corporate organization, market structure and corporate science.

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

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

Konstantin Kosenko
Bank of Israel
Jerusalem 91907, Israel
konstantin.kosenko@boi.org.il

Jungkyu Suh
Duke University
Fuqua School of Business
durham, nc 27708
jungkyu.suh@duke.edu

Yishay Yafeh
School of Business Administration
The Hebrew University, Mt. Scopus
Jerusalem 91905, ISRAEL
and CEPR and ECGI
msyafeh@huji.ac.il

# 1 Introduction

Vannevar Bush asserted that science is the "pacemaker of technological progress" (Bush, 1945). Indeed, advances in basic science have served as critical inputs for firms introducing radically novel products such as the transistor. However, in contrast to Bush's vision, private firms often carry out basic research themselves. At the same time, since scientific knowledge easily spills over to rivals (Arora, Belenzon, & Sheer, 2020b; Bloom, Schankerman, & Van Reenen, 2013), the reasons why firms may engage in basic science remain an enduring puzzle (Nelson, 1959a). Firms could decide to invest in science to increase absorptive capacity (Cohen & Levinthal, 1989), to attract competent inventors (Sauermann & Cohen, 2010; Stern, 2004), as well as to signal their product quality to buyers (Azoulay, 2002; Hicks, 1995). The present paper explores the role of underdeveloped U.S. universities in the interwar period in inducing firms close to the technological frontier to start investing in scientific research.

By the end of World War I, advances in technology were increasingly science-based, especially in industries such as chemicals, plastics, synthetic fibers, electricity and communications. Inventions required more than just trial and error and tinkering; a deeper scientific understanding of the materials, machines and the natural forces at work was needed. General Electric, for instance, had exhausted trial and error methods to reduce the blackening that occurred on the surface of the lightbulb. Irving Langmuir, an American chemist hired by GE after after a PhD from the University of Göttingen, made fundamental discoveries in surface chemistry in the course of diagnosing the source of the blackening as evaporation from the tungsten filament under extremely high temperatures. This led to a radically different solution to what had been proposed by technologists at the time: instead of trying to create a better vacuum, Langmuir proposed to fill the bulb with inert gases that would scatter the evaporated particles.[1] Elucidating the science behind existing products also allowed firms to develop valuable new products: DuPont, an explosives manufacturer that had by the 1920s diversified into paints, rubbers and rayon, all of which were based on polymeric materials, invested in polymer science, which yielded blockbuster products such as nylon and polyester.

Around the same time period, in the early decades of the twentieth century, several large American firms had reached the technological frontier in their respective markets and realized the need for a more fundamental understanding of how their flagship products worked. However, as we show below, the scientific knowledge available from American universities was inadequate for firms at the technological

[1]Langmuir recalled that his work on surface chemistry allowed him to "conclude with certainty that the life of the lamp would not be appreciably improved even if we could produce a perfect vacuum" (Reich, 1983).

frontier because of the relative backwardness (in comparison with Europe) of U.S. universities, especially in certain scientific disciplines within physics and chemistry.[2] We argue below that the firms that felt this need most acutely were large, publicly traded firms, which relied on novel inventions for their growth. There is also some evidence to suggest that they operated in relatively concentrated markets, where growth through mergers and acquisitions was increasingly circumscribed by antitrust laws.[3] Continued sources of growth had to be found in new products and markets. Because the public science system was underdeveloped, these firms chose to invest internally. Because they operated in concentrated markets, spillover of knowledge to rivals was apparently not a major concern.

The relationship between the state of academic research and corporate investment in science is complex. Academic science could complement or substitute for internally generated research by companies. Moreover, because academic science is potentially available to all firms in a market, the nature of the strategic interactions among competitors matters as well. Finally, human capital is jointly produced along with academic research. The combined effect is complicated and likely to be historically contingent. We develop a simple conceptual framework to study the private returns to investment in research. We distinguish between scientific knowledge and innovation. Innovation–the introduction of new products and processes–is the source of profits. Scientific knowledge, both from universities and from internal research, reduces the cost of innovation. Leading firms that are more dependent on innovation derive greater returns from investing in research. However, their incentives also depend upon the nature of strategic interactions, as well as on the state of academic science. Under some conditions, the incentives to invest may actually be higher when the supply of academic science is low.[4]

We study these issues empirically using newly assembled historical data, which combine novel data on corporate innovation with ownership and financial information during the interwar period. We assemble three measures of innovation: publications, industrial lab employment and patents. We collect data on scientific publications authored by researchers employed at corporations using Microsoft Academic Graph (MAG), including the scientific discipline they publish in, as well as citations received by other publications and patents. We match firms to the *Industrial Research Laboratories of the United States* survey (hereafter "the IRL directory") of corporate science laboratories. A total of 306 firms (65%) in

[2] Indeed, American oil companies had sponsored the establishment of chemical engineering as a discipline at MIT because they needed to increase refining efficiency.

[3] DuPont had grown by acquiring smaller innovative firms such as Hercules, and Giant, which had the license from Nobel. Following the breakup of the rail and oil trusts, DuPont had to divest its existing operations in smokeless powder in 1911.

[4] The model does not consider the effect of the production of human capital. See Arora, Belenzon, and Patacconi (2019) who analyze how the joint production of knowledge and human capital conditions the incentive of a single incumbent in a model where the incumbent may potentially buy inventions from startups.

our sample of patenting firms operate a research lab, while 196 (42%) publish at least once during our sample period. We complement the information on corporate science by matching our sample firms to U.S. patents[5] and link them to in-text citations made to scientific publications published during this period (Marx & Fuegi, 2020).[6] Our dataset includes not only the number of patents by company but also measures of patent quality, such as forward patent citations received and stock market values from Kogan, Papanikolaou, Seru, and Stoffman (2017).

To measure the public science that firms could draw upon, we develop measures of relative American backwardness by scientific field, as well as measures of a firm's proximity to the technological frontier. Scientific backwardness is measured using three alternative methods. In the featured method, we calculate for each scientific field the number of academic publications by European authors in the field, divided by all publications by U.S.-based scientists in the same field.[7] Scientific fields are weighted by their relevance to the firm's patenting activity. A firm's proximity to the technological frontier is measured through prior-art citations a firm's patents receive, in-text citations its patents make to scientific articles and the ratio of patents to assets. The combined data set comprises 469 firms and 7,035 firm-year observations.

We combine the data on corporate innovation with firm-level panel data, which include detailed financial information and ownership. We extend a dataset constructed by Kandel, Kosenko, Morck, and Yafeh (2019) (hereafter "KKMY"), which includes both publicly traded and private firms that were active in upstream R&D (academic publications) or downstream R&D (patenting) during the period 1926 to 1940.[8] We use information on corporate ownership and control. In particular, we identify firms that are controlled by other (parent) firms, or control their own subsidiaries, and firms which are part of business groups with multiple affiliates, often operating across industries. Such interfirm ties matter because, prior to 1940, U.S. corporate ownership was quite different from what it is today; business groups and conglomerates of various types dominated the U.S. economy.[9] We supplement the information on corporate ownership and organization with data on firm size (assets) and profitability, collected from Moody's Manuals and from the Center for Research in Securities Prices (CRSP).[10] We also include data

[5] Provided through Google Patents and IFI claims services.

[6] Front-page NPL citations were introduced in 1947 at the U.S. Patent Office.

[7] In the second measure, we calculate the share of citations (out of total backward citations) made to European journals by American journals in each scientific field. For the third method, we calculate the ratio of prominent American scientists trained in Europe divided by American scientists without experience in Europe.

[8] Companies were included only if they had at least one publication or one patent over the sample period.

[9] For instance, the DuPont company, a corporate pioneer in chemical research, also had a significant ownership stake in General Motors. Indeed, Pierre DuPont became President of GM in 1920 after having served as DuPont's President. DuPont invented and produced lacquers and enamels for the growing automotive sector. It also manufactured tetraethyl lead, the anti-knock agent discovered by Thomas Midgely, a chemist working at GM.

[10] Moody's Manuals provide balance sheet data, including sales and assets. CRSP contains information on a subset of

on the extent of competition in each firm's 3-digit industry using a cross-sectional classification of sectors into monopolistic, oligopolistic and competitive ones (Wilcox, 1940).

Our sample period, ranging from 1926 to 1940, is well suited to the issue at hand.[11] First and foremost, this is the period in which the phenomenon of corporate science emerged in the United States. Second, our sample period ends before the onset of World War II, when the U.S. government became deeply involved in scientific development both directly and through military procurement contracts (Gross & Sampat, 2020a, 2020b). During our sample period (much of which coincides with the Great Depression), the U.S. government had little influence on corporate research and development activities.

We find that firms in the technological frontier were more likely to invest in science. For instance, firms with forward patent citations above the sample average published around nine times more in scientific journals than firms whose patent citations were below average. The difference is similar for industrial lab sizes: firms with above-average forward patent citations employed around ninety lab personnel on average, while firms with below-average citations employed eleven. Our results also indicate that large firms, operating in concentrated markets, and affiliated with business groups published more. Moreover, these relationships are stronger for firms that relied on scientific areas in which American public science was relatively less developed. For instance, firms whose patents cited science themselves published more, especially when the firms relied upon scientific fields, such as chemistry and physics, that were relatively underdeveloped in the United States.

We make three contributions to the literature. First, we provide empirical evidence on, and explanations for, the rise of corporate research in America and the emergence of an innovation ecosystem, which played an important role in promoting American science and the U.S. economy to world prominence. Lamoreaux and Sokoloff (1997) study the internalization of R&D by U.S. firms around the turn of the twentieth century from the perspective of the independent inventors and their gradual conversion into salaried R&D employees in larger firms. Nicholas (2010, 2014) attributes one of the causes of this transition to the increasing complexity of the technology underlying the chemical and electric industries. Company histories (Hounshell & Smith, 1988; Jenkins & Chandler, 1975; Maclaurin & Harman, 1949; Reich, 1985) document the various motives behind the establishment of large industrial R&D laboratories. Mowery (2009) notes that companies were withdrawing from research by the 1980s, and Arora, Belenzon, and Patacconi (2019) show that this trend reflected changes in firm behavior and was not

listed firms, for which market capitalization is calculated.

[11] Our sample begins in 1926, as this is the first year in the KKMY sample and the year in which the CRSP data on stock market data becomes available.

simply due to changes in size or composition of industrial activity. We show that the rise of corporate research was spearheaded by firms operating on the technological frontier, typically large, publicly traded firms, some of them affiliated with business groups, especially in scientific fields that were underdeveloped in the United States. This is consistent with the interpretation that the weakness of university research provided an opportunity for firms to gain competitive advantage by internalizing scientific research. Our findings suggest that the dramatic growth of university research after WWII affected the private returns to firms from undertaking such research themselves.

Second, we provide evidence in support of the missing institutions (or institutional voids) framework in a context hitherto not documented in the literature. Khanna (2000, 2018), and Khanna and Palepu (2000) apply this idea to emerging markets; the present paper finds that this concept may be useful in understanding phenomena related to the U.S. economy in the first half of the twentieth century. Our results are consistent with the view that corporate research in America was driven by the attempts of large corporate entities to make up for the weakness of American academia at the time. This finding, as well as the finding that business group affiliates were prone to invest in basic scientific knowledge, are consistent with Nelson (1959b). In addition, these results provide a historical perspective on contemporary debates regarding the role of large U.S. corporations in advancing science and promoting innovation. Whereas Autor, Dorn, Katz, Patterson, and Van Reenen (2020) describe large U.S. corporations as highly innovative "superstar firms," others, such as Gutiérrez and Philippon (2020), view them as primarily inefficient entities shielded from competitive pressures.

Third, the dataset we construct, combining information on U.S. corporate ownership and U.S. corporate science in the interwar period, is the most comprehensive of its kind. This dataset should open the way to future research on the possible links between corporate characteristics, research and development, government policy and institutional context.

The paper is organized as follows. In the next section, we summarize the rise of American corporate science in the early twentieth century. Section 2 and 3, respectively, survey the historical context and set up the theoretical framework to for our analyses. Section 4 describes our data. Section 5 presents our econometric specifications and estimation results. Section 6 concludes.

## 2 The Rise of Corporate Science in America

Prior work has established that American firms in the early twentieth century steadily increased the scope of their activities, and some firms invested in internal R&D as well (Chandler Jr, 1993). Lamoreaux and Sokoloff (1997), in their study of American inventors between 1870 and 1911, show that independent inventors who had previously supplied inventions and contract research to firms began to be directly employed by firms. Between 1921 and 1940, the number of firms with labs increased more than sevenfold, from 297 to 2264 (Mowery & Rosenberg, 1998). The available evidence from company histories suggests that in the early years, corporate labs focused on quality control and solving operational problems rather than fundamental science.[12] For instance, Charles Dudley's tenure at the Pennsylvania Railroad company that began in 1875 was focused on examining the metallurgical properties of the rail tracks that were supplied to the firm by steel companies. Thomas Edison's Menlo Park facility was known for shunning fundamental investigations that Edison considered "purely aesthetic" (Wise, 1985).[13] However, by WWI, inventions were increasingly science-based. Initially, companies looked, as they had in the past, to external suppliers to fulfill this need.

This motivated the establishment of specialized contract research organizations, such as the Mellon Institute in 1913. The institute grew steadily in contract revenues ($300,000 to $800,000) between 1919 and 1929. Over the same period, the number of industrial fellows sponsored by firms grew from 83 to 145. Industrial fellows, such as George Curme, made crucial contributions to replacing coal tar with petroleum for certain fine chemicals, while Union Carbide's contract with the institute yielded ethylene gylcol, the antifreeze, which became a key product for the firm (Servos, 1994, p. 223). But contract research worked best for generic, well-specified problems. Outsourcing research entailed contracting problems and required the costly transfer of firm-specific information. Indeed, outsourcing research required that the firm itself had significant research capabilities (Mowery, 1983).

Research managers, such as Willis Whitney at GE Research Labs, Frank Jewett at AT&T Bell Labs, CEK Meese at Kodak and Charles Stine at DuPont, therefore chose to invest internally. Other firms, such as Westinghouse, Standard Oil, Western Union, RCA and Alcoa, soon followed by also instituting

[12] Robert Duncan, the founder of the Mellon Institute for Industrial Research (established in 1913, the premier contract R&D organization of its time) and an advocate of corporate research, lamented that the factories of American firms were dominated by foremen that stuck to traditional practices and that managers were too myopic to wait the "two, three, or even five years" for scientific projects to reach their potential (Servos, 1994, pp. 223–225).

[13] According to a foreign-trained lab employee, Edison argued that, "We can't be like those old German professors who, as long as they can get their black bread and beer, are content to spend their whole lives studying the fuzz on a bee." (Wise, 1985)

in-house research programs. As expected, the quantity of corporate research increased. As we show in Section 4, American firms published three times as many scientific papers in 1940 as they did in 1926. The quality of research also improved dramatically: Irving Langmuir (GE) was awarded his Nobel Prize in Chemistry (1932) and Clinton Davisson (AT&T) in Physics (1937).

## 2.1 Causes Behind Corporate Science

Why did firms invest in science? There are four explanations, which are not mutually exclusive and perhaps even complementary.[14] First, German chemical firms such as Bayer, Hoechst, and Agfa fared well in the international organic synthetic dyes market by building on their corporate research (Reich, 1985, p.41). This set a precedent for American firms to emulate.

Second, American inventions were being challenged by European competition, which was leveraging scientific advances to invent new products and processes. GE's electric lighting business that was started by Edison in 1879, for instance, was based on carbon-filament high-vacuum incandescents. However, Walther Nernst (the 1920 Nobel Laureate in Chemistry) at Göttingen invented a glower that required no vacuum to operate and was more efficient. Westinghouse eventually acquired the patent rights to the Nernst glower in a bid to compete with GE (Wise, 1985). This was one of the motivations behind the establishment of GE Research Laboratory (GERL) in 1900. Langmuir, one of Nernst's American doctoral students, was also recruited by Whitney to further his work on surface chemistry.

Third, American firms had often reached the technological "frontier," and improving their existing products and processes required a deeper scientific understanding of how they worked. New products in a range of industries, such as chemicals, electric lighting and communications, were even more deeply rooted in scientific advances, but few independent inventors had the required scientific capability and equipment. When firms attempted to acquire technology from abroad, a lack of scientific sophistication prevented them from fully exploiting such imports. For instance, after the U.S. entry into World War I, the federal government mandated compulsory licensing of German patents under the Trading with the Enemies Act (TWEA) of 1917. However, American firms found it difficult to replicate the products described in the patents. DuPont and other U.S. firms had to expend substantial R&D dollars to reproduce German dyestuffs, underscoring the need for internal research (Hounshell & Smith, 1988).[15]

[Figure 1 Here]

[14] This section draws on Arora, Belenzon, Patacconi, and Suh (2020a).

[15] Inadequate disclosure in patents did not help. A classic example was the Haber-Bosch process, which was critical for synthesizing nitrogen, where BASF had withheld information about the catalysts required (Haynes, 1945).

Finally, in contrast to the European system, American universities were less focused on basic research and lagged their European counterparts, particularly in quantum physics and organic chemistry in the interwar period.[16] Five out of the 42 Nobel Prizes in Physiology (12%), three out of the 39 in Chemistry (8%) and six out of the 46 in Physics (13%) awarded between 1901 (the first awards) and 1939 went to American scientists. This is in stark contrast to the post-1940 period (1940-2020), where the American share jumps to 55% for Physiology, 68% for Chemistry, and 51% for Physics (see Figure 1 for trends over time). The gap was wider in select fields of chemistry and physics: only two Americans (Irving Langmuir and Karl Compton) were invited to the famous 1927 (5th) Solvay Conference on Electrons and Photons. In mathematics, Germany produced well over half of all PhDs in the subject until 1920, while America was responsible for around a quarter (Castelvecchi, 2016).[17] This is indicative of the types of transatlantic intellectual interactions that occurred before the 1930s: American physicists would receive postdoctoral training in Germany, while German physicists would come on lecture tours in the United States (Fleming & Bailyn, 1968; Holton, 1981).[18]

University research budgets also reflect this gap. U.S. universities before WWII received very little federal funding for research. Geiger (1986, pp. 273-4) notes that the number of publications by the nation's top 16 universities at the end of World War I was less than a quarter of its level at the eve of World War II. In 1919, the budgets of these universities were less than half of their level in 1937. Firms may have therefore found research from American universities inadequate for their purpose.

A comparison of the research expenditures between the best corporate laboratories and universities of the era underlines this point. In its "Research: A National Resource" report published in 1938, the National Resources Planning Board under the NRC surveyed 1,450 American colleges and universities and found that the top 150 spent an average of $ 333,333 per university on research (Council, 1938). The University of Chicago ($2,557,803 in 1929-30), and the University of California ($2,350,000 in 1928-29) were the top research spenders. By comparison, Hounshell and Smith (1988, p. 612) note that DuPont's

[16]Irving Langmuir, for instance, was disenchanted with the lack of research support at Stevens Institute of Technology, where he joined as a faculty member in 1906: "To his chagrin, he found few students with an interest in science. His attempts to upgrade laboratory facilities and the quality of student work met with hostility from students and indifference from his colleagues. To make matters worse, he had little time for research. When he left in 1909 to join the GE Research Laboratory, Langmuir found a position that met his needs far better" (Reich, 1985, p.111).

[17]The U.S. would overtake Germany as the largest producer of mathematics PhDs by 1940.

[18]To be clear, there were areas of American scientific and technological excellence. Agricultural sciences and mechanical and civil engineering, for instance, were nurtured after the Morrill Act of 1862. There was a long tradition of applied research for specialized products, such as boilers and rubber in Purdue University and the University of Akron, respectively (Geiger, 1986). In machine making, American firms were technologically advanced to the extent that German competitors were playing catch-up by reverse engineering imported American machines (Richter & Streb, 2011). Moreover, American universities were clearly catching up with European universities during the interwar period (MacLeod & Urquiola, 2020; Urquiola, 2020).

1925, 1930 and 1935 budgets were \$1.99 million, \$5.5 million and \$6.6 million, respectively. This implies that DuPont's R&D budget in 1930 was as large as that of the Universities of Chicago and California put together. AT&T's R&D data from Maclaurin and Harman (1949, p. 158) show that the 1925, 1930 and 1935 budgets were \$11.7 million, \$23.2 million and \$15.4 million, respectively. Simply put, the weakness of American university research represented a gap, so companies such as Standard Oil, DuPont, AT&T, GE, Kodak and Alcoa could hope to gain sustained advantage by creating scientific expertise in-house.

## 3 Theoretical Framework

The foregoing account of the rise of corporate research stresses three factors: the imperative to innovate for the leading firms, the role of science in facilitating innovation and the weakness of American university science. To study more formally how these factors interact, we adapt the framework developed in Arora, Belenzon, and Sheer (2021). Whereas they analyze the impact of spillovers, we focus on the the differences across firms in the payoffs from innovation and the effect of public science on research investments.

There are two firms, indexed by 0 and 1. Both compete in the product market, and both invest in innovation, $d_0$ and $d_1$, respectively. There are three stages. In stage 3, the firms compete in the product market. Their product market performance depends on the quality of their products and the cost of producing them. We assume that cost and quality depend upon the innovation output, $d_i$, $i = 0, 1$. Their payoffs from stage 3 are $\Pi(d_0, d_1)$ and $\tilde{\Pi}(d_1, d_0)$, where the tilde indicates firm 1. Firm 0 is closer to the frontier, so that its marginal return from innovation is higher than that of its rival. Firms farther from the frontier can increase profits by imitation, reducing production bottlenecks and increasing scale, possibilities that the leaders have already exhausted. Instead, leaders have to introduce new and improved products and processes—to innovate. Accordingly, the marginal product of innovation for firm 0 is greater than that of firm 1. To represent this, we assume that the payoff from innovation for firm 0 has an additional term, $kd_0$, $k > 1$.

In stage 2, firms choose their innovation output. Firm 0 chooses $d_0$ and firm 1 chooses $d_1$. The cost of innovation for firm 0 is $\phi(r_0; u)d_0$, where $r_0$ represents investments in internal scientific research by the firm and $u$ indexes the stock of (relevant) public science. Innovation typically requires the invention of new products and processes. Internal research reduces the cost of invention by guiding the search for inventions in more promising directions. Innovations may also be based on inventions acquired from independent inventors, other firms or university researchers. Thus, the cost of innovation also depends on the state of public science. It is natural to assume that both internal research and public science reduce the unit cost of innovation, $\phi(r_0; u)$, i.e., $\frac{\partial \phi}{\partial r_0} < 0$, $\frac{\partial \phi}{\partial u} < 0$, and diminishing returns so that $\frac{\partial^2 \phi}{\partial r_0^2} > 0$.

In stage 1, firms may invest in research. Firm 0's research investment is denoted by $r_0$, and the cost of research is modelled simply as $\frac{\gamma}{2} r_0^2$, so that the value of the firm, $v = d_0 - \frac{c_{00}}{2} d_0^2 - bd_1 - \frac{c_{11}}{2} d_1^2 + c_{01} d_1 d_0 - \phi(r_0, \lambda) d_0 - \frac{\gamma}{2} r_0^2$.

### 3.1 Model Predictions and Empirical Implications

We provide details and proofs in Appendix A. Here we provide the main results and the intuition. Although the profit and cost functions are otherwise symmetric, this additional term for the leader can result in markedly different outcomes. The returns to investing in research depend on the scale of innovation because research reduces the unit cost of innovation. Firm 0 has a higher marginal return to innovation, in equilibrium, it will innovate more, and thus will have a higher marginal return from research. Furthermore, as $k$, representing the higher returns from innovation to the leader, increases, the marginal return to investing in research increases for the leader. The supply of public science will enhance the effect of $k$ on research if it complements research and will diminish it if the two are substitutes in reducing the cost of innovation.

As $k$ increases, the returns to research fall for the follower if innovations are strategic substitutes. Intuitively, an increase in innovation by the leader reduces innovation by the follower if innovations are strategic substitutes. The decline in its innovation reduces the follower's incentives to invest in research. In other words, the gap between the leader and follower in the value of innovation leads to a corresponding divergence in the marginal returns to investing in research if innovations are strategic substitutes.

The first empirical implication of the model is that there is likely to be an important extensive margin for research. Indeed, as we show in Section 4, the distribution of innovations across firms is highly skewed, and, further, only a minority of firms that innovate invest in research. We also find that firms with higher returns from innovation–e.g., more patent intensive firms and firms with higher quality patents–are more likely to invest in research.

Related to this, as the gap between the leader and follower increases, the leader's returns to investing in research will depend on the supply of public science, increasing with public science if public science and research are complements and decreasing with public science otherwise.

When we restrict attention to the intensive margin, research investment depends on two other factors: public science and strategic interactions in the product market. An increased supply of public science would tend to increase investment in research, unless public science is a substitute for internal research. Strategic interactions are important as well. If innovations are strategic substitutes, so that an innovation by the rival reduces the marginal returns to innovation by the focal firm, then public science would increase internal research even if public science and internal research were independent in their direct effect on innovation. Put differently, if we observe a negative relationship between public science and internal research, it implies that public science and research are substitutes, or that innovations are strategic complements, or both. Finally, public science can reduce the value of the leader, especially if public science and internal research are substitutes in reducing the cost of innovation.

# 4 Data

Our unbalanced panel of firms is constructed by matching several datasets: the corporate ownership and financial statements dataset assembled by KKMY,[19] augmented by market value data on other listed companies from CRSP; USPTO data from Google Patents and publication data from Microsoft Academic Graph (MAG). The combined dataset covers the period 1926-1940.

We begin with 234 firms from KKMY that patent at least once within our sample period in an IPC that cites at least five scientific articles between 1947 and 1957. This restricts our sample to firms that are "at risk" of beginning scientific research.[20] We augment KKMY, which consists of large industrial firms active during the 1920s, by including 235 listed firms from CRSP that patent (Kogan, Papanikolaou, Seru, & Stoffman, 2017). Therefore, our basic sample consists of 469 private and public American firms (7,035 firm-years) that patent at least once in our sample period in an IPC that cites at least five scientific articles between 1947 and 1957. Of these, there are 469 firms (and 4,305 firm-years) for which we have financial statement data.[21]

## 4.1 Scientific Publications

We use Microsoft Academic Graph (MAG) to source 283,992 peer-reviewed scientific publications between 1926 and 1940. We exclude papers in the humanities and the social sciences based on their OECD

[19] As KKMY collect data for 5 years (1926,29,32,37,40), we collect data for the intervening years through the Moody's Manuals.

[20] Examples of excluded patent classes include B27M (woodworking), B60P (loading transportation vehicles) and E03D (Water Closets or Flushing Valves thereof). Around 26% of patenting firms in our sample are lost to this restriction.

[21] The difference in firm-years is because patent data are available for all years in our sample period, but financial data have years without coverage.

subfields.[22] We calculate publication stock for a firm-year using a perpetual inventory method with a 15% depreciation rate. Similar to patents, we calculate normalized forward publication citations by dividing raw forward citations received by publications up until 2019 by the average number of forward citations received by the focal cohort (papers published in the same year). Using Marx and Fuegi (2020), we also identify which publications are cited by a U.S. patent in the future.[23]

We distinguish between university authors and corporate authors of scientific publications. We match 140,766 author affiliations from 283,992 papers to each sector. For universities, we filter publications that contain affiliations that indicate academic authorship, such as "University," "College," "Institute of Technology," "School" etc. This process yields 60,305 affiliations (123,657 papers) that are related to universities for 1926-1940. We limit our sample to affiliations in the United States only. For corporations, we use a fuzzy string matching algorithm that takes into account abbreviations frequent in the era (e.g., firms in the railroad sector may be abbreviated as RR (railroad), RW (railway), RC (rail company)), and name variants for certain companies (e.g., AT&T's Bell Labs, SOCONY for the Standard Oil Company of New York). We ensure that eponymous charitable foundations and hospitals (e.g., by DuPont, Carnegie and Rockefeller) are not erroneously classified as corporate publications. We match 3,194 corporate publications to 201 sample firms. Of these, 110 are found in KKMY sample, 162 are found in CRSP and 71 are found in both samples.[24]

In Table 1, we find that electrical engineering was the scientific field in which the most papers were published, with a total of 1,642 papers published by 156 firms during our sample period. The top publishers in Figure 2, GE, AT&T and Westinghouse, all publish heavily in this field. Electrical engineering is followed by physics (461 articles) and chemistry (344 articles). The field with the higher number of forward citations received from other papers is physics (1.64), followed by biology (1.37).

[Insert Table 1, Figure 2 Here]

### 4.2 Corporate Labs

We also obtain data on the size of R&D labs operated by firms from a national survey by the National Research Council (NRC) conducted since 1920 (Service, 1931). Data from these surveys have been used in Mowery and Rosenberg (1999), Nicholas (2011), Field (2003) and Furman and MacGarvie (2007). We

[22]These fields of science (FOS) have been defined in the 2002 revision (6th edition) of the Frascati Manual. See https://www.oecd.org/science/inno/38235147.pdf for full list of classifications.

[23]We use Version 24 of this dataset, available from https://zenodo.org/record/3976926#.YSFi2S2cZTY

[24]See Appendix C.5 for details on matching.

manually match our firm to firms in the 1927 (999 firms), 1931 (1,620 firms), 1933 (1,562 firms) and 1938 (1,769 firms) surveys. We collect the number of total personnel that were employed in labs. Figure 2, right panel, shows the distribution of lab personnel by firm. Some of the top publishers (shown in the left panel), such as AT&T and GE, also operate the largest labs, but the rank correspondence is not one-to-one. For instance, DuPont operates a large lab with over 3,096 personnel, but publishes a total of 21 papers in our sample period. This may reflect a heterogeneity in firm publishing policies, as well as field-specific publishing behavior.[25] This underlines the importance of normalizing the number of American scientific papers (from WoS) by European counterparts to account for field-specific differences in publication behaviors.

### 4.3 Gap in University Science

We measure the "void" or "gap" in university science in America compared to Europe by the citation-weighted scientific publications authored by scientists in each region. We also find broadly similar results using two alternative measures: the number of scientists trained at or affiliated with a European university, and the citations to European journals made by American journals.

**Scientific Publications: U.S. and Europe** — We use the country of correspondence for the authors of scientific publications. We first collect address information of authors for 44,355 publications published between 1900 and 1920 from WoS and classify addresses into US, Europe and "Rest of World" regions based on their country names.[26] For publications missing addresses, we match the authors' last and first names to the *American Men of Science* directory to identify 27,924 publications by prominent American scientists. The rest of the publications during this period are classified as European. We exclude papers in the social sciences and humanities and are left with 15 OECD subfields for which at least one "European" or "American" published between 1900 and 1920.[27] The above process yields 155,571 publications by Europeans and 60,605 publications by Americans in the sciences between 1900 and 1920. To adjust for quality differences, we weigh the publication counts by the number of forward paper citations received until 2019. These numbers are broken down by field in Appendix Table D9.

**American Men of Science** — Scientific areas where the United States is ahead will exhibit more homegrown scientific talent, while areas where the U.S. lags behind Europe will feature more scientists

[25] For instance, Western Union's 1931 publication in the Transactions of the American Institute of Electrical Engineers reports the construction of a new transatlantic cable the firm laid in 1928 (See https://ieeexplore.ieee.org/document/5055804).

[26] Microsoft Academic Graph does not contain an address field, whereas Web of Science contains a separate field dedicated for addresses and country classifications based on these. See Appendix Tables D1 and D2 for details on classifications.

[27] We use the correspondence in Marx and Fuegi (2019) to map Web of Science subject fields to 39 OECD subfields.

trained in Europe. For instance, the founder of the *American Journal of Chemistry* (Ira Remsen) studied at Göttingen, while the alma mater of the founder of the *American Journal of Mathematics* (James Sylvester) was the University of Cambridge (Kevles, 1979). On the other hand, areas such as agriculture and civil engineering, where the United States did not lag as far behind, did not require a similar import of overseas talent. We collect information on European education/affiliation by American scientists from the 1921 (3rd) edition of the Cattel Directory of American Men of Science.

Published by James McKeen Cattel since 1906 and running its 38th edition in 2020, the American Men of Science directory (hereafter AMS) is one of the oldest and most comprehensive listings of scientists active in the United States (Moser & San, 2020). To measure relative scientific strength before our sample period (which starts from 1926), we focus on the 1921 edition because the Optical Character Recognition (OCR) quality is highest, and it provides the most comprehensive listing of scientists.[28] The number of listed scientists increases from around 4,000 in 1906 to 5,500 in 1910 and 9,500 in 1921, likely reflecting both the growth of American science as well as better coverage by Cattel. We extract 8,232 author entries. Of the 7,245 scientists on whom we have the the required information, 1,649 are trained in European institutions.[29] It is possible that scientists who were exclusively trained in the United States up to the doctoral level are recruited by European institutions and show up as "European" due to their affiliations, though random checks suggest that this is quite rare. We manually classify each scientist into a scientific field (OECD subfield) based on the subject listed for each entry (Appendix Figure D1 shows an entry for a scientist listed under chemistry). This allows us to count the number of European-affiliated (and non-affiliated) scientists in America.

**Web of Science: Transatlantic Journal Citations** — Another way to measure scientific gaps is by the number of citations made to publications in European journals by American journals. We classify 244 journals indexed in the WoS Science Citation Index - Expanded (SCI-EXPANDED) as "American" or "European" based on name and web searches. Journal names in non-English languages, such as "Comptes Rendus" and "Zeithschrift für Physik," are first identified as non-American (with the exception of those in Latin such as "Acta Mathematica"). All other journals are searched online and classified based on the home country of the academic society. Where this information is not available, we use the home country of the journal's founders. 230 journals out of the 244 are classified, 111 (45%) of which are American.

For articles published between 1900 and 1920, we count the number of citations made by "American" journals to "European" journals in the same period. This constitutes a measure of European scientific

[28] https://catalog.hathitrust.org/Record/003255132 for details.
[29] See Appendix D.1 for details on cleaning the AMS data.

strength — if a field relies more on European science, citations to Euroepan journals would be higher. We also count the number of citations made to American journals, which constitutes a measure of American scientific strength.

**Calculating Scientific Gaps for Firms** — Our regional scientific activity data (from AMS and WoS) are encoded at the scientific field level. Therefore, we link them to firms based on how much each firm patents in a patent class, and on how much a patent class relies on a scientific field. For instance, for the first measure using publication addresses, we calculate the number of papers (European and American) published between 1900 and 1920 relevant for each 4-digit IPC based on the share of patents in that IPC that cite the OECD subfield in their front page Non-Patent Literature section:[30]

$$European\ Papers_{IPC,field} := \frac{NPL\ Citations_{IPC,field}}{NPL\ Citations_{IPC}} \times European\ Papers_{field} \tag{1}$$

We sum $European\ Papers_{IPC,field}$ over all OECD subfields to obtain the number of (European) papers "relevant" to a given IPC: $European\ Papers_{IPC}$. We then map this IPC level value to a firm using the share of patents the firm has in each IPC over the sample period, 1926-40.

$$European\ Papers_{firm,IPC} := \frac{Patents_{firm,IPC}}{Patents_{firm}} \times European\ Papers_{IPC} \tag{2}$$

Summing $European\ Papers_{firm,IPC}$ over the 4-digit IPCs, we obtain the number of European publications relevant to each firm. We repeat the same procedure for American publications published between 1900 and 1920 in WoS. We then divide the number of European publications by the sum of the American and European publications at the firm level to get our primary measure of scientific gap the firm faces.[31]

For the gap measures using scientist affiliations in the 1921 edition of the American Men of Science, we replace $European\ Papers_{field}$ in Equation 1 with the number of scientists in that OECD subfield that have been trained, at least in part, in Europe (while $American\ Papers_{field}$ are replaced with the

[30] We use data for patents granted for the first 10-year period from the time NPL citations were formalized in U.S. patent documents (i.e., between 1947 and 1957).

[31] For example, AT&T, which is in the 90th percentile in this score, 15% of its patents in IPC H01J (Electric discharge tubes or discharge lamps) between 1926 and 1940. Patents in this IPC, in turn, cite the Chemical Sciences most often (26%), followed by Electrical Engineering (23%) and Physical Sciences (21%) between 1947 and 1957. As we see in Table D9, Chemical Sciences and Physical Sciences have European-to-American ratios that are higher than the average, which contributes to the higher firm-level gap score for AT&T. In contrast, General Ice Cream Corp, which is below the 10th percentile in this score, patents most often in A23G (Cococa; Cocoa Products), where the highest number of NPL citations are made to Biological Sciences. Biological sciences, in turn, has a European-to-American ratio below the average, which contributes to the firm receiving a low gap score.

scientists without European experience). Scientists are assigned a firm-specific weight, as before, to reflect the importance of their scientific field to the firm. We divide Euroepan-affiliated scientists by the sum of European-affiliated and non-affiliated scientists to generate an additional AMS-based historical (1921) scientific gap measure.

To calculate the gap measure based on citations to European journals by American journals, we replace $European\ Papers_{field}$ in Equation 1 with the number of citations to European journals made by American journals in that OECD subfield ($American\ Papers_{field}$ is replaced with American citations to American journals). As before, each scientific field is weighted by its relevance to a firm, based on the patent classes the firm patents in and the rate at which patents in the class cite the scientific field. Dividing the European citations by the sum of American and European citations yields the journal citations-based scientific gap measure.

### 4.4 Patents

Our patent data are derived from Google's public patent dataset. There are 637,190 patents granted between 1926 and 1940 by the USPTO. We collect information on the grant date, assignee and inventor names, Cooperative Patent Classification (CPC) codes, as well as prior-art citations made to patents. Based on this, we calculate normalized forward patent citations by dividing the total number of prior-art citations received by a focal patent by the per-patent citations received by all patents granted in the focal patents' issue year.[32] To measure the extent to which a patent "relies" on science, we count citations to scientific publications in Microsoft Academic Graph in the text of the patent, from Marx and Fuegi (2020).[33] We also measure the "novelty" of a patent by counting the number of times the same CPC combination of a patent has been granted since 1790 (Fleming, 2001).[34] A Combination Familiarity score of zero implies that the technical combination has never appeared before. For instance, Wallace Carother's nylon patent for DuPont (US2130948A) combines eight different CPC subclasses, some for polyamides and others for fibers, which was an unprecedented combination at the time (hence, the resulting familiarity score is zero).[35]

[32] This cohort-based normalization is important because a procedural change at the USPTO starting from 1947 substantially increased citations afterwards.

[33] We use in-text citations because NPL citations are available only after 1947. We use references with a confidence score above 8. We find 237 patent citations to science by our sample firms between 1925 and 1940.

[34] "Combination Familiarity" of patent $i$, $R_i$ is $k \times exp(\frac{publication\ date\ as\ patent\ i - publication\ date\ as\ patent\ k}{time\ constant\ of\ knowledge\ loss})$ where $k =$ number of patents granted before patent i and *time constant of knowledge loss* is set to 5 years.

[35] These are D01F6/60,D01F6/58, D01F6/605, C08G69/26, C08G69/28, D01D5/06, Y10S8/21, Y10T428/2904, Y10T428/2976, and Y10T428/2978.

Patent assignees are matched to firms using the same fuzzy string matching algorithm used to match publications.[36] We match 89,328 patents to the 469 firms in our panel between 1926 and 1940. Of these, 234 firms are found in the KKMY sample, 350 are found in CRSP and 115 firms are found in both.

### 4.5 Corporate Ownership, Financial Statements and Industry Concentration

We collect corporate ownership data in order to link firms under common ultimate ownership from KKMY, who collect data on the control of U.S. nonfinancial corporations for the years of 1926, 1929, 1932, 1937, 1940 (and for 1950, after the end of our sample period). Using Moody's Manuals on nonfinancial sectors (Railroads, Public Utilities and Industrials Manuals, which are available from Mergent Online (http://webreports.mergent.com/), the authors start with the largest 200 nonfinancial corporations, ranked by total assets as reported in Berle and Means (1932), and construct ownership trees for these firms, their parents and subsidiaries. The chains of control, from the ultimate owner to all the subsidiaries, are based on Moody's definition of control, which uses both equity links and other considerations. Ultimate owners of control chains (individuals, families or, in some cases, a widely held apex company) are identified using a variety of archival data sources.[37]

**Financial Statement Variables** — Balance sheet data on earnings and assets are not available before 1950 from conventional sources such as S&P Compustat. Therefore, we build on KKMY, who collect data on firm assets and earnings for the sample firms for the years 1926, 29, 32, 37, 40 (and 50), using Moody's Manuals.[38] We expand this dataset for the intervening years from the same source. To classify the industries in which these firms operate, we use descriptions of firm "occupations" in Moody's Manuals. We then manually connect each industry name to one of the 85 3-digit industry codes in the revised 1947 SIC tables (reported by the BEA in 1958).[39] We augment the dataset by collecting all available end-of-the-year stock market value data for all listed firms using the CRSP Monthly Stock File for North American firms. For listed firms that appear on CRSP but not in the KKMY sample, we obtain data on their financials (but not ownership data) from Graham, Leary, and Roberts (2015).[40]

**Measures of Industry Concentration** — We use Wilcox (1940) to classify 3-digit industries into monopolistic, oligopolistic and competitive ones. Wilcox does not rely on any single test of monopoly or competition to measure the extent of competition in different industries (because of difficulties described

[36] See appendix C.5 for details.
[37] See Appendix C for details on the construction of ownership chains and the identification of ultimate owners.
[38] Figure C1 in Appendix C.2 reproduces the 1949 entry for the Porto Rico Telephone Company.
[39] Source: http://www.bea.gov/industry/io_histsic.htm.
[40] This dataset is also manually collected from Moody's Industrial Manuals.

in Wilcox, 1940, pp. 1-12; 19-20). Instead, he uses broad criteria, as well as the regional nature of the markets, to separate effectively competitive industries from effectively monopolistic ones as of 1934-1939 (the measures do not refer explicitly to a specific year). Monopolistic industries are further divided into cases where supply was predominately accounted for by one or two firms (monopoly and duopoly); by only a few firms (oligopoly); by one or a few dominant firms and many smaller ones (dominant-firm industries) and by several or many firms acting in collusion (cartels and effective trade associations). Industries may be characterized by more than one type of behavior. Industries in which none of these were present were classified as competitive. In our analysis, we focus on the distinction between competitive and non-competitive industries.[41]

### 4.6 Descriptive Statistics

[Table 2 Here]

Table 2 presents descriptive statistics at the firm-year level. The maximal number of observations is 7,035 (469 firms observed for 15 years). "Lab Size" counts the number of lab personnel reported in the IRL directory. There are only around a third of the total observations here because the IRL was collected for only five years by the NRC (1927, 31, 33, 37, 40). As observed from the difference in median and mean values (see also figure 4), scientific publications and lab personnel are skewed to the right. The average gross income and assets are $879 million and $1.4 billion, respectively. These are slightly larger than the values for the "pre-period" between 1926 and 1930, reflecting the growth of the U.S. economy.

[Figure 3 Here]

Figure 3 presents trends in corporate investment in science during the interwar period. Publications per firm exhibit an upward trend, reflecting an aggregate increase from just under 200 corporate-authored papers in 1926 to around 700 papers in 1940. In addition, about 12.5% of all firms published in 1926, increasing to 17.6% of all firms in 1940. The staff employed in corporate laboratories mirrors the same upward trend. These figures reflect an expansion of corporate science both on the extensive margin (measured by total publications) and intensive margin (measured by share of firms that publish). This rise in corporate science, however, is not monotonic, peaking in the early 1930s and declining for the next

[41] Both Stigler (1949) and Nutter and Einhorn (1969) follow, and validate, the Wilcox (1940) classification. Nutter and Einhorn (1969, pp. 94-97) supplement the Wilcox classification using other sources and derive a "concentration ratio" for 1939, albeit only for a subset of sectors.

four years, reflecting the cutbacks in research during the Great Depression. The increase in patenting is steadier, if less dramatic, with an aggregate rise from around 3,000 to 6,500 corporate patents per year.

[Figure 4 Here]

Figure 4 shows that these trends mask substantial heterogeneity. For instance, 158 firms out of our sample of 469 firms never operate an R&D lab, while more than half (268) of the firms never publish a scientific article. Perazich and Field (1940) estimate that less than 1% of all firms accounted for a third of all industrial research employment in 1921, 1927 and 1938, respectively. A mere 45 firms in 1938 employed half of the total research personnel (Perazich & Field, 1940).

[Figure 5 Here]

Figure 5 presents the correlation between the scientific gap and corporate science across industries.[42] Corporate investments in science seem to occur in industries where the U.S. lags behind European science. For instance, construction, which relies on civil engineering where the gap is small, exhibits less corporate science investment than communications, which relies partly on chemistry, where the gap is large. This pattern is consistent with our conjecture in Section 2; it also calls for the use of industry fixed effects throughout the empirical analysis.

# 5 Empirical Results

## 5.1 Who Invests in Science?

We first explore the existence of an extensive margin between "leader" and "follower" firms. If leading firms have higher returns from innovation and there is a fixed cost to corporate research, leaders will invest in research while followers will not. This accords with the history of corporate science surveyed in Section 2: select firms, such as General Electric and DuPont, had reached the technological frontier where the payoffs from fundamental, science-based innovation were higher. Since establishing a research organization to "routinize" such innovation was a costly endeavor, "follower" firms that were not as technologically advanced likely chose not to invest.

We find in the first three rows of Table 3 that firms closer to the technological frontier were more likely to publish in scientific journals and operate corporate labs. In particular, firms with patenting intensity

[42]We replicate the results with the other two measures in Appendix Figure B1

greater than the mean tend to have around 22 times greater publication stock and employ around 12 times more lab employees than firms with patent intensity below the mean. Similarly, firms with at least one patent citation to scientific articles during our sample period also tend to have publication stocks that are fiftyfold those of firms whose patents do not cite science, and laboratory personnel fifteenfold larger. Interestingly, these "frontier" firms are in the minority, accounting for barely 6% of the observations.[43]

The other variable heavily associated with investment in research is size. Firms with above average assets publish around 22 times more than those under average, while the difference in lab personnel is around 10 times. Similarly, we find that public firms and firms in the KKMY sample (both correlated with size) tend to invest more heavily in science. Business Group-affiliated firms publish around 16% more publications and employ around 73% more lab personnel.[44] Greater scale may allow firms to reap greater marginal returns from innovation (Cohen & Klepper, 1996). However, it is also possible that firm size and business group affiliation proxies for easier access to internal factor markets that enable conducting basic research (Schumpeter, 1939). Finally, we observe that firms in noncompetitive industries tend to engage more in corporate science (more publications and larger corporate laboratories). This is consistent with the conjecture that firms in concentrated or non-competitive industries, may be able to capture the rents from the provision of scientific knowledge, a quasi-public good.

[Table 3 Here]

### 5.2 Interaction with Gaps in University Science (OLS)

We next explore whether gaps in public science accentuate the incentives of those leading firms to invest further in research. A concern is that our findings are driven by unobserved heterogeneity in firm quality. Unobserved firm quality should be positively correlated with technological leadership and internal science, leading to an upward bias in their estimated relationship. To alleviate this concern, we exploit variation across technology fields differing in the availability of public science. Our theory predicts that technology leaders should find it attractive to invest in internal science, especially when relevant public science is relatively scarce. When public science is scarce, leading firms, which benefit more from science, have a strong incentive to invest. Yet, as public science becomes abundant, science is no longer a source of private value, and investment in internal science becomes less attractive. Therefore, using public science

[43] The total observations for the "Patenting Intensity Above Average?" row (4,305) is smaller than the "Patents Cite Science" rows (7,035) due to missing sales data.

[44] However, unlike the previous results, the difference between business group affiliates and non-affiliates is not statistically significant at the 5% level.

as a moderator of the relationship between technology leadership and internal science, and showing that our key results hold when public science is scarce, should alleviate, at least partly, the unobserved firm quality concern.

We estimate the following specification via OLS, where we measure the availability of public science using our scientific gap measure(s):

$$Corpsci_{it} = \beta_0 + \beta_1 MR\_Inno_i + \beta_2 Gap_i \times MR\_Inno_i + \beta_3 Gap_i + \boldsymbol{Z_i'\gamma} + \tau_t + \phi_c + \epsilon_{it} \tag{3}$$

where $Corpsci_{it}$ is defined as the investment in corporate science by our sample firms, measured by the scientific publication stock;[45] $MR\_Inno_i$ refers to the marginal returns to innovation for firms. One proxy for these returns is proximity to the technological frontier (need for science): i) patent intensity (patents granted divided by log of total assets), ii) average forward citations the firm's patents receive and iii) the number of citations to scientific publications the firm's patents make. We also replace $MR\_Inno$ with firm characteristics indirectly related to returns to innovation: i) the natural logarithm of the firm's total assets; ii) a dummy variable denoting whether the firm is part of a multi-firm business group and iii) a dummy variable denoting whether the firm's industry is classified as "competitive" by Wilcox (1940). $Gap_i$ is measured as the ratio of European publications to American publications relevant to each firm[46]; $\boldsymbol{Z_i'}$ is a vector of controls. We include level values for each science measure since our gap measure is a ratio. We also control for the size of the firm (assets) and include year ($\tau_t$) and 2-digit industry ($\phi_c$) fixed effects. We expect $\hat{\beta}_2 > 0$ if the gap in public science accentuates "leader" firm investment in science.

[Table 4 Here]

Table 4, Columns 1-3 show that firms with higher patenting intensity, whose patents receive more forward citations and whose patents cite the scientific literature were more likely to respond to a gap in university science. Specifically, firms with patents in the 75th percentile of forward patent citations respond around 11% more in terms of new scientific publications compared to firms whose patent citations are in the 25th percentile. We also find similar results when normalizing firm scale (e.g., dummy for whether the firm patents cite science, number of patent citations to science normalized by number of firm patents), but we prefer to independently control for size using assets to also understand the baseline

[45] Using number of lab employees yields similar results. However, with fewer than half the observations, the estimates are far less precise.

[46] We present results using the two alternative measures in Appendix B

relationship between investment in science and size. These results are consistent with our conjecture in Section 2 that, on average, more technologically advanced firms experienced a more acute need for internal research when faced with weaker domestic universities. In terms of the theoretical framework, these findings suggest that university science was a substitute for internal research or that there are strong strategic complementarities in innovation, or both.

We also examine interactions between size and scientific gaps. Column 4 shows that the publication sensitivity to the gap in science of firms with a one standard deviation larger-than-the mean assets is around 1.2 times the estimated sensitivity of a firm whose assets equal the sample mean. This suggests that larger firms were more likely to internalize the effects of corporate research.[47] Similarly, firms affiliated with business groups published about 9 times more in response to gaps in university science (Column 5). We test whether firms operating in more concentrated (less competitive) industries are willing to invest more in science. Other things equal, firms in industries with higher concentration may be less concerned about spillovers, since they can more easily appropriate the returns to their research. Consistent with this, Column 6 shows that being in a competitive industry is negatively correlated with engaging in corporate research in response to scientific gaps.

We find that the non-interacted, standalone coefficients measuring technological frontier and firm size are negative (for instance, -64.052 for patenting intensity and -94.513 for assets). However, the inflection points for the marginal "effects" of these variables occur at fairly low values of the scientific gap measures. For instance, the marginal difference in publication stock related to patenting intensity becomes positive starting from a scientific gap of 0.65, which is in the lower fifth percentile of this measure.[48]

We conclude that firms most likely to respond to gaps in science are the technologically advanced firms. Firm size and organization and, in particular, affiliation with business groups and other multi-firm entities, are related to the likelihood of engaging in research, presumably because of the ability of such corporate structures to internalize the benefits generated by the creation of basic knowledge.

### 5.3 Performance Consequences of Corporate Science

Nelson (1959a, p.119) noted that "Research laboratories may be created and maintained by firms for many purposes, including (...) quality control, (...) improvement of manufacturing methods, improvement of

[47] In unreported robustness checks, we replicate this result by replacing the continuous measure of size with a dummy equal to one if a firm has assets above the sample average, finding similar results

[48] In unreported robustness checks, we replicate the results with only level coefficients (excluding interaction terms, but including identical controls and fixed effects) to find that the average marginal effects on publication stock of these variables are positive, consistent with Table 3.

existing products and development of new uses for them, development of new products and processes, and *scientific research to acquire knowledge enabling more effective work to be done to achieve the above purposes*" (emphasis added). If so, firms investing in scientific research ought also to have better and more valuable new inventions. We use two measures: patents deemed valuable by investors (i.e., whose issuance is associated with increases in the firm's stock price), and novel patents, which combine patent classes which have rarely been combined before. As a benchmark, we first estimate the following OLS specification, which includes firm publications as a proxy for corporate science and controls for the current (contemporaneous) WoS gap measure:

$$HomeRun_{it} = \beta_1 Corpsci_{it} + \boldsymbol{Z_i'}\boldsymbol{\gamma} + \tau_t + \phi_c + \epsilon_{it} \tag{4}$$

$HomeRun_{it}$ is measured using the number of patents that are in the top 5% of stock market value (Kogan, Papanikolaou, Seru, & Stoffman, 2017) and novelty scores (Fleming, 2001).[49] One possible concern is that the OLS estimate of $\beta_1$ may be upward biased if unmeasured technical opportunity drive both corporate publication activity and the value (or novelty) of patents. On the other hand, it is possible that other firm characteristics, such as the quality of the firm, may be negatively correlated with the current technical opportunities, leading to a downward bias. That is, it is possible that technical opportunity is reflected in greater public, not private, science. We instrument for corporate publications using historical public science gap, which can also purge the measurement error in using publication stock as a measure of investment in corporate research.[50]

Our instrument is calculated using American and European scientific publications authored before our sample period.[51] Historical gaps that predate the sample period affect investment in science but are unlikely to affect concurrent firm inventions and value. In addition, patent value and inventive activity might be affected by the concurrent availability of public science. Therefore, we include concurrent gap measures as controls. We use two-stage least squares, where in the first stage, we regress $Corpsci_{it}$ on the pre-sample period public science gap measure (from Equation 3) and other controls, and in the second stage we regress $HomeRun_{it}$ on the fitted values of investment in corporate science ($\widehat{Corpsci_{it}}$) obtained

[49]Results are not sensitive to the use of alternative thresholds, such as top 1% or 10%.

[50]Highly novel patents are correlated with patent quality (patent intensity, forward citations, citations to science used in Table 4, Columns 1-3), since both are patent-based measures. Because higher patent quality is also correlated with higher publication stock (per first three rows of Table 3), this may lead to a spurious relationship between publication stock and patent novelty.

[51]Our sample period is between 1926 and 1940, and the publications for the gaps collected from 1900 to 1920. Similarly, the AMS directory data for the gap measure calculation is from the 1921 edition, while the journals used for the citation share gap measure are those published between 1900 and 1920.

in the first stage.

Market value data for patents are only available for public firms; nevertheless, as shown in Table 3, the bulk of corporate research is carried out by publicly traded firms. Indeed, public firms account for the majority of firms in our sample and constitute the vast majority (around 80%) of firms that publish or operate labs.

[Table 5 Here]

OLS estimates from Table 5, Columns 1 and 4, show that firm publication stock is positively correlated with the number of highly valuable (within top 5% of stock market value per Kogan, Papanikolaou, Seru, and Stoffman (2017)) and highly novel (within 5% of novelty scores measured by the count of subclass combinations per Fleming (2001)) patents. In Columns 2 and 5 we instrument for firm publication stock using the historical (pre-sample period) gaps in university science. The first stage regressions of publication stock against gaps in university science are significant, with an F-stat of 37. Because American universities were catching up to European standards during the interwar period, we also control for the "current" gaps calculated for each firm and year and find it to be negatively correlated with the corporate publication stock.[52] In the second stage regressions (Columns 3 and 6), we find that an increase in the (instrumented) publication stock increases the number of valuable patents. A one standard deviation larger publication stock (due to historical gaps in university science) leads to around three more patents in the top 5% of stock market value (or around 3.5 times the sample mean). The estimate in Column 3 is about 3.4 times larger than the OLS estimate in Column 1 (similarly, the estimate in Column 6 is 1.8 times larger than OLS estimate in Column 4). One interpretation is that the corporate investments in research undertaken when public science lags are particularly potent sources of competitive advantage, as reflected in more valuable and distinct inventions. Put differently, publication stocks might have heterogeneous effects on invention outcomes. Publication stocks that reflect investments in response to gaps appear to have a larger impact on invention outcomes. We find similar results in Appendix Tables B3 and B4 using the two alternative measures of public science gaps between America and Europe. In sum, investments in internal research do result in more valuable inventions.

As an alternative to estimating "technological" (patent-based) returns, we also estimate the contri-

[52]The $European\ Papers_{subfield}$ in the context of Section 4.3 are now calculated for each year, instead of summed up for papers published between 1900 and 1920. The mapping from scientific field to firm are identical, and the gap measure is calculated at the firm-year level, not the firm level.

bution of corporate research to firm value of public firms:

$$ln(Q)_{it} = \beta_1 ln(Pubstock_{it-1}) + \boldsymbol{Z_i'\gamma} + \tau_t + \phi_c + \epsilon_{it} \tag{5}$$

where $Q$ is the market-to-book ratio (or Tobin's Q) and $ln(Pubstock_{it-1})$ refers to the natural log of one plus lagged publication stock.

[Table 6 Here]

Column 1 of Table 6 shows that publication stock is positively related to the market-to-book ratio of firms. A one standard deviation larger publication stock is associated with an increase of Tobin's Q by around 0.01 (2% of the sample mean). We also split the sample by gap measures to probe whether "responding firms" benefited more from science. Comparing Columns 2 and 3, we find that, for firms whose gaps in university science (based on author affiliations) are smaller than the sample mean, there is a statistically insignificant correlation between their publications and their market-to-book ratios. In contrast, for firms with gaps above the sample mean there are positive and significant effects of the publication stock on Q. These results suggest that investments in science are positively related to market value, and that this relationship is driven by firms with large scientific gaps.

We also instrument publication stock by the gap in university science. As expected, we find in Column 4 that the first stage regression coefficient of publication stock against gaps in university science is positive and statistically significant. However, we do not find significant results for the second stage. This may be due to the fact that the market-to-book ratio is affected by many firm-level characteristics and is imprecisely measured in the historical data, making it difficult to estimate the effect of corporate research on it. When we look at a more narrow definition of financial returns, we find that the stock market value of patents is positively related in the second stage to publication stock predicted by our scientific gap measures (Column 3 of Table 5).

## 6 Discussion and Conclusion

We argue that the rise of corporate research in America in the interwar period is related to the weakness of American academia in certain scientific fields. For some firms - large, group affiliates and close to the technological frontier - investment in internal research was the way to overcome this institutional weakness, or void, as well as a source of competitive advantage: competitors would not be able to readily acquire the needed scientific knowledge from universities.

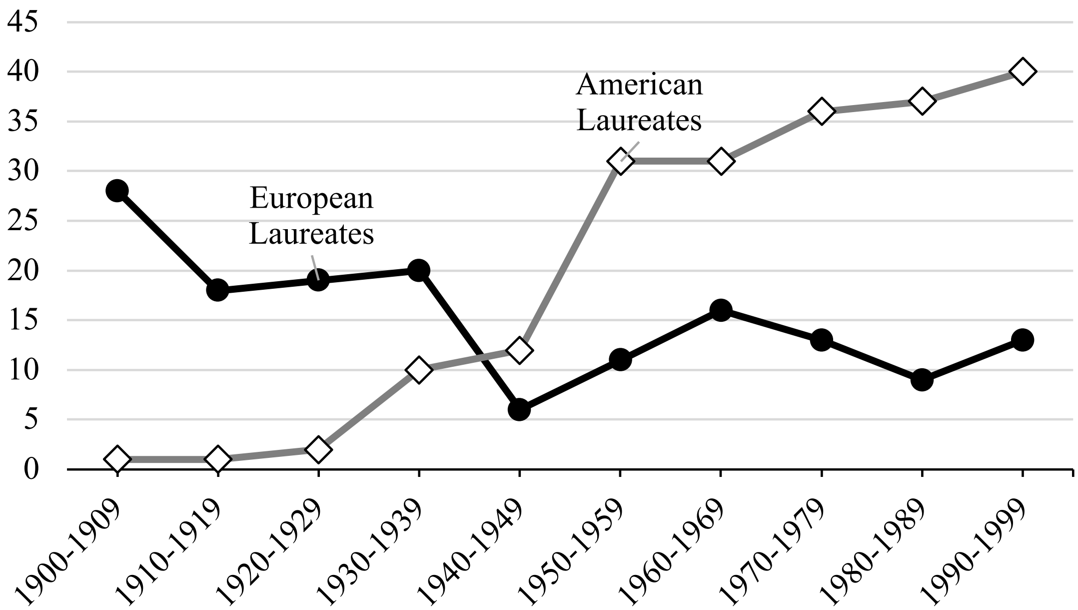
This historical evidence on research carried out by private corporations may be of relevance to the present-day debate in the U.S. and other advanced economies about the costs and benefits of large, technologically advanced firms. Our results suggest that such firms can play an important role in advancing knowledge, but that this knowledge is likely to grant them considerable advantage over their competitors. In emerging economies, our historical analysis sheds light on a relatively little-explored mechanism by which large corporate entities attempt to make up for institutional voids, in this context, voids related to the accessibility of science and the quality of domestic academic research. Our historical evidence suggests that, in some circumstances, private corporations can and do substitute for institutional weaknesses in science, as some corporate giants in contemporary emerging markets (e.g., India, Turkey and Korea in its early stages of development) have done. Naturally, such historical parallels should be used with caution.

In the decades since the end of our sample period, American universities have increased the quality and quantity of their scientific output, yet the implications of this change for corporate research remain poorly understood. As noted at the outset, universities produce both new knowledge as well as human capital, which affect both the costs and benefits of private investment in research and innovation. Following World War-II, the growth of university research was paralleled by growing investments in corporate research. However, by the 1980s, the two trends have diverged, leading to a growing division of labor between academia and universities (Arora, Belenzon, & Patacconi, 2018).[53] How the private value of an input changes as the supply of the input expands is an important but understudied topic. When the input in question is knowledge, whose use by one firm does not preclude its use by another but may affect the private value that accrues to the firms, the issue becomes even more complex.

In addition to providing new evidence on corporate research in America in the interwar period, we assemble the most extensive historical sample of American firms that were involved in innovation during that period, including information on the scientific output of these firms and on the relative gap between American and European universities. We hope that these newly developed data will contribute to future research on the open questions we raised.

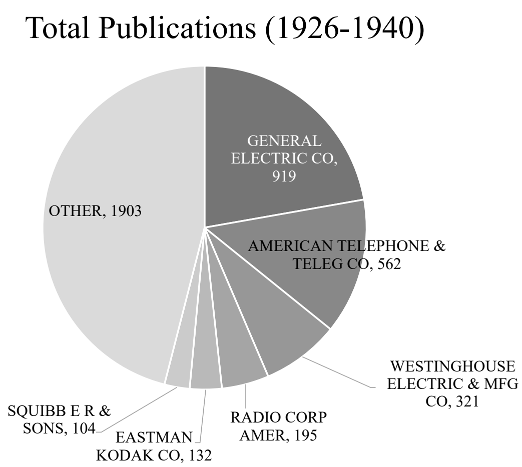
[53] Besides the growing scientific might of American universities, it is possible that corporate research in the 1930s was easier protect from rivals (few could effectively use it), whereas by the 1980s knowledge spillovers may have become increasingly costly. This conjecture is consistent with Arora, Belenzon, and Sheer (2020b), who show that companies cut back on research when spillovers to rivals increased relative to the value from internal use.

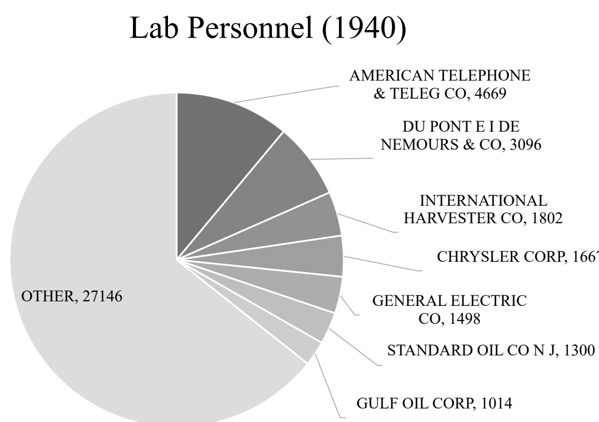
Figure 1: Number of Natural Science Nobel Prize Laureates, by Citizenship at Award




*Notes:* The line graph plots the number of total laureates in the Nobel Prize for Physics, Chemistry, and Physiology/Medicine. The home countries of the winners are coded based on the classification by the Encyclopedia Britannica (please see https://www.britannica.com/topic/Winners-of-the-Nobel-Prize-for-Physics-1856942 for page for Physics). According to the source, "Nationality given is the citizenship of recipient at the time award was made."

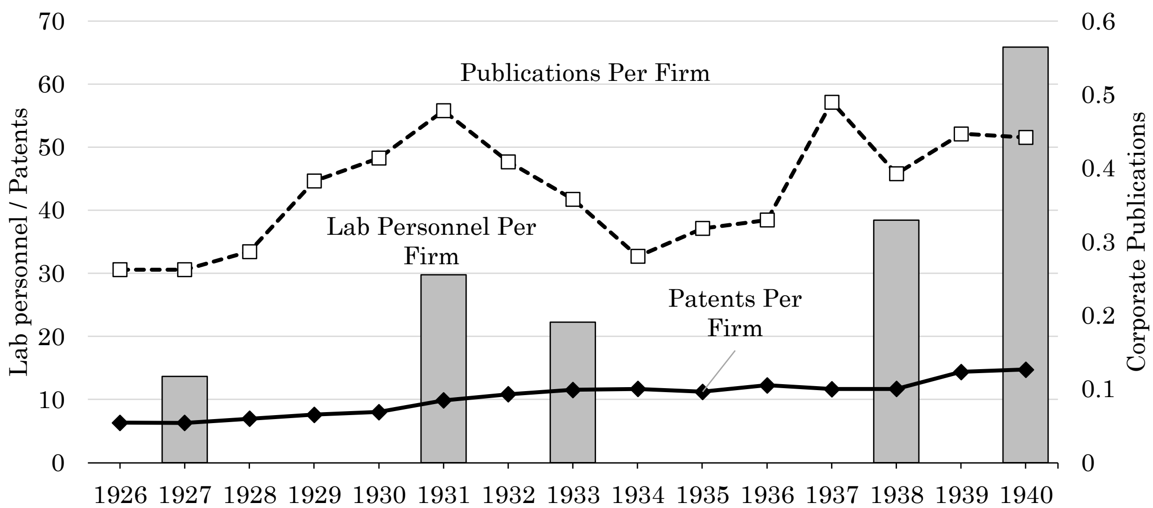
Figure 2: Firm Share of Publications and Personnel



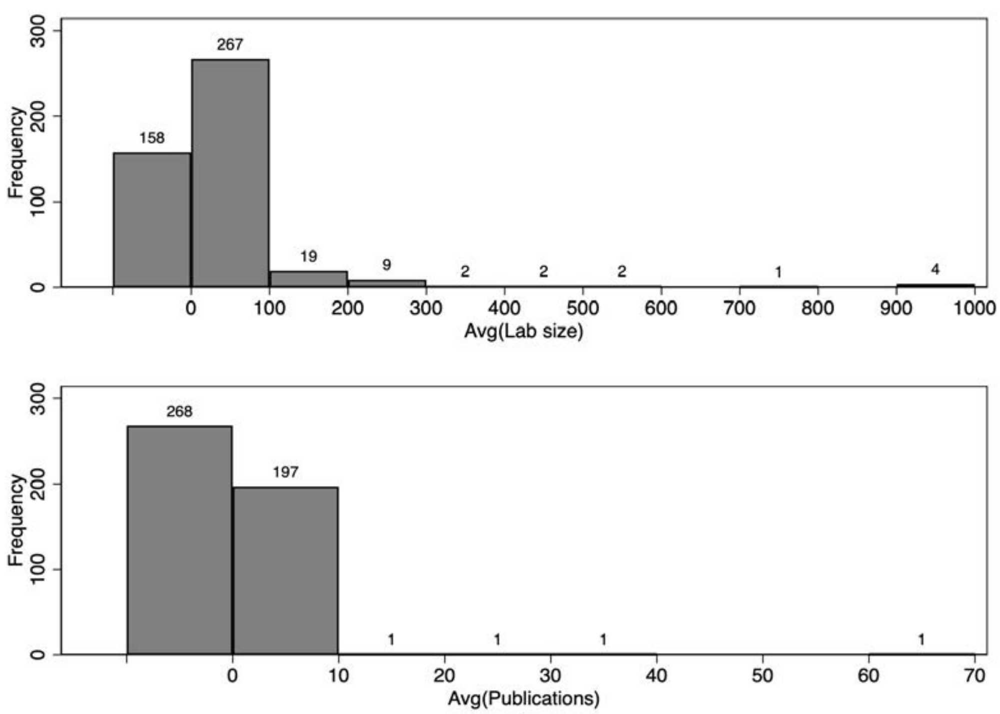

*Notes:* The left pie chart downward sorts the number of total publications by firms in our sample clockwise. The right pie chart sorts the number of lab personnel reported by firms in the 1940 edition of the Industrial Research Laboratory directory clockwise.

Figure 3: Emergence of Corporate Science, 1926-1940



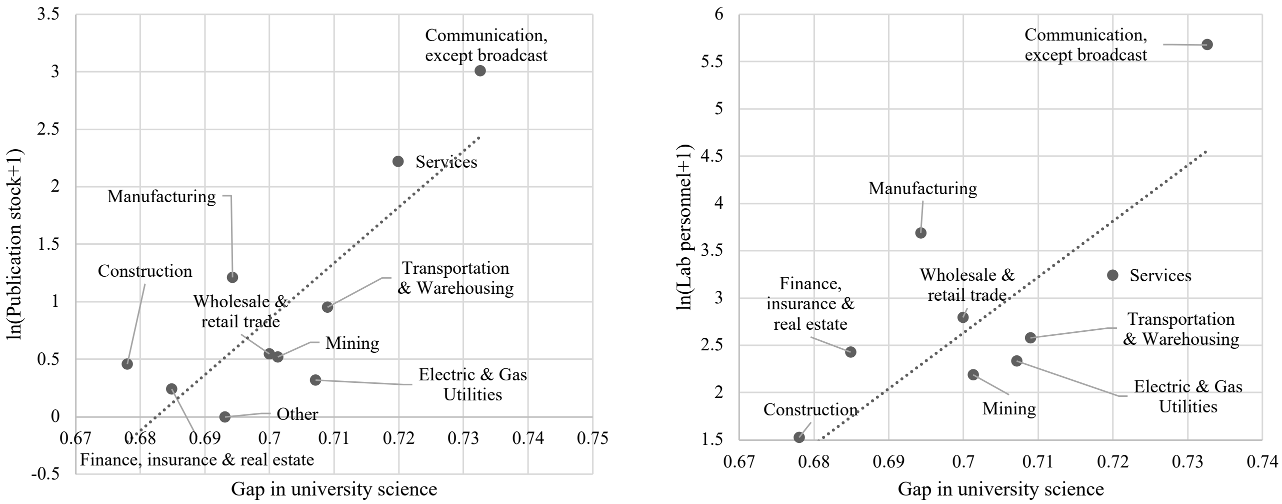

*Notes:* The bar graph indicates the number of personnel employed at corporate laboratories per firm from the Industrial Research Laboratories Directory. The broken line indicates the publications per firm in our sample matched to Microsoft Academic Graph. The solid line indicates the number of patents by firms in our sample matched to USPTO utility patents.

Figure 4: Heterogeneity of Corporate Science




*Notes:* The upper histogram bins the number of personnel employed at corporate laboratories for firms in our sample. 158 firms (the leftmost bar) report no employed lab personnel in our sample period. The lower histogram bins the number of publications authored by firms in our sample. 268 firms (the leftmost bar) do not author any scientific publications in our sample period.

Figure 5: Corporate Science vs Gaps in University Science, by Industry




*Notes:* Industry-level scatter plots of firm investment in science and the gaps in the relevant academic discipline. The left panel plots the natural log of one plus the average publication stock against the gaps in public science measure. The right panel replaces the publication stock with the number of personnel at R&D labs, from the IRL directory.

Table 1: Corporate Scientific Publications, by OECD Subfield

| OECD Subfield | Number of Firms | Number of Papers | Avg Forward Publication Cites |
|---|---|---|---|
| 1.03 Physical sciences and astronomy | 63 | 461 | 1.64 |
| 1.06 Biological sciences | 32 | 83 | 1.37 |
| 2.03 Mechanical engineering | 89 | 186 | 1.12 |
| 2.02 Electrical eng, electronic eng | 156 | 1642 | 0.75 |
| 1.04 Chemical sciences | 76 | 344 | 0.72 |
| 1.07 Other natural sciences | 2 | 2 | 0.60 |
| 1.02 Computer and information sciences | 36 | 85 | 0.59 |
| 2.05 Materials engineering | 56 | 186 | 0.54 |
| 1.01 Mathematics | 36 | 95 | 0.52 |
| 3.02 Clinical medicine | 50 | 139 | 0.47 |
| 4.01 Agriculture, forestry, fisheries | 13 | 19 | 0.45 |
| 1.05 Earth and related environmental sciences | 64 | 136 | 0.35 |
| 2.08 Environmental biotechnology | 9 | 9 | 0.34 |
| 2.11 Other engineering and technologies | 61 | 119 | 0.33 |
| 2.06 Medical engineering | 8 | 25 | 0.27 |
| 3.01 Basic medical research | 23 | 30 | 0.22 |
| 2.07 Environmental engineering | 96 | 259 | 0.20 |
| 2.04 Chemical engineering | 21 | 24 | 0.18 |
| 3.03 Health sciences | 17 | 27 | 0.17 |
| 4.02 Animal and dairy science | 8 | 11 | 0.14 |
| 2.01 Civil engineering | 57 | 106 | 0.14 |
| 4.03 Veterinary science | 2 | 3 | 0.04 |
| 4.05 Other agricultural science | 12 | 14 | 0.02 |
| Not Available | 62 | 131 | 0.01 |

*Notes:* Observations are at OECD subfield level for years betwen 1926 and 1940. "Number of Firms" counts the number of firms publishing at least one article in the focal field. "Number of Papers" counts the number of total papers in the focal field. "Average Forward Publication Cites" take the field-level average of the normalized forward citations. Forward citations are normalized by the average number of forward citations received by all publications published in the focal publication's year.

Table 2: Summary Statistics of Main Variables

| | Observations | Mean | Median | Std. Dev. | Min | Max |
|---|---|---|---|---|---|---|
| Gap in university science (1900-1920) | 7035 | 0.70 | 0.70 | 0.03 | 0.58 | 0.79 |
| Lab Size | 2320 | 43.49 | 0.00 | 214.86 | 0.00 | 4669.00 |
| Patents Granted Per Year | 7035 | 13.13 | 1.00 | 54.88 | 0.00 | 838.00 |
| Publications Authored Per Year | 7035 | 0.46 | 0.00 | 3.72 | 0.00 | 88.00 |
| Patent Stock | 7035 | 71.18 | 7.54 | 316.85 | 0.00 | 4441.06 |
| Publication Stock | 7035 | 2.70 | 0.00 | 22.16 | 0.00 | 440.33 |
| Forward Patent Citations | 4035 | 0.77 | 0.62 | 0.80 | 0.00 | 18.42 |
| KPSS Patent Value | 2629 | 2.58 | 1.03 | 4.37 | 0.03 | 56.65 |
| Total Assets ($MM) | 4305 | 1369.71 | 418.87 | 3272.52 | 7.43 | 60114.66 |
| Gross Income ($MM) | 2789 | 879.27 | 270.66 | 1864.59 | -1.98 | 20655.93 |
| Market Capitalization ($MM) | 3856 | 1103.23 | 248.91 | 2903.02 | 0.69 | 37352.08 |
| Business Group Affiliated = 1 | 3104 | 0.41 | 0.00 | 0.49 | 0.00 | 1.00 |
| Total Assets ($MM) (1926-1930) | 1330 | 1258.17 | 445.75 | 2723.14 | 12.39 | 36047.36 |
| Gross Income ($MM) (1926-1930) | 728 | 833.95 | 235.17 | 1865.36 | -1.98 | 14366.30 |

*Notes:* Observations are at the firm-year level, and the sample period is 1926-1940. Forward Patent Citations is first defined at the patent level as the number of forward prior-art citations received normalized by the average number of forward citations for the patent's grant year cohort. This value is averaged at the focal firm-year level to produce the Forward Patent Citations measure in the table. Patent and publication stock are calculated using a perpetual inventory method with a 15% rate of depreciation. KPSS Patent Value is the value of a patent (in million dollars) based on the cumulative abnormal returns in the firm's market value at the issuance event of the patent (Kogan, Papanikolaou, Seru, & Stoffman, 2017).

Table 3: Investment in Corporate Science, by Firm Characteristics

| | | Avg(Publication Stock) | | Avg(Lab Personnel) | | Observations |
|---|---|---|---|---|---|---|
| Patenting Intensity Above Average? | Yes | 16.89 | (2.22) | 224.64 | (31.78) | 768 |
| | No | 0.75 | (0.04) | 19.41 | (1.55) | 3537 |
| Patent Forward Cites Above Average? | Yes | 5.82 | (0.66) | 89.70 | (10.55) | 2788 |
| | No | 0.65 | (0.04) | 11.28 | (1.17) | 4247 |
| Patents Cite Science? | Yes | 32.20 | (3.86) | 343.50 | (57.90) | 450 |
| | No | 0.68 | (0.03) | 22.75 | (1.94) | 6585 |
| Public? | Yes | 3.37 | (0.35) | 47.19 | (5.29) | 5250 |
| | No | 0.73 | (0.05) | 32.64 | (8.18) | 1785 |
| KKMY? | Yes | 4.90 | (0.53) | 73.30 | (8.87) | 3510 |
| | No | 0.51 | (0.03) | 14.45 | (1.22) | 3525 |
| Assets Above Average? | Yes | 7.38 | (0.87) | 109.46 | (13.20) | 2014 |
| | No | 0.34 | (0.03) | 11.28 | (1.09) | 2291 |
| Business Group? | Yes | 5.35 | (1.41) | 175.29 | (67.69) | 477 |
| | No | 4.63 | (1.20) | 101.07 | (23.15) | 693 |
| Competitive Industry? | Yes | 1.12 | (0.10) | 28.68 | (3.07) | 2865 |
| | No | 3.89 | (0.45) | 54.85 | (7.40) | 4050 |

*Notes:* Unit of analysis is the firm-year. Standard errors are indicated in parentheses. "Patenting Intensity" is defined as number of patents granted divided by log of assets. "Patents Cites Science" is equal to "Yes" if the firm's patents cite at least one scientific article and "No" otherwise (data from Marx and Fuegi (2020)). Observations with zero patents also enter the "No" category. "Public?" and "KKMY?" respectively ask whether the firm is a listed firm found in the CRSP dataset and the (Kandel, Kosenko, Morck, & Yafeh, 2019) dataset from Section 4. Observations enter the "Yes" row for "Assets Above Average?" if the total assets of the observations are above average. "Business Group?" asks whether firms are part of ownership chains with more than three firms per the definition in Kandel, Kosenko, Morck, and Yafeh (2019). "Competitive Industry?" rows classify firms based on their industry's competition classifications in Wilcox (1940); firms in industries classified as oligopolies and monopolies are classified in the "No" group.

Table 4: Corporate Publications and Gaps in University Science, by Firm Characteristics

| | DV: Publication stock | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Gap in university science, 1900-20 × Patenting Intensity | 99.098 | | | | | |
| | (27.004) | | | | | |
| Gap in university science, 1900-20 × Forward Patent Citations | | 45.762 | | | | |
| | | (11.949) | | | | |
| Gap in university science, 1900-20 × ln(Patent Cites to Science) | | | 1466.288 | | | |
| | | | (220.445) | | | |
| Gap in university science, 1900-20 × ln(Assets) (1926-1930) | | | | 140.875 | | |
| | | | | (23.226) | | |
| Gap in university science, 1900-20 × Business Group Dummy (Imputed) | | | | | 282.032 | |
| | | | | | (81.008) | |
| Gap in university science, 1900-20 × Competitive market | | | | | | -139.601 |
| | | | | | | (22.996) |
| Gap in university science, 1900-20 | -7.364 | 64.508 | 66.172 | -3614.816 | 31.827 | 132.589 |
| | (9.829) | (16.099) | (14.400) | (601.497) | (47.901) | (24.099) |
| Patenting Intensity | -64.052 | | | | | |
| | (19.291) | | | | | |
| Forward Patent Citations | | -29.700 | | | | |
| | | (8.143) | | | | |
| ln(Patent Cites to Science) | | | -994.021 | | | |
| | | | (156.863) | | | |
| ln(Assets) (1926-1930) | | | | -94.513 | | |
| | | | | (15.771) | | |
| Business Group Dummy (Imputed) | | | | | -198.801 | |
| | | | | | (56.328) | |
| ln(Assets) | -0.352 | 4.168 | 2.842 | 0.392 | 6.599 | 4.326 |
| | (0.248) | (0.584) | (0.429) | (0.643) | (1.081) | (0.601) |
| ln(American pubs), 1900-20 | -0.348 | 2.615 | 2.239 | 2.354 | 6.334 | 3.278 |
| | (0.580) | (1.016) | (0.796) | (1.303) | (3.197) | (1.050) |
| ln(European pubs), 1900-20 | 1.160 | 0.214 | 0.238 | 0.513 | 0.160 | -0.362 |
| | (0.556) | (0.880) | (0.645) | (1.193) | (2.779) | (0.888) |
| Average of Dependent Variable | 3.642 | 3.642 | 3.642 | 3.843 | 6.625 | 3.657 |
| Year Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.683 | 0.151 | 0.291 | 0.183 | 0.180 | 0.152 |
| Number of Firms | 422 | 422 | 422 | 358 | 199 | 420 |
| Number of Observations | 4,293 | 4,293 | 4,293 | 4,028 | 1,702 | 4,275 |

*Notes:* Unit of analysis is at the firm-year level. "Gap in university science, 1900-20" is defined at the firm level as the number of European scientific publications relevant to a firm divided by all (European & American) publications relevant to a firm (winsorized at the 1% and 99% level). "Patenting Intensity" is defined as number of patents granted divided by log of assets. "Patent Cites to Science" counts the number of in-text citations made by the focal firm's patents to scientific publications in MAG. "Business Group Dummy (Imputed)" refers to firms that are Business Group affiliates (defined as ownership chains with at least three firms). "ln(Assets) (1926-1930)" takes the natural log of average assets between 1926-1930. "Competitive market" is a dummy equal to one if the firm's industry is classified as "competitive" by Wilcox (1940). The rest of the variable definitions are identical to those in Table 2. Industry fixed effects are applied at 2-digit SIC codes. Standard errors are robust to arbitrary heteroscedasticity.

Table 5: Corporate Science and "Home-Run" Patents

| | DV: Top 5% Xi | | | DV: Top 5% Novelty | | |
|---|---|---|---|---|---|---|
| | (1) OLS | (2) 1st Stage IV | (3) 2nd Stage IV | (4) OLS | (5) 1st Stage IV | (6) 2nd Stage IV |
| Publication Stock (100s) | 0.040 | | 0.137 | 0.273 | | 0.484 |
| | (0.012) | | (0.027) | (0.020) | | (0.051) |
| Gap in university science, 1900-20 | | 95.653 | | | 74.670 | |
| | | (15.705) | | | (13.514) | |
| Gap in university science, current | -2.728 | -5.580 | -1.832 | 0.974 | -2.219 | 1.926 |
| | (1.103) | (5.115) | (0.984) | (1.627) | (4.079) | (1.585) |
| ln(Assets) | 1.419 | 2.529 | 1.188 | 0.761 | 2.408 | 0.276 |
| | (0.218) | (0.471) | (0.202) | (0.160) | (0.400) | (0.153) |
| ln(Patent stock) | 0.589 | 5.222 | 0.094 | 1.866 | 4.204 | 0.989 |
| | (0.081) | (0.563) | (0.091) | (0.152) | (0.477) | (0.156) |
| Average of Dependent Variable | 0.864 | | 0.861 | 2.903 | | 2.895 |
| Kleibergen-Paap rk Wald F statistic | | 37.097 | | | 30.530 | |
| Year Fixed Effects | Yes | | Yes | Yes | | Yes |
| Industry Fixed Effects | Yes | | Yes | Yes | | Yes |
| $R^2$ | 0.248 | | 0.023 | 0.635 | | 0.378 |
| Number of Firms | 327 | | 327 | 425 | | 425 |
| Number of Observations | 3,569 | 3,569 | 3,569 | 4,293 | 4,293 | 4,293 |

*Notes:* Analysis is at the firm-year level. The dependent variable for Columns 1 and 3 is the number of firm patents in the top 5% of stock market value (Kogan, Papanikolaou, Seru, & Stoffman, 2017). The dependent variable for Columns 4, 6 is the number of firm patents in the top 5% of novelty scores (Fleming, 2001). Columns 2 and 5 present first stage estimation results where dependent variable is publication stock. Instrument for Columns 2, 3, 5 and 6 is the share of European papers ("Gap in university science, 1900-20") for papers published between 1900 and 1920. "Gap in university science, current" calculates the share of European publications in the focal year. Industry fixed effects are applied at 2-digit SIC codes. Standard errors are robust to arbitrary heteroscedasticity.

Table 6: Corporate Science and Market-to-Book Ratios

| | Baseline (OLS) | Gap Split (OLS) | | IVE | |
|---|---|---|---|---|---|
| | (1) | (2) Small | (3) Large | (4) 1st Stage | (5) 2nd Stage |
| $\ln(\text{Pubstock}_{t-1})$ | 0.032 | -0.001 | 0.045 | | -0.040 |
| | (0.008) | (0.019) | (0.009) | | (0.081) |
| Gap in university science, 1900-20 | | | | 2.867 | |
| | | | | (0.492) | |
| $\ln(\text{Patstock}_{t-1})$ | 0.023 | 0.042 | 0.010 | 0.222 | 0.038 |
| | (0.004) | (0.007) | (0.006) | (0.013) | (0.018) |
| Gap in university science, current | -0.177 | 0.084 | -0.204 | -0.021 | -0.187 |
| | (0.083) | (0.153) | (0.103) | (0.186) | (0.082) |
| Average of Dependent Variable | 0.591 | 0.607 | 0.575 | | 0.591 |
| Kleibergen-Paap rk Wald F statistic | | | | 28.916 | |
| Year Fixed Effects | Yes | Yes | Yes | | Yes |
| Industry Fixed Effects | Yes | Yes | Yes | | Yes |
| $R^2$ | 0.346 | 0.400 | 0.396 | | 0.003 |
| Number of Firms | 325 | 170 | 155 | | 325 |
| Number of Observations | 3,399 | 1,740 | 1,659 | 3,399 | 3,399 |

*Notes:* Unit of analysis is at the firm-year level. Columns 1-3 present results from estimating the Tobin's Q equation against lagged publication stock. Columns 2 and 3 are split by mean values of the "Gap in university science, 1900-20" measure (2 being below average and 3 above). Column 4 presents the first stage IV estimates, where lagged publication stock is predicted by share of European publications published between 1900 and 1920 ("Gap in university science, 1900-20"). Column 5 regresses Tobin's Q against the predicted lagged publication stock from column 4. Industry fixed effects are applied at 2-digit SIC codes. Standard errors are robust to arbitrary heteroscedasticity.

# References


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

Arora, A., Belenzon, S., & Patacconi, A. (2019). A theory of the U.S. innovation ecosystem: Evolution and the social value of diversity [Publisher: Oxford University Press]. *Industrial and Corporate Change*, *28*(2), 289–307.

Arora, A., Belenzon, S., Patacconi, A., & Suh, J. (2020a). The changing structure of American innovation: Some cautionary remarks for economic growth. *Innovation Policy and the Economy*, *20*(1), 39–93.

Arora, A., Belenzon, S., & Sheer, L. (2020b). Back to Basics: Why do Firms Invest in Research? *American Economic Review*.

Arora, A., Belenzon, S., & Sheer, L. (2021). Knowledge spillovers and corporate investment in scientific research. *American Economic Review*, *111*(3), 871–98.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., & Van Reenen, J. (2020). The Fall of the Labor Share and the Rise of Superstar Firms*. *The Quarterly Journal of Economics*, *135*(2), 645–709. https://doi.org/10.1093/qje/qjaa004

Azoulay, P. (2002). Do pharmaceutical sales respond to scientific evidence? *Journal of Economics & Management Strategy*, *11*(4), 551–594.

Berle, A., & Means, G. (1932). Private property and the modern corporation. *New York: Mac-millan.*

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

Bonbright, J. C., & Means, G. C. (1932). *The holding company: Its public significance and its regulation.* McGraw-Hill Book Company.

Buchanan, N. S. (1936). The public utility holding company problem [Publisher: HeinOnline]. *Calif. L. Rev.*, *25*, 517.

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

Castelvecchi, D. (2016). Majority of mathematicians hail from just 24 scientific 'families'. *Nature News*, *537*(7618), 20.

Chandler Jr, A. D. (1993). *The visible hand*. Harvard University Press.

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

Cohen, W. M., & Levinthal, D. A. (1989). Innovation and learning: The two faces of R & D. *The economic journal*, *99*(397), 569–596.

Council, N. R. (1938). *Research: A National Resource : II. Industrial Research.* Washington, DC, The National Academies Press. https://doi.org/10.17226/20233

Field, A. J. (2003). The most technologically progressive decade of the century. *American Economic Review*, *93*(4), 1399–1413.

Fleming, D., & Bailyn, B. (1968). *The intellectual migration: Europe and America, 1930-1960* (Vol. 2). Charles Warren Center for Studies in American History, Harvard University.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management science*, *47*(1), 117–132.

Furman, J. L., & MacGarvie, M. J. (2007). Academic science and the birth of industrial research laboratories in the U.S. pharmaceutical industry [Publisher: Elsevier]. *Journal of Economic Behavior & Organization*, *63*(4), 756–776.

Geiger, R. L. (1986). *To advance knowledge : The growth of American research universities, 1900-1940.* New York : Oxford University Press, 1986.
Bibliography: p. 279-320.

Graham, J. R., Leary, M. T., & Roberts, M. R. (2015). A century of capital structure: The leveraging of corporate America [Publisher: Elsevier]. *Journal of Financial Economics*, *118*(3), 658–683.

Gross, D. P., & Sampat, B. N. (2020a). Inventing the Endless Frontier: The Effects of the World War II Research Effort on Post-War Innovation. *Harvard Business School Strategy Unit Working Paper*, (20-126).

Gross, D. P., & Sampat, B. N. (2020b). *Organizing Crisis Innovation: Lessons from World War II* (tech. rep.). National Bureau of Economic Research.

Gutiérrez, G., & Philippon, T. (2020). *Some facts about dominant firms* (tech. rep.). National Bureau of Economic Research.

Haynes, W. (1945). *American Chemical Industry: The World War I period: 1912-1922* (Vol. 3). Van Nostrand.

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.

Holton, G. (1981). The formation of the American physics Community in the 1920s and the coming of Albert Einstein [Publisher: Springer]. *Minerva*, *19*(4), 569–581.

Hounshell, D., & Smith, J. K. J. (1988). *Science and corporate strategy : Du Pont R&D, 1902-1980*. Cambridge [Cambridgeshire] ; New York : Cambridge University Press, 1988. https://find.library.duke.edu/catalog/DUKE000776227
Bibliography: p. 613-728.

Iaria, A., Schwarz, C., & Waldinger, F. (2018). Frontier knowledge and scientific production: Evidence from the collapse of international science [Publisher: Oxford University Press]. *The Quarterly Journal of Economics*, *133*(2), 927–991.

Jenkins, R., & Chandler, A. D. (1975). *Images and enterprise: Technology and the American photographic industry, 1839 to 1925*. Johns Hopkins University Press Baltimore.

Kandel, E., Kosenko, K., Morck, R., & Yafeh, Y. (2019). The great pyramids of America: A revised history of U.S. business groups, corporate ownership, and regulation, 1926–1950. *Strategic Management Journal*, *40*(5), 781–808.

Kevles, D. (1979). The physics, mathematics, and chemistry communities: A comparative analysis (A. Oleson & J. Voss, Eds.). In A. Oleson & J. Voss (Eds.), *The Organization of Knowledge in Modern America, 1860-1920*. Baltimore, Johns Hopkins University Press. Retrieved January 28, 2019, from http://resolver.caltech.edu/CaltechAUTHORS:20171127-134808791

Khanna, T. (2000). Business groups and social welfare in emerging markets: Existing evidence and unanswered questions [Publisher: Elsevier]. *European Economic Review*, *44*(4-6), 748–761.

Khanna, T. (2018). When technology gets ahead of society. *Harvard Business Review*, *96*(4), 86–95.

Khanna, T., & Palepu, K. (2000). Is group affiliation profitable in emerging markets? An analysis of diversified Indian business groups [Publisher: Wiley Online Library]. *The Journal of Finance*, *55*(2), 867–891.

Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics*, *132*(2), 665–712.

Lamoreaux, N. R., & Sokoloff, K. L. (1997). *Inventors, Firms, and the Market for Technology in the Late Nineteenth and Early Twentieth Centuries*. National Bureau of Economic Research Cambridge, Mass., USA.

Lundberg, F. (1937). *America's 60 families*. Vanguard Press, New York.

Maclaurin, W. R., & Harman, R. J. (1949). *Invention & innovation in the radio industry*. Macmillian.

MacLeod, W. B., & Urquiola, M. S. (2020). Why Does the U.S. Have the Best Research Universities? Incentives, Resources, and Virtuous Circles. *NBER Working Paper*.

Marx, M., & Fuegi, A. (2019). Reliance on science in patenting. *Available at SSRN*.

Marx, M., & Fuegi, A. (2020). Reliance on science: Worldwide front-page patent citations to scientific articles [Publisher: Wiley Online Library]. *Strategic Management Journal*.

Moser, P., & San, S. (2020). Immigration, Science, and Invention. Lessons from the Quota Acts. *Lessons from the Quota Acts (March 21, 2020)*.

Mowery, D., & Rosenberg, N. (1998). *Paths of innovation : Technological change in 20th-century America*. Cambridge ; New York : Cambridge University Press, 1998.
Includes bibliographical references (p. 181-199) and index.

Mowery, D. C. (1983). The relationship between intrafirm and contractual forms of industrial research in American manufacturing, 1900–1940 [Publisher: Elsevier]. *Explorations in Economic History*, *20*(4), 351–374.

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

Mowery, D. C., & Rosenberg, N. (1999). *Paths of innovation: Technological change in 20th-century America*. Cambridge University Press.

Nelson, R. R. (1959a). The economics of invention: A survey of the literature [Publisher: JSTOR]. *The Journal of Business*, *32*(2), 101–127.

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

Nicholas, T. (2010). The role of independent invention in U.S. technological development, 1880–1930 [Publisher: JSTOR]. *The Journal of Economic History*, 57–82.

Nicholas, T. (2011). Did R&D firms used to patent? Evidence from the first innovation surveys [Publisher: Cambridge University Press]. *The Journal of Economic History*, *71*(4), 1032–1059.

Nicholas, T. (2014). *Scale and innovation during two U.S. breakthrough eras*. Harvard Business School.

Nutter, G. W., & Einhorn, H. A. (1969). *Enterprise monopoly in the United States: 1899-1958*. Columbia University Press.

Perazich, G., & Field, P. M. (1940). *Industrial research and changing technology*. WPA, Nat. Research Project.

Pinchot, G. (1928). The power monopoly; its make-up and its menace.

Reich, L. S. (1983). Irving Langmuir and the Pursuit of Science and Technology in the Corporate Environment [The Johns Hopkins University Press, Society for the History of Technology]. *Technology and Culture*, *24*(2), 199–221. https://doi.org/10.2307/3104037

Reich, L. S. (1985). *The making of American industrial research : Science and business at GE and Bell, 1876-1926*. Cambridge Cambridgeshire ; New York, Cambridge University Press
Leonard S. Reich. ill. ; 24 cm. Includes index. Bibliography: p. 258-301.

Richter, R., & Streb, J. (2011). Catching-up and falling behind: Knowledge spillover from American to German machine toolmakers [Publisher: Cambridge University Press]. *The Journal of Economic History*, *71*(4), 1006–1031.

Rüegg, W. (2004). *A history of the university in Europe: Volume 3, universities in the nineteenth and early twentieth centuries (1800–1945)* (Vol. 3). Cambridge University Press.

Sauermann, H., & Cohen, W. M. (2010). What makes them tick? Employee motives and firm innovation. *Management Science*, *56*(12), 2134–2153.

Schumpeter, J. A. (1939). *Business cycles* (Vol. 1). McGraw-Hill New York.

Service, N. R. C. U. R. I. (1931). *Industrial Research Laboratories of the United States, Including Consulting Research Laboratories*. RR Bowker Company.

Servos, J. W. (1994). Changing partners: The Mellon Institute, private industry, and the federal patron [Publisher: JSTOR]. *Technology and Culture*, *35*(2), 221–257.

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

Stigler, G. J. (1949). Competition in the United States, In *Five Lectures on Economic Problems*. London, Longmans, Green; Co.

Urquiola, M. (2020). *Markets, Minds, and Money: Why America Leads the World in University Research*. Harvard University Press.

Wilcox, C. (1940). Competition and Monopoly in American Industry. *(Investigation of Concentration of Economic Power, T.N.E.C, Monograph No. 21)*.

Wise, G. (1985). *Willis R. Whitney, General Electric, and the origins of U.S. industrial research.* Columbia University Press.

# Appendix A Model

## A.1 Setup

There are three stages. In stage 3, the firms compete in the product market. Their product market performance depends on the quality of their products and the cost of producing them. We assume that cost and quality depend upon the innovation output, $d_i$, $i = 0, 1$. Their payoffs from stage 3 are $\Pi(d_0, d_1)$ and $\tilde{\Pi}(d_1, d_0)$, where the tilde indicates firm 1. We assume that $\Pi(d_0, d_1)$ is increasing in the first argument and decreasing in the second, and concave in its arguments, so that the firm's profit increases in its innovation output, albeit at a diminishing rate. To avoid the need for assumptions on third order derivatives, we assume

$$\Pi(d_0, d_1) = kd_0 - \frac{c_{00}}{2}d_0^2 - bd_1 - \frac{c_{11}}{2}d_1^2 + c_{01}d_1d_0, \quad k > 1$$
$$\tilde{\Pi}(d_1, d_0) = d_1 - \frac{c_{00}}{2}d_1^2 - bd_0 - \frac{c_{11}}{2}d_0^2 + c_{01}d_1d_0$$

Firms farther from the frontier (e.g., smaller firms) can increase profits by imitation and by increasing scale, possibilities that the leaders have already exhausted. Instead, leaders have to introduce new and improved products and processes-to innovate. Accordingly, the marginal product of innovation for firm 0 is greater than that of firm 1 because $k > 1$.

The coefficient $c_{01}$ is positive under strategic complementarity and negative under substitutability. Concavity of $\Pi$ implies $c_{00} > 0, c_{11} > 0, c_{00}c_{11} - c_{01}^2 \geq 0$. We assume that $b > 0$ so that $\frac{\partial \Pi}{\partial d_1} = -b - c_{11}d_1 < 0$, i.e., innovation by rivals reduces payoff. We also assume that $c_{00} \geq c_{11}$. This assumption implies that the marginal returns to internal invention decline faster than the rate at which profits decline due to invention by rivals.

In stage 2, firms choose their innovation output. Firm 0 chooses $d_0$ and firm 1 chooses $d_1$. The cost of innovation for firm 0 is $\phi(r_0; u)d_0$, where $r_0$ represents investments in internal scientific research by the firm, and $u$ indexes the stock of (relevant) public science. The cost of innovation includes the cost of inventing new products and processes or improving them. Internal research may directly lead to such inventions, but may also indirectly reduce the cost of invention by guiding the search for inventions in more promising directions. Innovations may also be based on inventions acquired from independent inventors, other firms or university researchers. Thus the cost of innovation also depends on the state of public science. It is natural to assume that both internal research and public science reduce the unit cost of innovation, $\phi(r_0; u)$, i.e., $\frac{\partial \phi}{\partial r_0} < 0, \frac{\partial \phi}{\partial u} < 0$, and diminishing returns so that $\frac{\partial^2 \phi}{\partial r_0^2} > 0$.

As we show below, the relationship between public science and internal research in the reduction in the unit cost of innovation will be important in how research investments relate to the stock of public science. The relationship may be one of strategic complementarity (in the sense of Milgrom and Roberts 1989). For instance, it is typically believed that public science would complement internal research efforts. However, public science may also lead to startups and independent inventors, who can license or sell their inventions, which can substitute for internally generated inventions. If so, the relationship may be one of strategic substitutability. Strategic complementarity exists if $-\frac{\partial^2 \phi}{\partial r_0 \partial u} > 0$, and substitutability exists if -$\frac{\partial^2 \phi}{\partial r_0 \partial u} < 0$. If $\frac{\partial^2 \phi}{\partial r_0 \partial u} = 0$, public science and research have independent effects on the cost of innovation.

The cost of innovation for firm 1 is $\tilde{\phi(u)}d_1$. As noted, innovations may be based on external discoveries and inventions. Thus, we assume that $\tilde{\phi(u)}$ decreases with $u$.

In stage 1, firm 0 choose its research investments, $r_0$, and the cost of research is modelled simply as $\frac{\gamma}{2}r_0^2$, so $v_0 = d_0 - \frac{c_{00}}{2}d_0^2 - bd_1 - \frac{c_{11}}{2}d_1^2 + c_{01}d_1d_0 - \phi(r_0, \lambda)d_0 - \frac{\gamma}{2}r_0^2$.

## A.2 Stage 2: Innovation

We assume a stable Nash Equilibrium exists. For a stable equilibrium, we require that $D = c_{00}^2 - c_{01}^2 > 0 \iff |c_{00}| > |c_{01}|$.

Note that as long as $k \geq 1 + (\phi - \tilde{\phi})$, $d_0 \geq d_1$. In particular, if neither firm invests in research, so that $\phi = \tilde{\phi}$, firm 0 would innovate more, and the gap increases the larger is $k$. This would imply that firm 0 has a greater incentive to invest in research. The following intermediate results are helpful for later results.

### A.2.1 Focal Firm Research and Innovation

The response of innovation output to the focal firm's research is

$$\begin{aligned} \frac{\partial d_0}{\partial r_0} &= \frac{c_{00}}{D}(-\frac{\partial \phi}{\partial r_0}) \\ \frac{\partial d_1}{\partial r_0} &= \frac{c_{01}}{D}(-\frac{\partial \phi}{\partial r_0}) \end{aligned} \tag{A1}$$

Note that if $c_{01} \geq 0$, firm 1 also increases its innovation in response to an increase in research by firm 0. Furthermore, $\frac{\partial^2 d_0}{\partial r_0 \partial u} = -\frac{c_{00}}{D}\frac{\partial^2 \phi}{\partial r_0 \partial u} \geq 0$ if $\frac{\partial^2 \phi}{\partial r_0 \partial u} \leq 0$, i.e., if public science and internal research are complements.

### A.2.2 Public Science and Innovation

The response of innovation output to public science is

$$\begin{aligned} \frac{\partial d_0}{\partial u} &= \frac{-1}{D}\left(c_{00}\frac{\partial \phi}{\partial u} + c_{01}\frac{\partial \tilde{\phi}}{\partial u}\right) \\ \frac{\partial d_1}{\partial u} &= \frac{-1}{D}\left(c_{00}\frac{\partial \tilde{\phi}}{\partial u} + c_{01}\frac{\partial \phi}{\partial u}\right) \end{aligned} \tag{A2}$$

If there is strategic complementarity, i.e., $c_{01} \geq 0$, both firms innovate more in response to an increase in public science. However, if there is strategic substitutability, then one (but not both) firms may reduce innovation. In particular, if the innovation costs of a firm are not very responsive to public science, the effect of a rival increasing its innovation may cause the firm to reduce its innovation.

## A.3 Stage 1: Research

Suppose firm 1 does not invest in research. Firm 0 chooses $r_0$, taking into account how its choice will affect the equilibrium choices of $d_0$ and $d_1$ in the stage 2 game. For firm 0, the first-order condition for optimal $r_0$, is

$$-\frac{\partial \phi}{\partial r_0}d_0 + \frac{\partial \Pi}{\partial d_1}\frac{\partial d_1}{\partial r_0} = \gamma r_0 \tag{A3}$$

The marginal return to research has a direct benefit represented by the first term: the reduction in the unit cost of innovation, which is proportional to the scale of innovation. The second term represents the feedback effect from competition in the innovation stage. By increasing innovation, research has a secondary benefit if it reduces innovation by the rival, which would be the case if there is strategic substitution in the innovation, so that $c_{01} \leq 0$. If innovations are strategic complements, then there is a secondary cost, because the second term would be negative. However, the first term is always larger than the second term. Substituting for $\frac{\partial d_1}{\partial r_0}$ from Equation A1 and gathering terms, Equation A3 can be rewritten as

$$-\frac{\partial\phi}{\partial r_0}\left(\frac{\partial\Pi}{\partial d_1}\frac{c_{01}}{D}+d_0\right)=\gamma r_0 \tag{A4}$$

Therefore, $\frac{\partial\Pi}{\partial d_1}\frac{c_{01}}{D}+d_0$ must be positive. A sufficient condition for this is strategic substitutability in innovation, $c_{01}\leq 0$. [54]

## A.4 Result 1: Innovation Leadership

Leaders earn higher profits. Conversely, the profits of the follower fall with the lead of firm 0. Formally,

$$\begin{aligned}\frac{\partial v}{\partial k}&=d_0+\frac{\partial\Pi}{\partial d_1}\frac{\partial d_1}{\partial k}\\&=d_0+\frac{\partial\Pi}{\partial d_1}\frac{c_{01}}{D}>0\text{ at an interior maximum}\\\frac{\partial\tilde{v}}{\partial k}&=\frac{\partial\tilde{\Pi}}{\partial d_0}\frac{\partial d_0}{\partial k}=\frac{\partial\tilde{\Pi}}{\partial d_0}\frac{c_{00}}{D}<0\end{aligned}\tag{A5}$$

Importantly, the returns to research of the innovation leader increase with its lead $k$. Those of the follower decrease if innovations are strategic substitutes and increase otherwise. Intuitively, as $k$ increases, the leader increases innovation. With strategic substitutes, the marginal return to innovation for the follower decreases. Given that research reduces the cost of innovation, the marginal return to research for the follower decreases.

$$\begin{aligned}\frac{\partial^2 v}{\partial k\partial r_0}&=\frac{\partial d_0}{\partial r_0}+\frac{c_{01}}{D}(-c_{11}\frac{\partial d_1}{\partial r_0}+c_{00}\frac{\partial d_0}{\partial r_0})\\&=\frac{c_{00}}{D}(-\frac{\partial\phi}{\partial r_0})+(-\frac{\partial\phi}{\partial r_0})\frac{c_{01}^2}{D}(c_{00}-c_{11})>0\\\frac{\partial^2\tilde{v}}{\partial k\partial r_1}&=\frac{c_{00}}{D}(-c_{11}\frac{\partial d_0}{\partial r_0}+c_{01}\frac{\partial d_1}{\partial r_1})=\frac{c_{00}}{D}(-\frac{\partial\tilde{\phi}}{\partial r_1})c_{01}(c_{00}-c_{11})\leq 0\iff c_{01}\leq 0\end{aligned}\tag{A6}$$

This result points to why firm 1 may not invest in research. If $k$ is large and there is strategic substitutability, firm 1's scale of innovation is small, thereby reducing its returns to innovation.

## A.5 Public Science

Equation A6 implies that if innovations are strategic substitutes, as the gap between leaders and followers grows, their incentives to invest in research diverge: leaders are more likely to invest in research, and followers are less likely to do so. If there is a fixed cost to such investment, then, for a range of such costs, we will have only firm 0 invest in research while firm 1 does not. In this section, we focus on the equilibrium where only firm 0 invests in research.

### A.5.1 The Value of the Firm

The value of the firm, $v$, may decrease with public science if public science substitutes for internal research, particularly if innovations are strategic complements. Intuitively, although public science reduces the cost of innovation, the innovation cost of the rival also declines. Increased innovation by the rival reduces value for the focal firm. If public science substitutes for internal research, it will be less effective in reducing the innovation cost of firm 0, i.e., $|\frac{\partial\phi}{\partial u}|<|\frac{\partial\tilde{\phi}}{\partial u}|$. Formally, the value of the firm is $v=\max_{r_0}\{\Pi-\gamma\frac{r_0^2}{2}\}$.

[54] We assume that the second order condition for an interior maximum holds. This requires that $\gamma$ be large.

Applying the envelope theorem, the effect of public science is given by

$$\begin{aligned}\frac{\partial v}{\partial u} &= -d_0\frac{\partial \phi}{\partial u} + \frac{\partial \Pi}{\partial d_1}\frac{\partial d_1}{\partial u} \\ &= -\frac{\partial \phi}{\partial u}\left(d_0 + \frac{\partial \Pi}{\partial d_1}\frac{c_{01}}{D}\right) - c_{00}\frac{\partial \Pi}{\partial d_1}\frac{\partial \tilde{\phi}}{\partial u}\end{aligned} \tag{A7}$$

Although the first term is positive by Equation A3, its magnitude depends on $|\frac{\partial \phi}{\partial u}|$. The second term is negative, and represents the effect due to the reduction in the rival's innovation cost. It is larger in magnitude the larger is $|\frac{\partial \tilde{\phi}}{\partial u}|$. Note that rivalry also matters. If $\dfrac{\partial \Pi}{\partial d_1} = -b + c_{01}d_0$ is large in magnitude (as would be the case for $b$ large and $c_{01} < 0$, the firm's value can decline with public science.

## A.6 Internal Research and Public Science

At an interior maximum, the direction of the effect of public science on internal research is given by $\dfrac{\partial^2 v}{\partial r_0 \partial u}$. Research increases with public science if $\dfrac{\partial^2 v}{\partial r_0 \partial u} \geq 0$ and decreases otherwise.

$$\begin{aligned}\frac{\partial^2 v}{\partial r_0 \partial u} &= \left(-\frac{\partial \phi}{\partial r_0}\right)\frac{\partial d_0}{\partial u} + d_0\left(-\frac{\partial^2 \phi}{\partial r_0 \partial u}\right) + \frac{\partial \Pi}{\partial d_1}\frac{\partial^2 d_1}{\partial r_0 \partial u} + \frac{\partial d_1}{\partial r_0}\frac{\partial^2 \Pi}{\partial d_1 \partial u} \\ &\text{substituting and collecting terms} \\ &= \left(-\frac{\partial \phi}{\partial r_0}\right)\frac{\partial d_0}{\partial u} - \frac{\partial^2 \phi}{\partial r_0 \partial u}\left(d_0 + \frac{\partial \Pi}{\partial d_1}\frac{c_{01}}{D}\right) + \frac{\partial d_1}{\partial r_0}\frac{\partial^2 \Pi}{\partial d_1 \partial u}\end{aligned} \tag{A8}$$

The first term in Equation A8 is positive. The second is positive if public science and research are strategic complements and negative otherwise. The third term is negative only if innovations are strategic complements and positive otherwise. Put differently, the first term reflects a direct effect: public science reduces innovation costs, and the resulting increase in innovation increases the marginal return to research. The second term represents the interaction between public science and research in reducing innovation costs. If they are complements, the second term also implies that the marginal return to research increases with public science. The third term captures the strategic interaction in innovation. If innovations are strategic substitutes, this term is also positive. Strategic complementarity is a necessary, but not sufficient, condition for this term to be negative. Thus, if internal research falls with public science, it implies that public science is a strategic substitute for research, or innovations are strategic complements, or both. These are one-way implications; even if they hold, public science could increase internal research if the direct effect, represented by the first term, is large.

To see this more fully, consider the case where there is neither complementarity nor substitution in the innovation stage, and where public science and research are independent. The latter implies that $\frac{\partial^2 \phi}{\partial r_0 \partial u} = 0$, and the former implies that $\frac{\partial d_1}{\partial r_0} = 0$. In that case, Equation A8 has a single term $\left(-\frac{\partial \phi}{\partial r_0}\right)\frac{\partial d_0}{\partial u} \geq 0$. *That is, if public science and research are independent and there are no strategic interactions in the innovation stage, internal research increases with public science because public science increases the scale of innovation, thereby increasing the marginal return to research.*

If there are no strategic interactions in innovation, Equation A8 is $\left(-\frac{\partial \phi}{\partial r_0}\right)\frac{\partial d_0}{\partial u} - \frac{\partial^2 \phi}{\partial r_0 \partial u}\left(d_0 + \frac{\partial \Pi}{\partial d_1}\frac{c_{01}}{D}\right)$. The second term is non-negative if $-\frac{\partial^2 \phi}{\partial r_0 \partial u} \geq 0$, i.e., if public science and internal research are complements and negative otherwise. *Therefore, if internal research declines with public science, and there are not strategic interactions in innovation, it implies that public science and internal research are strategic substitutes.*

The third term can be written as

$$\frac{\partial d_1}{\partial r_0}\frac{\partial^2 \Pi}{\partial d_1 \partial u} = \frac{\partial d_1}{\partial r_0}\left[\frac{\partial^2 \Pi}{\partial d_1 \partial d_0}\frac{\partial d_0}{\partial u} + \frac{\partial^2 \Pi}{\partial d_1^2}\frac{\partial d_2}{\partial u}\right]$$

$$= \frac{\partial d_1}{\partial r_0}\frac{1}{D}\left[-c_{11}c_{00}(-\frac{\partial \tilde{\phi}}{\partial u}) + -c_{11}c_{01}(-\frac{\partial \phi}{\partial u}) + c_{01}c_{00}(-\frac{\partial \phi}{\partial u}) + c_{01}^2(-\frac{\partial \tilde{\phi}}{\partial u})\right] \tag{A9}$$

collecting terms and substituting

$$\frac{\partial d_1}{\partial r_0}\frac{\partial^2 \Pi}{\partial d_1 \partial u} = \frac{c_{01}}{D^2}(-\frac{\partial \tilde{\phi}}{\partial u})(c_{01}^2 - c_{00}c_{11}) + \frac{c_{01}^2}{D^2}(-\frac{\partial \phi}{\partial u})(c_{00} - c_{11})$$

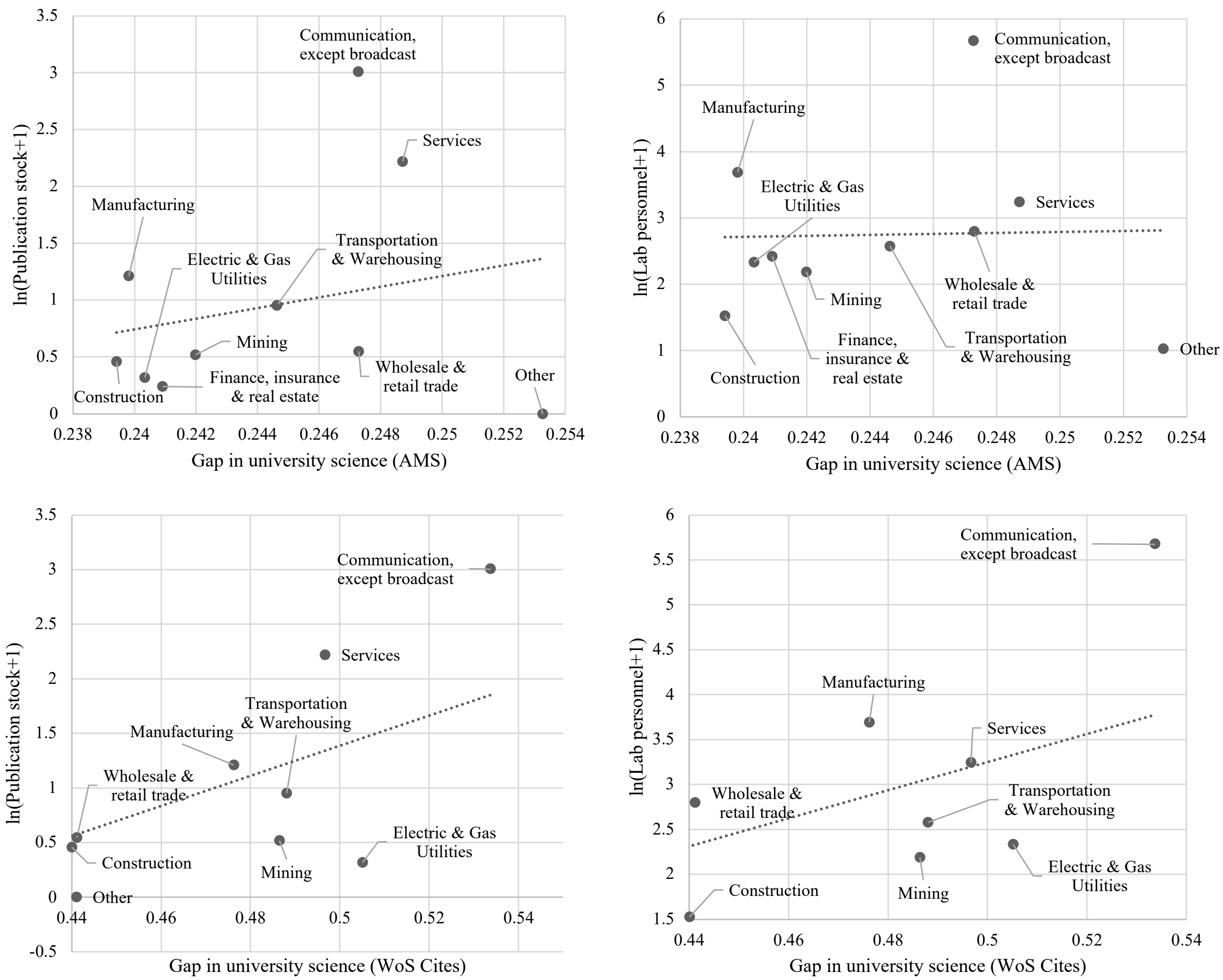
Note that $c_{00} \geq c_{11}$, so that $\frac{c_{01}^2}{D^2}(-\frac{\partial \phi}{\partial u})(c_{00} - c_{11}) \geq 0$. Also, $-\frac{\partial \tilde{\phi}}{\partial u}(c_{01}^2 - c_{00}c_{11}) \leq 0$ by the concavity of $\Pi$. Thus, $\frac{c_{01}}{D^2}(-\frac{\partial \tilde{\phi}}{\partial u})(c_{01}^2 - c_{00}c_{11}) > 0$ if $c_{01} < 0$ and negative otherwise. Therefore, a necessary condition for the expression in Equation A9 to be negative is that innovations be strategic complements. The conclusion is that for public science to reduce research, it would require that either innovations be strategic complements, or that public science be a strategic substitute for internal research. Else, public science will increase research by the leading firm.

### A.6.1 The Gap Between the Leader and Follower, the Returns to Research, and Public Science

Recall from Equation A6 that the marginal returns from research to the leader as $k$ increases is given by $\frac{c_{00}}{D}(-\frac{\partial \phi}{\partial r_0}) + (-\frac{\partial \phi}{\partial r_0})\frac{c_{01}^2}{D}(c_{00} - c_{11})$. It is easy to see that this expression is increasing in $u$ if public science and internal research are complements – $\frac{\partial^2 \phi}{\partial r_0 \partial u} \geq 0$–and decreasing otherwise. Similarly, the effect on the marginal returns from research to the follower of $k$ increases is given by $\frac{c_{00}}{D}(-\frac{\partial \tilde{\phi}}{\partial r_1})c_{01}(c_{00} - c_{11})$. This expression falls with $u$ if innovations are strategic complements, $c_{01} > 0$, and public science and research are substitutes, $\frac{\partial^2 \tilde{\phi}}{\partial r_1 \partial u} \geq 0$, or if innovations are strategic substitutes but public science and research are complements. Otherwise, the marginal returns of the follower also increase with $u$.

# Appendix B Auxiliary Results

Figure B1: Corporate Science vs Gaps in University Science, by Industry




*Notes:* Industry-level scatter plots of firm investment in science and the gaps in the relevant academic discipline. The left panels plot the natural log of one plus the average publication stock against the gaps in public science measures. The right panel replaces the publication stock with the number of personnel at R&D labs from the IRL directory. The upper panels measure gaps using the number of foreign-trained American scientists in the AMS directory. The lower panels measure gaps using share of transatlantic citations made by American journals.

In Table B1, we replicate the result in Table 4 by using the AMS scientist-based measure of the gap. The directions of the interaction term coefficient for forward patent citations is identical, though we fail to replicate statistical significance. However, the coefficient for patent citations to science is positive and significant. The effects of corporate size, group affiliation and competitive industries all have the same signs and are statistically significant, as in Table 4.

Table B2 uses the journal citations-based gap measure. All results except the interaction with patent citations to science (Column 2) are also statistically significant at the 5% level (in Column 2, the interaction coefficient is significant at the 10% level (t= 1.95)).

Table B1: Corporate Publications and Gaps in University Science (Scientists), by Firm Characteristics

| | DV: Publication stock | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Gap in university science, 1921 (Scientists) × Patenting Intensity | 183.739 | | | | | |
| | (103.126) | | | | | |
| Gap in university science, 1921 (Scientists) × Forward Patent Citations | | 129.790 | | | | |
| | | (40.331) | | | | |
| Gap in university science, 1921 (Scientists) × ln(Patent Cites to Science) | | | 4171.920 | | | |
| | | | (792.562) | | | |
| Gap in university science, 1921 (Scientists) × ln(Assets) (1926-1930) | | | | 439.018 | | |
| | | | | (71.239) | | |
| Gap in university science, 1921 (Scientists) × Business Group Dummy (Imputed) | | | | | 623.660 | |
| | | | | | (243.421) | |
| Gap in university science, 1921 (Scientists) × Competitive market | | | | | | -447.074 |
| | | | | | | (67.875) |
| Gap in university science, 1921 (Scientists) | -21.778 | 193.082 | 213.533 | -11315.717 | 233.304 | 420.305 |
| | (40.215) | (65.536) | (61.361) | (1840.963) | (167.575) | (83.332) |
| Forward Patent Citations | | -29.004 | | | | |
| | | (9.543) | | | | |
| ln(Patent Cites to Science) | | | -981.860 | | | |
| | | | (197.172) | | | |
| ln(Assets) (1926-1930) | | | | -101.985 | | |
| | | | | (16.710) | | |
| Business Group Dummy (Imputed) | | | | | -152.217 | |
| | | | | | (58.545) | |
| Patenting Intensity | -38.056 | | | | | |
| | (25.508) | | | | | |
| ln(Assets) | -0.668 | 4.062 | 1.231 | 0.817 | 6.478 | 4.110 |
| | (0.280) | (0.570) | (0.228) | (0.619) | (1.052) | (0.577) |
| ln(American scientists), 1921 | 8.372 | 16.108 | 21.474 | 15.549 | 18.959 | 11.425 |
| | (4.388) | (7.211) | (7.082) | (7.360) | (18.070) | (7.135) |
| ln(European scientists), 1921 | -7.155 | -13.294 | -18.677 | -12.742 | -10.873 | -8.497 |
| | (4.432) | (7.141) | (6.989) | (7.302) | (17.976) | (7.088) |
| ln(Patent Granted) | | | 4.255 | | | |
| | | | (0.672) | | | |
| Average of Dependent Variable | 3.642 | 3.642 | 3.642 | 3.843 | 6.625 | 3.657 |
| Year Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.666 | 0.151 | 0.317 | 0.180 | 0.184 | 0.151 |
| Number of Firms | 422 | 422 | 422 | 358 | 199 | 420 |
| Number of Observations | 4,293 | 4,293 | 4,293 | 4,028 | 1,702 | 4,275 |

*Notes:* Unit of analysis is at the firm-year level. "Gap in university science, 1921 (Scientists)" is defined as the share of of American scientists with a European affiliation (winsorized at 1% and 99%). The other variable definitions are identical to those in Tables 2 and 4. Industry fixed effects are applied at 2-digit SIC codes. Standard errors are robust to arbitrary heteroscedasticity.

Table B2: Corporate Publications and Gaps in University Science (Cites), by Firm Characteristics

| | DV: Publication stock | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Gap in university science, 1900-20 (Cites) × Patenting Intensity | 39.254 | | | | | |
| | (7.780) | | | | | |
| Gap in university science, 1900-20 (Cites) × Forward Patent Citations | | 16.649 | | | | |
| | | (4.808) | | | | |
| Gap in university science, 1900-20 (Cites) × ln(Patent Cites to Science) | | | 508.747 | | | |
| | | | (260.366) | | | |
| Gap in university science, 1900-20 (Cites) × ln(Assets) (1926-1930) | | | | 25.936 | | |
| | | | | (4.890) | | |
| Gap in university science, 1900-20 (Cites) × Business Group Dummy (Imputed) | | | | | 36.773 | |
| | | | | | (17.107) | |
| Gap in university science, 1900-20 (Cites) × Competitive market | | | | | | -13.994 |
| | | | | | | (5.512) |
| Gap in university science, 1900-20 (Cites) | 29.832 | 58.684 | 52.539 | -579.706 | 183.879 | 77.714 |
| | (5.756) | (13.663) | (10.589) | (117.909) | (32.884) | (15.127) |
| Patenting Intensity | -12.383 | | | | | |
| | (3.627) | | | | | |
| Forward Patent Citations | | -6.302 | | | | |
| | | (2.181) | | | | |
| ln(Patent Cites to Science) | | | -187.051 | | | |
| | | | (122.358) | | | |
| ln(Assets) (1926-1930) | | | | -8.169 | | |
| | | | | (1.903) | | |
| Business Group Dummy (Imputed) | | | | | -16.687 | |
| | | | | | (7.683) | |
| ln(Assets) | -0.440 | 4.170 | 2.900 | 0.565 | 7.341 | 4.264 |
| | (0.224) | (0.585) | (0.426) | (0.602) | (1.191) | (0.595) |
| ln(Cites to America), 1900-20 | 7.434 | 10.426 | 7.596 | 20.114 | 33.367 | 12.700 |
| | (1.229) | (3.438) | (2.006) | (4.022) | (6.924) | (3.473) |
| ln(Cites to Europe), 1900-20 | -6.433 | -7.276 | -4.938 | -16.569 | -25.661 | -9.625 |
| | (1.206) | (3.278) | (1.777) | (3.710) | (6.380) | (3.275) |
| Average of Dependent Variable | 3.642 | 3.642 | 3.642 | 3.843 | 6.625 | 3.657 |
| Year Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.683 | 0.147 | 0.273 | 0.159 | 0.177 | 0.146 |
| Number of Firms | 422 | 422 | 422 | 358 | 199 | 420 |
| Number of Observations | 4,293 | 4,293 | 4,293 | 4,028 | 1,702 | 4,275 |

*Notes:* Unit of analysis is at the firm-year level. "Gap in university science, 1900-20 (Cites)" is defined at the firm level as American journal backward citations to Europe divided by total American backward journal citations (winsorized at 1% and 99%). The other variable definitions are identical to those in Tables 2 and 4. Industry fixed effects are applied at 2-digit SIC codes. Standard errors are robust to arbitrary heteroscedasticity.

Table B3: Corporate Science and "Home-Run" Patents (AMS Gaps)

| | DV: Top 5% Xi | | | DV: Top 5% Novelty | | |
|---|---|---|---|---|---|---|
| | (1) OLS | (2) 1st Stage IV | (3) 2nd Stage IV | (4) OLS | (5) 1st Stage IV | (6) 2nd Stage IV |
| Publication Stock (100s) | 0.040 | | 0.182 | 0.273 | | 0.557 |
| | (0.012) | | (0.041) | (0.020) | | (0.075) |
| Gap in university science, 1921 (Scientists) | | 250.154 | | | 202.045 | |
| | | (49.513) | | | (42.007) | |
| Gap in university science, current (Pubs) | -2.728 | -2.779 | -1.410 | 0.974 | 0.018 | 2.252 |
| | (1.103) | (5.298) | (1.047) | (1.627) | (4.155) | (1.744) |
| ln(Assets) | 1.419 | 2.383 | 1.080 | 0.761 | 2.273 | 0.110 |
| | (0.218) | (0.458) | (0.186) | (0.160) | (0.384) | (0.210) |
| ln(Patent stock) | 0.589 | 5.172 | -0.139 | 1.866 | 4.170 | 0.689 |
| | (0.081) | (0.559) | (0.144) | (0.152) | (0.473) | (0.239) |
| Average of Dependent Variable | 0.864 | | 0.861 | 2.903 | | 2.895 |
| Kleibergen-Paap rk Wald F statistic | | 25.525 | | | 23.134 | |
| Year Fixed Effects | Yes | | Yes | Yes | | Yes |
| Industry Fixed Effects | Yes | | Yes | Yes | | Yes |
| $R^2$ | 0.248 | | -0.177 | 0.635 | | 0.207 |
| Number of Firms | 327 | | 327 | 425 | | 425 |
| Number of Observations | 3,569 | 3,569 | 3,569 | 4,293 | 4,293 | 4,293 |

*Notes:* Analysis is at the firm-year level. The dependent variable for Columns 1 and 3 is the number of firm patents in the top 5% of stock market value (Kogan, Papanikolaou, Seru, & Stoffman, 2017). The dependent variable for Columns 4, 6 is the number of firm patents in the top 5% of novelty scores (Fleming, 2001). Columns 2 and 5 present first stage estimation results where dependent variable is publication stock. Instrument for Columns 2, 3, 5, 6 is the share of European-affiliated scientists ("Gap in university science, 1921 (Scientists)"). "Gap in university science, current (Pubs)" calculates the share of European papers for the focal year. Industry fixed effects are applied at 2-digit SIC codes. Standard errors are robust to arbitrary heteroscedasticity.

Table B4: Corporate Science and "Home-Run" Patents (WOS Citation Gaps)

| | DV: Top 5% Xi | | | DV: Top 5% Novelty | | |
|---|---|---|---|---|---|---|
| | (1) OLS | (2) 1st Stage IV | (3) 2nd Stage IV | (4) OLS | (5) 1st Stage IV | (6) 2nd Stage IV |
| Publication Stock (100s) | 0.041 | | 0.108 | 0.272 | | 0.268 |
| | (0.012) | | (0.034) | (0.020) | | (0.094) |
| Gap in university science, 1900-20 (Cites) | | 20.698 | | | 10.927 | |
| | | (4.848) | | | (3.642) | |
| Gap in university science, current (Cites) | -3.087 | 11.028 | -4.774 | 1.788 | 14.544 | 1.874 |
| | (1.040) | (4.910) | (1.356) | (1.082) | (3.996) | (2.282) |
| ln(Assets) | 1.428 | 2.327 | 1.274 | 0.758 | 2.279 | 0.767 |
| | (0.220) | (0.451) | (0.222) | (0.161) | (0.386) | (0.233) |
| ln(Patent stock) | 0.582 | 5.261 | 0.232 | 1.871 | 4.238 | 1.888 |
| | (0.080) | (0.568) | (0.167) | (0.153) | (0.482) | (0.411) |
| Average of Dependent Variable | 0.864 | | 0.861 | 2.903 | | 2.895 |
| Kleibergen-Paap rk Wald F statistic | | 18.225 | | | 9.001 | |
| Year Fixed Effects | Yes | | Yes | Yes | | Yes |
| Industry Fixed Effects | Yes | | Yes | Yes | | Yes |
| $R^2$ | 0.248 | | 0.112 | 0.636 | | 0.592 |
| Number of Firms | 327 | | 327 | 425 | | 425 |
| Number of Observations | 3,569 | 3,569 | 3,569 | 4,293 | 4,293 | 4,293 |

*Notes:* Analysis is at the firm-year level. The dependent variable for Columns 1 and 3 is the number of firm patents in the top 5% of stock market value (Kogan, Papanikolaou, Seru, & Stoffman, 2017). The dependent variable for Columns 4, 6 is the number of firm patents in the top 5% of novelty scores (Fleming, 2001). Columns 2 and 5 present first stage estimation results where dependent variable is publication stock. Instrument for Columns 2, 3, 5, 6 is the share of American journal citations made to European journals ("Gap in university science, 1900-20 (Cites)"). "Gap in university science, current (Cites)" calculates the backward citation share of European journals in American journals published in the focal year. Industry fixed effects are applied at 2-digit SIC codes. Standard errors are robust to arbitrary heteroscedasticity.

Table B5: Corporate Science and Market-to-Book Ratios

| | Cite Gap (OLS) | | Scientist Gap (OLS) | |
|---|---|---|---|---|
| | (1) Small | (2) Large | (3) Small | (4) Large |
| $\ln(\text{Pubstock}_{t-1})$ | 0.041 | 0.067 | 0.016 | 0.052 |
| | (0.014) | (0.011) | (0.019) | (0.009) |
| $\ln(\text{Patstock}_{t-1})$ | 0.051 | -0.019 | 0.026 | 0.009 |
| | (0.006) | (0.006) | (0.008) | (0.005) |
| Gap in university science, current (Cites) | -0.332 | 0.175 | -0.233 | -0.101 |
| | (0.185) | (0.185) | (0.163) | (0.121) |
| Average of Dependent Variable | 0.617 | 0.563 | 0.605 | 0.580 |
| Year Fixed Effects | Yes | Yes | Yes | Yes |
| Industry Fixed Effects | Yes | Yes | Yes | Yes |
| $R^2$ | 0.402 | 0.429 | 0.394 | 0.402 |
| Number of Firms | 169 | 156 | 149 | 176 |
| Number of Observations | 1,797 | 1,601 | 1,532 | 1,867 |

*Notes:* Unit of analysis is at the firm-year level. Dependent variable is Tobin's Q. Columns 1 and 2 are split by mean values of "Gap in university science, 1900-20 (Cites)" measures (1 being below average and 2 being above). Columns 3 and 4 are split by mean values of the the "Gap in university science, 1921 (Scientists)" measures (3 being below average and 4 being above). Industry fixed effects are applied at 2-digit SIC codes. Standard errors are robust to arbitrary heteroscedasticity.

We replicate the same results of Table 6 in Columns 1 and 2 by splitting the sample by average citation gap scores. We find that the correlation between Q and gaps are 1.6 times larger for those above average gap scores compared to those below the average. A similar mean-split based on average AMS gaps (based on scientist bios) in Columns 3 and 4 shows the coefficients to be 3.3 times for higher gap firms.

# Appendix C Details on Data Construction

## C.1 Corporate Historical Documents and Data Sources

- Bureau of Economic Analysis (BEA, 1958), U.S. Department of Commerce, Benchmark Federal Trade Commission (FTC) Annual Reports: www.ftc.gov/os/annualreports/index.shtm
- Input-Output Data: Historical SIC Data, www.bea.gov/industry/io_histsic.htm
- Interstate Commerce Commission (ICC) Reports
- Moody's Manuals, 1926-1940: http://webreports.mergent.com/
- Statistics of Income: http://www.irs.gov/pub/irs-soi/
- National Association of Railroad and Utility Commissioners
- National Resources Committee (NRC) (1939), The Structure of the American Economy (Washington, DC: U.S. Government Print Office)
- Regulation of Stock Ownership in Railroads, 71st Congress, 3d Session, House Report No. 2789, Vol.2, February 1931
- Securities and Exchange Commission (SEC) Annual Reports: www.sec.gov/about/annrep.shtml
- Survey of American Listed Corporations: Reported Information on Registrants with the SEC under the Securities Exchange Act of 1934, 1939-40
- Temporary National Economic Committee (TNEC), (1940), The Distribution of Ownership in the 200 Largest Nonfinancial Corporations, monograph 29 (1-2) (Washington, DC: U.S. Government Printing Office): http://www.bpl.org/govinfo/online-collections/federal-executive-branch/temporary-national-e
- Twentieth Century Fund, Committee on Taxation (1937), Facing the Tax Problem (New York: Twentieth Century Fund)

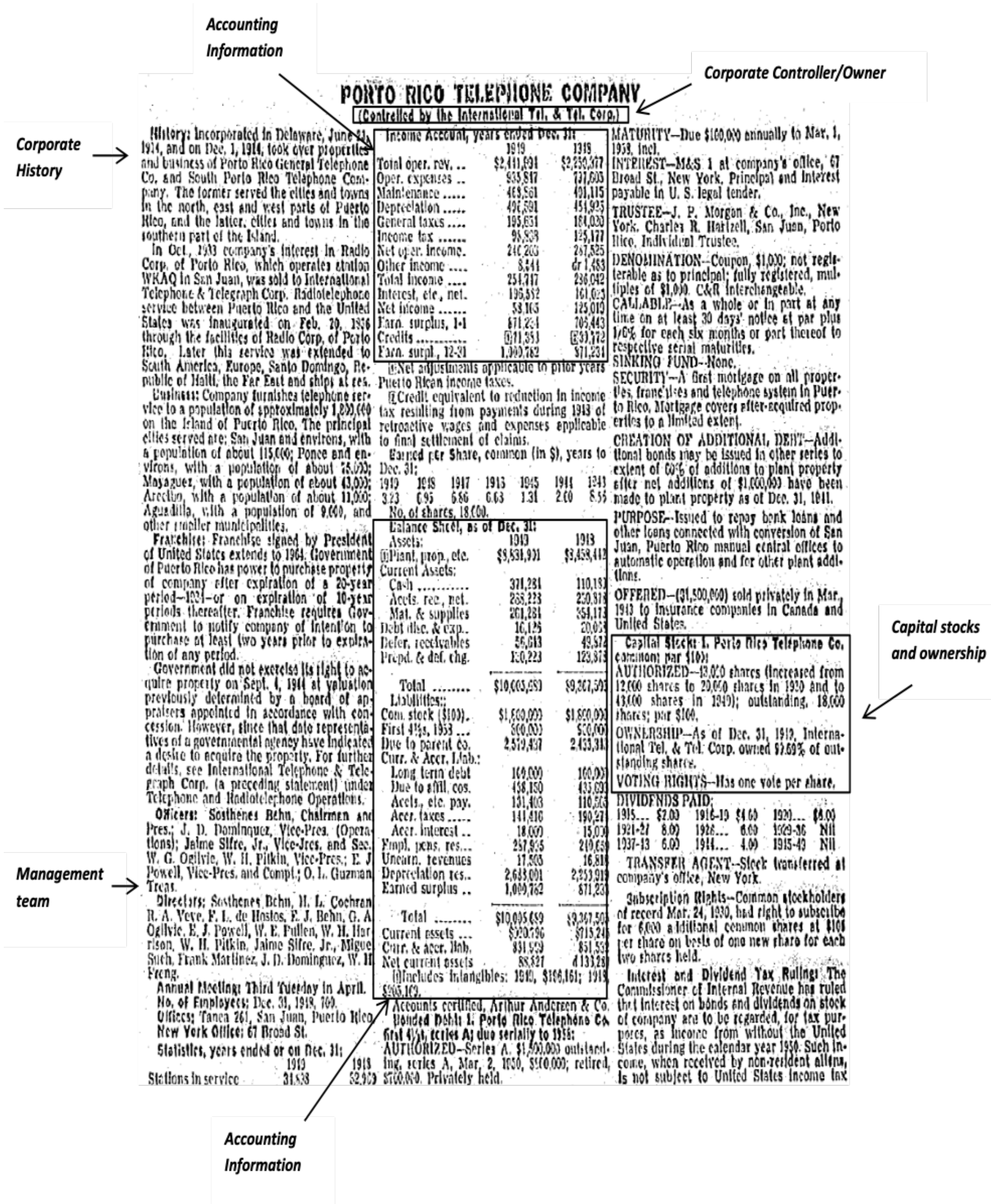
## C.2 Corporate Histories

- http://www.Archive.org
- Encyclopedia of American Business History (Facts on File, 2005): http://www.Fundinguniverse.com
- The New York Times Archives: http://www.nytimes.com/ref/membercenter/nytarchive.html
- The Wall Street Journal Archives: http://pqasb.pqarchiver.com/wsj/search.html

## C.3 Control Chains

We use Moody's Manuals to track companies controlling, or controlled by, the 200 companies on the B&M list. In each volume, a company report is followed by reports on its controlled subsidiaries (which are identified without an explicitly specified control threshold held by the controlling company). For example, if company A controls company B and company B, in turn, controls company C, and all three firms belong to the railroad sector, the A-B-C control chain will appear in Moody's Railroads Manual in the same sequence with the identity of the corporate controller usually reported next to the company name. We examine if one or more companies are controlled by another corporation included in the original list and, if this is the case, combine their control chains. Therefore, each control chain in our

Figure C1: A Moody's Manuals entry: The Porto Rico Telephone Company, 1949




# PORTO RICO TELEPHONE COMPANY

(Controlled by the International Tel. & Tel. Corp.)

**History:** Incorporated in Delaware, June 21, 1914, and on Dec. 1, 1914, took over properties and business of Porto Rico General Telephone Co. and South Porto Rico Telephone Company. The former served the cities and towns in the north, east and west parts of Puerto Rico, and the latter, cities and towns in the southern part of the Island.

In Oct., 1933 company's interest in Radio Corp. of Porto Rico, which operates station WKAQ in San Juan, was sold to International Telephone & Telegraph Corp. Radiotelephone service between Puerto Rico and the United States was inaugurated on Feb. 20, 1936 through the facilities of Radio Corp. of Porto Rico. Later this service was extended to South America, Europe, Santo Domingo, Republic of Haiti, the Far East and ships at sea.

**Business:** Company furnishes telephone service to a population of approximately 1,800,000 on the Island of Puerto Rico. The principal cities served are: San Juan and environs, with a population of about 115,000; Ponce and environs, with a population of about 75,000; Mayaguez, with a population of about 43,000; Arecibo, with a population of about 11,000; Aguadilla, with a population of 9,000, and other smaller municipalities.

**Franchise:** Franchise signed by President of United States extends to 1964. Government of Puerto Rico has power to purchase property of company after expiration of a 20-year period—1934—or on expiration of 10-year periods thereafter. Franchise requires Government to notify company of intention to purchase at least two years prior to expiration of any period.

Government did not exercise its right to acquire property on Sept. 1, 1944 at valuation previously determined by a board of appraisers appointed in accordance with concession. However, since that date representatives of a governmental agency have indicated a desire to acquire the property. For further details, see International Telephone & Telegraph Corp. (a preceding statement) under Telephone and Radiotelephone Operations.

**Officers:** Sosthenes Behn, Chairman and Pres.; J. D. Dominquez, Vice-Pres. (Operations); Jaime Sifre, Jr., Vice-Pres. and Sec.; W. G. Ogilvie, W. H. Pitkin, Vice-Pres.; E. J. Powell, Vice-Pres. and Compt.; O. L. Guzman, Treas.

**Directors:** Sosthenes Behn, H. L. Cochran, R. A. Veve, F. L. de Hostos, E. J. Behn, G. A. Ogilvie, E. J. Powell, W. E. Pullen, W. H. Harrison, W. H. Pitkin, Jaime Sifre, Jr., Miguel Such, Frank Martinez, J. D. Dominguez, W. H. Freng.

**Annual Meeting:** Third Tuesday in April.

**No. of Employees:** Dec. 31, 1948, 700.

**Offices:** Tanca 261, San Juan, Puerto Rico. **New York Office:** 67 Broad St.

**Statistics, years ended or on Dec. 31:**

| | 1949 | 1948 |
|---|---|---|
| Stations in service | 31,838 | 32,909 |

**Income Account, years ended Dec. 31:**

| | 1949 | 1948 |
|---|---|---|
| Total oper. rev. .. | $2,411,891 | $2,250,377 |
| Oper. expenses .. | 935,817 | 737,603 |
| Maintenance ..... | 463,561 | 401,115 |
| Depreciation ..... | 496,591 | 454,925 |
| General taxes .... | 195,651 | 184,030 |
| Income tax ...... | 98,808 | 125,177 |
| Net oper. income. | 240,203 | 267,525 |
| Other income .... | 8,544 | dr 1,483 |
| Total income .... | 254,747 | 236,042 |
| Interest, etc., net. | 196,582 | 161,023 |
| Net income ...... | 58,165 | 125,019 |
| Earn. surplus, 1-1 | 871,234 | 705,443 |
| Credits ........... | [a]71,353 | [b]39,772 |
| Earn. surpl., 12-31 | 1,000,762 | 871,231 |

[a]Net adjustments applicable to prior years Puerto Rican income taxes.

[b]Credit equivalent to reduction in income tax resulting from payments during 1948 of retroactive wages and expenses applicable to final settlement of claims.

**Earned per Share, common (in $), years to Dec. 31:**

| 1949 | 1948 | 1947 | 1946 | 1945 | 1944 | 1943 |
|---|---|---|---|---|---|---|
| 3.23 | 6.95 | 6.86 | 6.63 | 1.31 | 2.00 | 8.55 |

No. of shares, 18,000.

**Balance Sheet, as of Dec. 31:**

| Assets: | 1949 | 1948 |
|---|---|---|
| [c]Plant, prop., etc. | $8,831,901 | $8,458,412 |
| Current Assets: | | |
| Cash ........... | 371,281 | 110,181 |
| Accts. rec., net. | 288,223 | 250,818 |
| Mat. & supplies | 261,281 | 354,173 |
| Debt disc. & exp.. | 16,125 | 20,063 |
| Defer. receivables | 56,613 | 49,574 |
| Prepd. & def. chg. | 130,223 | 123,873 |
| Total ........ | $10,005,689 | $9,367,503 |
| Liabilities: | | |
| Com. stock ($100). | $1,800,000 | $1,800,000 |
| First 4½s, 1953 ... | 800,000 | 800,000 |
| Due to parent co. | 2,579,437 | 2,433,311 |
| Curr. & Accr. Liab.: | | |
| Long term debt | 100,000 | 100,000 |
| Due to affil. cos. | 458,150 | 435,691 |
| Accts., etc. pay. | 131,403 | 110,563 |
| Accr. taxes ..... | 141,410 | 190,271 |
| Accr. interest .. | 18,000 | 15,000 |
| Empl. pens. res... | 257,935 | 210,659 |
| Unearn. revenues | 17,505 | 16,818 |
| Depreciation res.. | 2,683,001 | 2,253,919 |
| Earned surplus .. | 1,000,762 | 871,231 |
| Total ........ | $10,005,689 | $9,367,503 |
| Current assets ... | $920,796 | $715,246 |
| Curr. & accr. liab. | 831,969 | 851,533 |
| Net current assets | 88,827 | d 133,287 |

[c]Includes intangibles: 1949, $196,161; 1948, $205,169.

Accounts certified, Arthur Andersen & Co.

**Bonded Debt: 1. Porto Rico Telephone Co. first 4½s, series A; due serially to 1953:**

AUTHORIZED—Series A, $1,500,000 outstanding, series A, Mar. 2, 1950, $800,000; retired, $700,000. Privately held.

MATURITY—Due $100,000 annually to Mar. 1, 1958, incl.

INTEREST—M&S 1 at company's office, 67 Broad St., New York. Principal and interest payable in U. S. legal tender.

TRUSTEE—J. P. Morgan & Co., Inc., New York. Charles R. Hartzell, San Juan, Porto Rico, Individual Trustee.

DENOMINATION—Coupon, $1,000; not registerable as to principal; fully registered, multiples of $1,000. C&R interchangeable.

CALLABLE—As a whole or in part at any time on at least 30 days' notice at par plus 1/6% for each six months or part thereof to respective serial maturities.

SINKING FUND—None.

SECURITY—A first mortgage on all properties, franchises and telephone system in Puerto Rico. Mortgage covers after-acquired properties to a limited extent.

CREATION OF ADDITIONAL DEBT—Additional bonds may be issued in other series to extent of 60% of additions to plant property after net additions of $1,000,000 have been made to plant property as of Dec. 31, 1941.

PURPOSE—Issued to repay bank loans and other loans connected with conversion of San Juan, Puerto Rico manual central offices to automatic operation and for other plant additions.

OFFERED—($1,500,000) sold privately in Mar., 1943 to insurance companies in Canada and United States.

**Capital Stock: 1. Porto Rico Telephone Co. common; par $100:**

AUTHORIZED—43,000 shares (increased from 12,000 shares to 20,000 shares in 1930 and to 43,000 shares in 1949); outstanding, 18,000 shares; par $100.

OWNERSHIP—As of Dec. 31, 1949, International Tel. & Tel. Corp. owned 99.69% of outstanding shares.

VOTING RIGHTS—Has one vote per share.

DIVIDENDS PAID:

| | | | | | |
|---|---|---|---|---|---|
| 1915... | $2.00 | 1916-19 | $4.00 | 1920... | $8.00 |
| 1921-27 | 8.00 | 1928... | 6.00 | 1929-36 | Nil |
| 1937-43 | 6.00 | 1944... | 4.00 | 1945-49 | Nil |

TRANSFER AGENT—Stock transferred at company's office, New York.

**Subscription Rights**—Common stockholders of record Mar. 24, 1930, had right to subscribe for 6,000 additional common shares at $105 per share on basis of one new share for each two shares held.

**Interest and Dividend Tax Ruling:** The Commissioner of Internal Revenue has ruled that interest on bonds and dividends on stock of company are to be regarded, for tax purposes, as income from without the United States during the calendar year 1930. Such income, when received by non-resident aliens, is not subject to United States income tax

*Source:* http://webreports.mergent.com

sample is a long sequence of firms consisting of an apex corporation and its subsidiaries, each of which has control over the next one. In most cases, control chains include firms belonging to the same industrial category (e.g., railroads), but there are occasionally multiple control chains in different categories with the same ultimate owner as well (e.g., a few cases of public utility apex companies controlling industrial companies).

## C.4 Ultimate Controlling Shareholders

Moody's Manuals do not provide any information on the identity of the controllers of apex firms. To identify the owners of apex corporations that are not controlled by any other entity, we use the following sources:

1. For the 1926-1929 period: Pinchot (1928), the Wall Street Journal (WSJ) and the New York Times (NYT) archives, as well as additional sources, such as internet searches, historical documents, corporate files, www.archives.org and www.fundinguniverse.com.

2. For the 1929-1932 period: Table XII, Berle and Means (1932), Bonbright and Means (1932), Buchanan (1936), Lundberg (1937), the Encyclopedia of American Business History (2006), the WSJ and NYT archives and www.fundinguniverse.com.

3. For the 1937-1940 period: National Resources Committee (1939, Chapter IX and Appendix 13) and TNEC (1940).

## C.5 Matching Corporations to Patents, Publications and Labs

### C.5.1 Matching Corporations to Patents

Our patent data is sourced from the Google Patents dataset via Google BigQuery.[55] We cross-check the number of utility patents granted each year with the official USPTO statistics for our sample period in Figure C4 to ensure that our data source does not have coverage issues.[56] We find that the missing rate is around 3.43%; there are an average of 42,476 utility patents granted per year between 1926 and 1940.

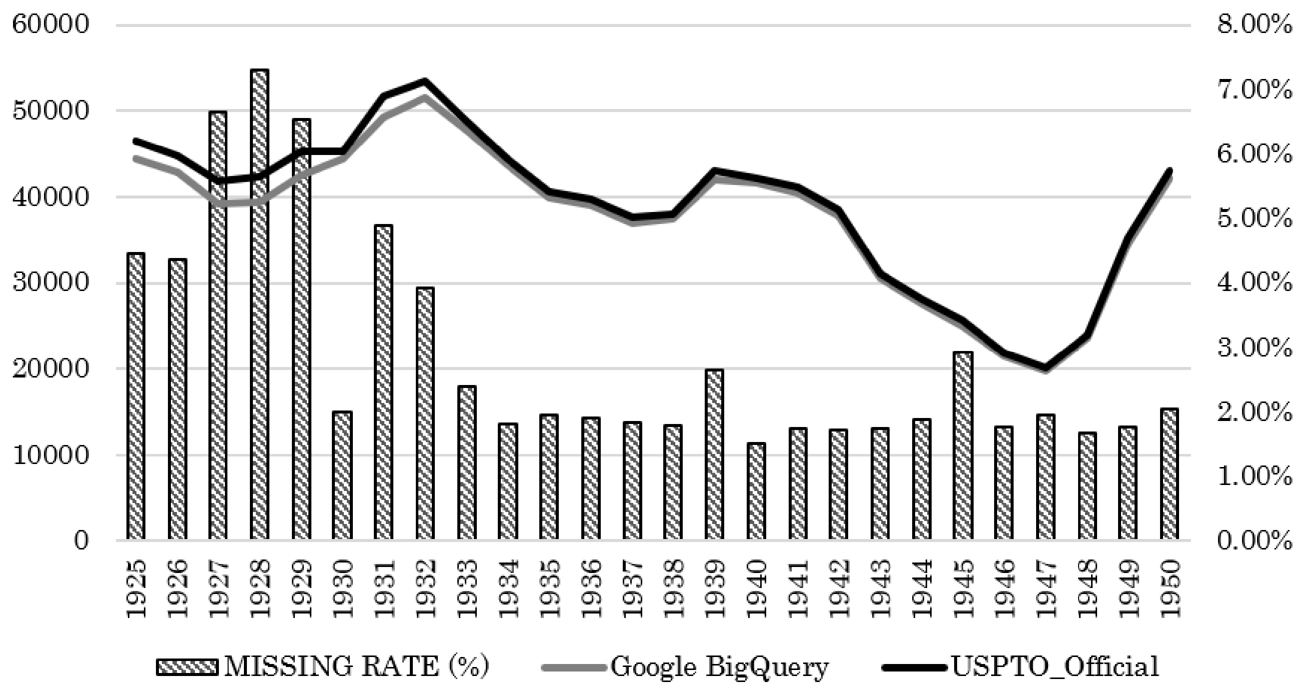
We extract the assignee field of the patents and standardize the names. We remove common prefixes and suffixes, such as 'The," "LLC," "INC," "A CORP OF". We also standardize names common in certain industries such as petroleum (sometimes abbreviated as "petr"), utilities ("power" abbreviated as "pwr"), rail ("railway," "railroad," "rail" used interchangeably and variously abbreviated as "RC," "RW," "RD," and "RC") as well as more common names, such as "manufacturing" ("MFG"), "National" ("Nat'l Steel Corp."), "American" ("Radio Corp of Amer") and state abbreviations. The last standardization is important for our sample period because companies then were more often named after the states they operated in (for instance, "Delaware Lackwanna  Western Coal Co." or the "Pennsylvania Electric Company"). Furthermore, we find alternative names specific certain firms such as the Standard Oil Company of Indiana (STANOLIND) and lab names for large companies such as AT&T's Bell Laboratories. Common abbreviations, such as RCA (Radio Corporation of America) and GE (General Electric), are also included. We then use a fuzzy string matching algorithm that calculates a length-adjusted Levenshtein distance. Using a fuzzy string matching algorithm is critical for patents from this period, as assignee names were not input electronically and are parsed through OCR.[57] Moreover, we manually check the

[55] Please see https://cloud.google.com/blog/topics/public-datasets/google-patents-public-datasets-connecting-public-paid-and-private-
for a brief overview of the dataset.

[56] USPTO official statistics for this period come from https://www.uspto.gov/web/offices/ac/ido/oeip/taf/h_counts.htm.

[57] As an example, the SOCONY Vacuum Oil Company is "misspelled" in the Google Patent data as: SCONY VACUUM OIL CO INC, SOCCNY VACUUM OIL CO INC, SOCENY VACUUM OIL CO IN, SOCONEY VACUUM OIL CO INC, SOCONY VACUNM OIL CO INC, SOCONY VAEUUM OIL CO INC, SOCONYVACUUM OIL CO INC, SOECNY VACUUM OIL CO INC, SOEONY VACUUM OIL CO INC, and SONCONY VACUUM OIL CO INC. The fuzzy string matching algorithm is still able to recover these matches.

Figure C2: Number of Published Utility Patents, 1925-1950




*Source:* The bar graph (right axis) plots the missing rate, defined as the difference in annual patent numbers between the USPTO official statistics and the Utility Patent (inventions) Column in the following source: https://www.uspto.gov/web/offices/ac/ido/oeip/taf/h_counts.htm.

names of 620 patentees with above 100 patents to include any matches that the string matching algorithm may still have missed.

We match 318 firms found in the Moody's directories to 64,523 patents. We also add 2,344 additional patents matched to 38 CRSP firms that were not matched in Kogan, Papanikolaou, Seru, and Stoffman (2017).[58]

[58] Kogan, Papanikolaou, Seru, and Stoffman (2017) match 60,493 patents to 368 CRPS firms between 1926 and 1940, which we also add to our sample.

### C.5.2 Matching Corporations to Publications

Table C1: American Corporate Publications (Top 20)

| Firm Name | Paper Count |
|---|---|
| General Electric Co. | 919 |
| American Telephone & Telegraph Co. | 562 |
| Westinghouse Elec. & Mfg. Co. | 321 |
| Radio Corp. of America | 195 |
| Eastman Kodak Co. | 132 |
| Humble Oil and Refining Company | 45 |
| Commonwealth Edison Co. | 44 |
| Swift & Co. | 42 |
| SHARP & DOHME INC | 36 |
| Procter & Gamble Co. | 34 |
| Western Union Tel. Co. | 34 |
| westinghouse lamp company | 32 |
| PARKE DAVIS & CO | 31 |
| Western Electric Company, Inc | 30 |
| Detroit Edison Co. | 29 |
| General Motors Corp. | 28 |
| National Carbon Co., Inc. | 25 |
| Texas Corp. | 25 |
| Aluminum Company of America | 24 |
| CORNING GLASS WORKS | 24 |

*Notes:* The table presents the number of scientific publications in MAG between 1925 and 1940 matched to our sample of American firms. The top 20 publishing firms are included.

Our publication data is sourced from Microsoft Academic Graph. We first download all author affiliations for papers published between 1926 and 1940. We run the same fuzzy string matching algorithm as above and manually check matches above a threshold score. Unlike patents, corporate publications are also often published under the name of the lab, which may not always correspond to the name of the firm. Therefore, we add names of prominent corporate laboratories such as Bell Labs and the Edgar C Bain Lab (for U.S. Steel) as name variants. To prevent false positive matches, we check that charitable organizations and university labs are not mismatched to the company. For instance, a 1934 publication by the "Eastman Laboratory of Physics" has high textual similarity to Eastman Kodak, but is actually part of the Massachusetts Institute of Technology. We also cross-tabulated the publication field of the company with its industry as a sanity check: we confirm, for instance, the wholesale and retail industry has scientific publications because the Boots Pure Drug Company (classified under this industry) published 29 articles ranging from the chemical sciences to clinical medicine.

Table C2: American Corporate Publications, by Scientific Field

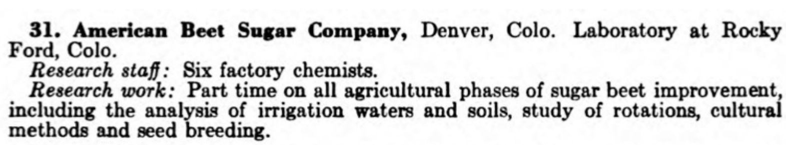
| OECD Subfield | Paper Count |
|---|---|
| 1.01 Mathematics | 77 |
| 1.02 Computer and information sciences | 77 |
| 1.03 Physical sciences and astronomy | 433 |
| 1.04 Chemical sciences | 267 |
| 1.05 Earth and related environmental sciences | 96 |
| 1.06 Biological sciences | 71 |
| 1.07 Other natural sciences | 2 |
| 2.01 Civil engineering | 73 |
| 2.02 Electrical eng, electronic eng | 1268 |
| 2.03 Mechanical engineering | 148 |
| 2.04 Chemical engineering | 22 |
| 2.05 Materials engineering | 152 |
| 2.06 Medical engineering | 21 |
| 2.07 Environmental engineering | 169 |
| 2.08 Environmental biotechnology | 6 |
| 2.11 Other engineering and technologies | 97 |
| 3.01 Basic medical research | 23 |
| 3.02 Clinical medicine | 112 |
| 3.03 Health sciences | 19 |
| 4.01 Agriculture, forestry, fisheries | 8 |
| 4.02 Animal and dairy science | 10 |
| 4.03 Veterinary science | 3 |
| 4.05 Other agricultural science | 10 |
| N/A | 6386 |

*Notes:* The table tabulates the scientific fields of MAG publications from 1925 to 1940 matched to our sample of American firms.

### C.5.3 Matching Corporations to Industrial Research Laboratories

We download the PDF files for the 1927, 1931, 1933 and 1938 editions of the NRC's Industrial Research Laboratory directory from Hathitrust. Since lab entries in the directory are of varying length (e.g., a stub for a leather company vs 2 pages for DuPont) and the fields are not sorted into metadata, the use of automated string matching algorithms is inefficient. However, since the entries are listed alphabetically, the directories are still amenable to manual matching. We enlisted two research assistants that manually searched through the directory to gather the name of the lab and the number of personnel employed at them. Though the directory also lists the type of personnel employed (e.g., chemists, physicists, etc.), these are not standardized by training or salary level, making it difficult to compare across firms. Therefore, we only use the total number of personnel as the indicator of investment in science for the analysis.

Figure C3: 1933 IRL Entry for American Beet & Sugar Company



**31. American Beet Sugar Company,** Denver, Colo. Laboratory at Rocky Ford, Colo.
*Research staff:* Six factory chemists.
*Research work:* Part time on all agricultural phases of sugar beet improvement, including the analysis of irrigation waters and soils, study of rotations, cultural methods and seed breeding.

Figure C4: 1933 IRL Entry for AT&T Bell Labs

**170. Bell Telephone Laboratories, Inc.,** 463 West Street, New York, N. Y. This company, a unit in the Bell Telephone System, engages in fundamental research in accordance with the research program of the American Telephone and Telegraph Company and carries out developments, designs and engineering services for the Western Electric Company, which latter company is the manufacturing unit of the Bell System.

*Company officers and department heads:* F. B. Jewett, President; P. Norton, Assistant to President; H. P. Charlesworth, Vice President. *Heads of functional activities:* O. E. Buckley, Director of Research; A. F. Dixon, Director of Systems Development; R. L. Jones, Director of Apparatus Development; J. G. Roberts, General Patent Attorney. *General staff:* S. P. Grace, Assistant Vice President; J. E. Moravec, Assistant Vice President; G. B. Thomas, Personnel Director; John Mills, Director of Publication.

In its functional organization the Laboratories divide into two main groups, the first of which is the technical staff including approximately 2000 research physicists, chemists, engineers, and other technicians, and the second, a somewhat smaller personnel concerned with the commercial operation of the Company and the rendering of service to the technical staff. In the second group fall such activities as the maintenance of the buildings, the operation of a well-equipped model shop, the purchase of equipment, accounting, library service, transcription, photographing, blue printing and personnel activities of education, employment and medical service.

The Laboratories carries on its technical work at the address above, and at several other locations, the most important of which are: 180 Varich Street and 480 Canal Street, New York, N. Y.; Holmdel, Deal, Summit, Whippany and Chester, N. J.

*Research work:* Researches in electronic physics, chemistry, magnetism, optics, radio and applied mathematics; in speech, hearing, conversion of energy between acoustic and electrical systems, the generation and modulation of electrical currents and instruments for the transmission of intelligence.

Development and design of apparatus for electrical communication, both wire and radio; studies of apparatus with a view to cost reduction either in manufacture, maintenance and repair, or through improved service; investigation of materials, maintenance of standards and methods of measurement, preparation of specifications for the manufacturer.

Development and design of communication systems combining economically for efficient operation communication apparatus and circuits, power equipment and other apparatus and circuits essential to the control, switching and supervision of communication circuits; continuing studies of current design; preparation of information necessary for manufacturer and installer.

Development and design of apparatus and investigation of materials for outside telephone plant; specification for manufacture or purchase.

Development of statistical methods of inspection and their adaptation for use by installer and manufacturer; development and application of standards of quality for communication apparatus and systems; study of inspection results; continuing study of service performance of the Laboratories' designs.

# Appendix D Details on Scientific Gap Calculations

## D.1 American Men of Science Directory

The AMS directory lists information on each scientist in a consistent manner: the last name is followed by the title, first name, current employment and residence and main discipline. Information on date and place of birth, alma mater, past employment and membership in professional societies follow. The final item in each entry is a detailed list of keywords that describe the focal scientist's research interests. We wish to extract i) the main discipline in which each scientist works and ii) any European degrees conferred.

Figure D1: American Men of Science Entry for Gilbert Lewis (1921)

**Lewis, Dr. G(ilbert) N(ewton),** University of California, Berkeley, Calif. **Chemistry.* Weymouth, Mass, Oct. 23, 75. Nebraska, 90–93; A.B, Harvard, 96, A.M, 98, Ph.D, 99; Leipzig and Göttingen, 00–01. Teacher, Phillips Acad, 96–97; instr. chem, Harvard, 99–00, 01–06, on leave in charge weights and measures, Bur. Govt. Laboratories, P. I, 04–05; asst. prof. physico-chem. research, Mass. Inst. Tech, 07–08, assoc. prof, 08–11, prof, 11–12, acting director, research lab, 07–09; *prof. chem. and dean col. chem, California, 12–* Major, lieut. col, chief of defense div, gas service, A.E.F, and chief of training div, C.W.S. Chevalier Légion d'honneur. Nat. Acad; Physical Soc; Chem. Soc; Philos. Soc; Am. Acad. Thermodynamic theory and its application to chemistry; free energy tables; equilibrium in numerous reactions; electric potentials of the common elements; properties of solutions and the activity of ions; distribution of thermal energy; specific heat of electrons; the principle of relativity and non-Newtonian mechanics; application of four-dimensional vector analysis to electro-magnetic theory; the geometry of the space time manifold of relativity; ultimate rational units; calculation of Stefan's constant; the structure of the atom and the molecule and the theory of valence; entropy of elements; third law of thermodynamics.

*Source:* Entry on Gilbert Lewis from the 1921 edition of the American Men of Science Directory.

The general data challenge is that the OCR on the image files, while relatively high quality, still has high error rates when classifying punctuation marks (commas, periods and semicolons) that are essential for separating out the entries into their constituent parts. Therefore, rather than splitting the text into its constituents, we directly search for the information we need. For main disciplines, we collect 131 scientific fields from a list of deceased scientists listed at the end of the 1906 and 1921 editions of AMS.[59] We conduct regular expressions (regex) on each AMS entry to determine which disciplines correspond to each scientist.[60] We further clean this data by determining the location of the regex match: if the matched discipline occurs after the birth date (Oct. 23, 75, for Gilbert Lewis in Figure D1), we remove the match. This prevents descriptions for research interests that occur later ("Thermodynamic theory and its application to chemistry; ... ; entropy of elements; third law of thermodynamics") from matching as the main discipline in which the scientist works. In the case of Gilbert Lewis, we prevent terms such as

[59]These are more feasible to collect manually, as the entries are structured as names, discipline, years of birth and death.

[60]"Technology," "General Science," and "Engineering" without specifying a field (mechanical, civil, mechanical, chemical) are excluded, as they are too general.

"electro-magnetic theory" or "non-Newtonian mechanics" to match with stemmed tokens for "Electrical Engineering" and "Mechanical Engineering." Afterwards, we manually map disciplines found in AMS into their equivalents in OECD subfields.

For alma maters and professional experience, we collect the list of all universities that were active in Europe between 1801 and 1945 from Wikipedia,[61] which in turn is heavily based on Rüegg (2004). Similar to main disciplines, we use regular expressions to determine whether each entry contains a match to at least one of these universities.[62] We further clean this data by removing matches for migrants that were born in Europe but trained exclusively in America: any match that occurs before the birth date of the scientist is excluded. Even after this cleaning, there will remain cases where an American is trained (until his doctoral degree) in the United States, only to be recognized by foreign institutions. We therefore complement this with scientific publication output data from Clarivate Web of Science.

## D.2 Web of Science Affiliations Coding

For the period between 1900 to 1920, the Microsoft Academic Graph data do not record the country of publication. Also, we find that the affiliations sections rarely list the full address of the author for this period, which leads MAG to omit country data from affiliation data. We therefore rely on Clarivate Web of Science, which has previously been used for research on the impacts of World War I on scientific production (Iaria, Schwarz, & Waldinger, 2018). Of 307,847 publications listed in Web of Science, 15% (44,356) have country data. We code each country as American, European and Rest of the World per Table D1 and D2. For the remaining 85% of publications without country information, we match the names of the authors to the 1906 and 1921 versions of the Cattell directory and classify those authors found in the directory as American (and the rest as European).

[61] https://en.wikipedia.org/wiki/List_of_modern_universities_in_Europe_(1801\OT1\textendash1945). We also collect data on early modern universities (established between 1501-1800) (https://en.wikipedia.org/wiki/List_of_early_modern_universities_in_Europe) and medieval universities (established before 1500) (https://en.wikipedia.org/wiki/List_of_medieval_universities) that were likely active in the early twentieth century.

[62] Schools with very short names, such as the University of Pau (France, 1722) and Literary University of Vic (Spain, 1599) are excluded because of high false positive match rates.

Table D1: WoS Countries and Regions (1/1)

| country | region | country | region | country | region | country | region |
|---|---|---|---|---|---|---|---|
| Africa | ROW | London | EUR | Switzerland | EUR | Uganda | ROW |
| Argentina | ROW | Malta | EUR | Syria | ROW | Rwanda | ROW |
| Australia | ROW | Mexico | ROW | Thailand | ROW | Ruanda Urundi | ROW |
| Austria | EUR | Mozambique | ROW | The Netherlands | EUR | Nigeria | ROW |
| Bahamas | ROW | N WALES | EUR | Turkey | ROW | Manchuria | ROW |
| Barbados | ROW | N Z | EUR | UK | EUR | Esthonie | EUR |
| Belgium | EUR | Netherlands | EUR | Ukraine | EUR | Ecudor | ROW |
| Belize | ROW | New Zealand | ROW | Uruguay | ROW | BURMA | ROW |
| Bermuda | ROW | Nicaragua | ROW | USA | USA | West Africa | ROW |
| Brazil | ROW | North Ireland | EUR | USSR | EUR | Ukriane | EUR |
| British | EUR | North Wales | EUR | Venezuela | ROW | Southern India | ROW |
| British East Af | EUR | Norway | EUR | W Indies | ROW | Palestine | ROW |
| British Hondurs | EUR | NS Wales | EUR | Wales | EUR | NY | USA |
| Bulgaria | EUR | Nyasaland | ROW | Western Austral | ROW | Kenya Colony | ROW |
| BWI | EUR | NZ | ROW | WIA | ROW | ISA | ? |
| Canada | ROW | P I | ROW | Yemen | ROW | Iraq | ROW |
| CEYLON | ROW | Panama | ROW | Yugoslavia | EUR | Great Britain | EUR |
| Chile | ROW | Peoples R China | ROW | SUISSE | EUR | East Africa | ROW |
| CHINA | ROW | Persia | ROW | Finnland | EUR | Yugoslavie | EUR |
| Colombia | ROW | Peru | ROW | BW1 | EUR | Western Samoa | ROW |
| Costa Rica | ROW | Philippine Isl | ROW | West Indies | ROW | Saskatchewan | ROW |
| Croatia | EUR | Philippine Isla | ROW | Russland | EUR | Russian Turkest | EUR |
| Cuba | ROW | Philippines | ROW | Prague | EUR | No Ireland | EUR |
| Czech Republic | EUR | Philippines Isl | ROW | Pakistan | ROW | Jugoslavia | EUR |
| CZECHOSLOVAKIA | EUR | Phillipine Isla | ROW | Malaysia | ROW | Johannesburg | ROW |
| Denmark | EUR | PI | ROW | Argentine | ROW | Inida | ROW |
| Egypt | ROW | Poland | EUR | Taiwan | ROW | Indien | ROW |
| England | EUR | Portugal | EUR | Kenya | ROW | Estonie | EUR |
| Federated Malay | ROW | Prussia | EUR | Bengal | ROW | Cook Islands | EUR |
| Fiji | ROW | Romania | EUR | Fed Malay State | ROW | BRITISH W INDIES | EUR |
| Finland | EUR | Russia | EUR | South America | ROW | Sri Lanka | ROW |
| FMS | ? | S AFRICA | ROW | Philippline Isl | ROW | Siberia | EUR |
| France | EUR | S Australia | ROW | Morocco | ROW | Lithuania | EUR |
| Germany | EUR | S India | ROW | Korea | ROW | Isle Wright | EUR |
| Greece | EUR | S Wales | EUR | Isle Of Man | EUR | Byelarus | EUR |
| Guatemala | ROW | Schweden | EUR | Engalnd | EUR | British West In | EUR |
| Guyana | ROW | Scotland | EUR | Ecuador | ROW | Philippine | ROW |
| HOLLAND | EUR | Senegal | ROW | Czechoslovakio | EUR | Belgian Congo | EUR |
| Honduras | ROW | Siam | ROW | Czechoslovak Re | EUR | Turkestan | ROW |
| Hong Kong | ROW | Sierra Leone | ROW | Columbia | ROW | Tunisia | ROW |
| Hungary | EUR | Singapore | ROW | Trinidad | ROW | Paris | EUR |
| India | ROW | South Africa | ROW | Tasmania | ROW | Maroc | ROW |
| Ireland | EUR | South Australia | ROW | Mauritius | ROW | Hongrie | EUR |
| Italien | EUR | South India | ROW | Estonia | EUR | Chili | ROW |
| Italy | EUR | South Korea | ROW | Esthonia | EUR | Tchecoslovaquie | EUR |
| Jamaica | ROW | Spain | EUR | Dutch E Indies | EUR | Haiti | ROW |
| Japan | ROW | Sudan | ROW | Dominican Repub | ROW | Berlin | EUR |
| Latvia | EUR | Sweden | EUR | Union Of South | ROW | Belguim | EUR |

*Notes:* The table lists the country affiliations of publications found in Clarivate Web of Science's Science Citation Index-Expanded between 1900 and 1920. "Region" has been imputed by the authors.

Table D2: WOS Countries and Regions (2/2)

| country | region | country | region | country | region |
|---|---|---|---|---|---|
| Union S Africa | ROW | Cihina | ROW | Paraguay | ROW |
| Ukraina | EUR | Britsh India | EUR | New Mexico | USA |
| Serbia | EUR | British India | EUR | Lebanaon | ROW |
| ROUMANIA | EUR | USRR | EUR | Istanbul | ROW |
| Polen | EUR | URSS | EUR | Isreal | ROW |
| Haut Congo Belg | EUR | UA | EUR | Estland | EUR |
| Gr Britain | EUR | Ruanda | ROW | E Indies | ROW |
| Chilli | EUR | R De P | ROW | E Africa | ROW |
| Cananda | ROW | N Ireland | EUR | Breslau | EUR |
| UdSSR | EUR | Dutch East Indi | EUR | Azerbaidjan | ROW |
| Schweiz | EUR | Czecho Solvakia | EUR | Venezuella | ROW |
| New Zeland | ROW | Cyprus | EUR | UL | ? |
| Lebanon | ROW | Chechoslovakia | EUR | Slovenia | EUR |
| Israel | ROW | Cairo | ROW | Republic Chili | ROW |
| Iceland | EUR | Ukrainia | EUR | Phillipine Isl | ROW |
| Hawaii | USA | Scothland | EUR | Nothern Ireland | EUR |
| Czechoslvakia | EUR | Puerto Rico | ROW | Netherlands Ind | EUR |
| Abyssinia | ROW | Oslo | EUR | Lebanan | ROW |
| W Africa | ROW | Irlande | EUR | Georgian SSR | EUR |
| Sud Mandschurei | ROW | Guadeloupe | EUR | Denamrk | EUR |
| Porto Rico | ROW | Ethiopie | ROW | CSR | EUR |
| North Africa | ROW | C I | ? | Britain | EUR |
| Netherland | EUR | BRASIL | ROW | Anglo Egyptian | EUR |
| Luxembourg | EUR | BELGIQUE | EUR | Yugoslavija | EUR |
| Irish Free Stat | EUR | Begium | EUR | Union South Afr | ROW |
| Iran | ROW | Bangladesh | ROW | UKx | EUR |
| Dominican Rep | ROW | W Germany | EUR | Sumatra | ROW |
| Central India | ROW | USRS | EUR | Slovakia | EUR |
| BWA | EUR | TH | ? | RHODESIA | ROW |
| USAa | USA | S Africia | ROW | Northern Ireland | EUR |
| Ungarn | ROW | Republic Panama | ROW | | |
| RUMANIA | EUR | Madras | ROW | | |
| R Argentina | ROW | LURSS | EUR | | |
| N Nigeria | ROW | Lettonia | EUR | | |
| Mailand | ROW | Jugoslawien | EUR | | |
| Jerusalem | ROW | Cameroon | ROW | | |
| Czlchoslovakia | EUR | ARSSR | EUR | | |
| Czechoslavakia | EUR | Zwitzerland | EUR | | |
| United Kingdom | EUR | USS | USA | | |
| Republ Libanaise | ROW | Rumanien | EUR | | |
| Norwegen | EUR | N Rhodesia | ROW | | |
| Northern Irelan | EUR | Czechoslovak | EUR | | |
| Java | ROW | Cent India | ROW | | |
| Hungry | EUR | Bolivia | ROW | | |
| GSSR | EUR | Belgien | EUR | | |
| Ethiopia | ROW | Armenia | EUR | | |
| Egpyt | ROW | Uzbekistan | EUR | | |
| Czecho Slovakia | EUR | Rep of Georgia | EUR | | |

*Notes:* The table lists the country affiliations of publications found in Clarivate Web of Science's Science Citation Index-Expanded between 1900 and 1920. "Region" has been imputed by the authors.

## D.3 Journal Country Coding and Citation Flows

We first classify journals with non-English and non-Latin names (e.g., Zeitshcrift für Physik) as European. We also classify journals with the name "American" in it as American (e.g., the American Heart Journal). We then manually classify the remaining journals by web searches. Where a full history of the journal is available, we classify the journal's home country as the place where its publisher/publishing academic society is. For instance, "Bacteriological Reviews" is a journal that was published by the American Society of Microbiology.[63] When publisher information is not available, we use the nationality of the founding members to classify the journal. Out of the 293 journals published between 1925 and 1940, we are able to classify 272 as American or European.

[63]https://en.wikipedia.org/wiki/Microbiology_and_Molecular_Biology_Reviews

Table D3: Country Coding of Journals (1/6)

| | Journal Name | Country |
|---|---|---|
| 1 | AMERICAN HEART JOURNAL | USA |
| 2 | AMERICAN JOURNAL OF ANATOMY | USA |
| 3 | AMERICAN JOURNAL OF BOTANY | USA |
| 4 | AMERICAN JOURNAL OF DISEASES OF CHILDREN | USA |
| 5 | AMERICAN JOURNAL OF HYGIENE | USA |
| 6 | AMERICAN JOURNAL OF HYGIENE-MONOGRAPHIC SERIES | USA |
| 7 | AMERICAN JOURNAL OF INSANITY | USA |
| 8 | AMERICAN JOURNAL OF MATHEMATICS | USA |
| 9 | AMERICAN JOURNAL OF NURSING | USA |
| 10 | AMERICAN JOURNAL OF OBSTETRICS AND GYNECOLOGY | USA |
| 11 | AMERICAN JOURNAL OF PATHOLOGY | USA |
| 12 | AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY | USA |
| 13 | AMERICAN JOURNAL OF PHYSIOLOGY | USA |
| 14 | AMERICAN JOURNAL OF PSYCHIATRY | USA |
| 15 | AMERICAN JOURNAL OF PSYCHOLOGY | USA |
| 16 | AMERICAN JOURNAL OF PUBLIC HEALTH | USA |
| 17 | AMERICAN JOURNAL OF PUBLIC HEALTH AND THE NATIONS HEALTH | USA |
| 18 | AMERICAN JOURNAL OF ROENTGENOLOGY | USA |
| 19 | AMERICAN JOURNAL OF ROENTGENOLOGY AND RADIUM THERAPY | USA |
| 20 | AMERICAN JOURNAL OF SCIENCE | USA |
| 21 | AMERICAN JOURNAL OF THE MEDICAL SCIENCES | USA |
| 22 | AMERICAN MINERALOGIST | USA |
| 23 | AMERICAN NATURALIST | USA |
| 24 | ANATOMICAL RECORD | USA |
| 25 | ANNALS OF INTERNAL MEDICINE | USA |
| 26 | ANNALS OF MATHEMATICS | USA |
| 27 | ANNALS OF SURGERY | USA |
| 28 | ARCHIVES OF INTERNAL MEDICINE | USA |
| 29 | ARCHIVES OF NEUROLOGY AND PSYCHIATRY | USA |
| 30 | ARCHIVES OF OPHTHALMOLOGY | USA |
| 31 | ARCHIVES OF OTOLARYNGOLOGY | USA |
| 32 | ARCHIVES OF PATHOLOGY | USA |
| 33 | ARCHIVES OF PATHOLOGY & LABORATORY MEDICINE | USA |
| 34 | ARCHIVES OF SURGERY | USA |
| 35 | ASTROPHYSICAL JOURNAL | USA |
| 36 | BELL SYSTEM TECHNICAL JOURNAL | USA |
| 37 | BIOLOGICAL BULLETIN | USA |
| 38 | BOSTON MEDICAL AND SURGICAL JOURNAL | USA |
| 39 | BOTANICAL GAZETTE | USA |
| 40 | BULLETIN OF THE AMERICAN MUSEUM OF NATURAL HISTORY | USA |
| 41 | BULLETIN OF THE GEOLOGICAL SOCIETY OF AMERICA | USA |
| 42 | BULLETIN OF THE JOHNS HOPKINS HOSPITAL | USA |
| 43 | BULLETIN OF THE TORREY BOTANICAL CLUB | USA |
| 44 | BUREAU OF STANDARDS JOURNAL OF RESEARCH | USA |
| 45 | CHEMICAL REVIEWS | USA |
| 46 | ECOLOGY | USA |
| 47 | ENDOCRINOLOGY | USA |
| 48 | GENETICS | USA |
| 49 | INDUSTRIAL AND ENGINEERING CHEMISTRY | USA |
| 50 | JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY | USA |
| 51 | JOURNAL OF BACTERIOLOGY | USA |
| 52 | JOURNAL OF BIOLOGICAL CHEMISTRY | USA |
| 53 | JOURNAL OF BONE AND JOINT SURGERY | USA |
| 54 | JOURNAL OF CLINICAL ENDOCRINOLOGY | USA |
| 55 | JOURNAL OF CLINICAL INVESTIGATION | USA |
| 56 | JOURNAL OF COMPARATIVE NEUROLOGY | USA |
| 57 | JOURNAL OF COMPARATIVE NEUROLOGY AND PSYCHOLOGY | USA |
| 58 | JOURNAL OF COMPARATIVE PSYCHOLOGY | USA |
| 59 | JOURNAL OF ECONOMIC ENTOMOLOGY | USA |

Table D4: Country Coding of Journals (2/6)

| | Journal Name | Country |
|---|---|---|
| 60 | JOURNAL OF EDUCATIONAL PSYCHOLOGY | USA |
| 61 | JOURNAL OF EXPERIMENTAL MEDICINE | USA |
| 62 | JOURNAL OF EXPERIMENTAL PSYCHOLOGY | USA |
| 63 | JOURNAL OF FARM ECONOMICS | USA |
| 64 | JOURNAL OF GENERAL PHYSIOLOGY | USA |
| 65 | JOURNAL OF GEOLOGY | USA |
| 66 | JOURNAL OF HEREDITY | USA |
| 67 | JOURNAL OF IMMUNOLOGY | USA |
| 68 | JOURNAL OF INDUSTRIAL AND ENGINEERING CHEMISTRY-US | USA |
| 69 | JOURNAL OF INFECTIOUS DISEASES | USA |
| 70 | JOURNAL OF LABORATORY AND CLINICAL MEDICINE | USA |
| 71 | JOURNAL OF MEDICAL RESEARCH | USA |
| 72 | JOURNAL OF MORPHOLOGY | USA |
| 73 | JOURNAL OF MORPHOLOGY AND PHYSIOLOGY | USA |
| 74 | JOURNAL OF NERVOUS AND MENTAL DISEASE | USA |
| 75 | JOURNAL OF NEUROLOGY AND PSYCHOPATHOLOGY | USA |
| 76 | JOURNAL OF NUTRITION | USA |
| 77 | JOURNAL OF PHARMACOLOGY AND EXPERIMENTAL THERAPEUTICS | USA |
| 78 | JOURNAL OF PHYSICAL CHEMISTRY | USA |
| 79 | JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA | USA |
| 80 | JOURNAL OF THE AMERICAN CERAMIC SOCIETY | USA |
| 81 | JOURNAL OF THE AMERICAN CHEMICAL SOCIETY | USA |
| 82 | JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION | USA |
| 83 | JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION | USA |
| 84 | JOURNAL OF THE FRANKLIN INSTITUTE | USA |
| 85 | JOURNAL OF THE OPTICAL SOCIETY OF AMERICA | USA |
| 86 | JOURNAL OF THE OPTICAL SOCIETY OF AMERICA AND REVIEW OF SCIENTIFIC INSTRUMENTS | USA |
| 87 | JOURNAL OF UROLOGY | USA |
| 88 | NEW ENGLAND JOURNAL OF MEDICINE | USA |
| 89 | ORGANIC SYNTHESES | USA |
| 90 | PEDAGOGICAL SEMINARY | USA |
| 91 | PEDAGOGICAL SEMINARY AND JOURNAL OF GENETIC PSYCHOLOGY | USA |
| 92 | PHYSICAL REVIEW | USA |
| 93 | PHYSIOLOGICAL REVIEWS | USA |
| 94 | PHYTOPATHOLOGY | USA |
| 95 | PLANT PHYSIOLOGY | USA |
| 96 | PROCEEDINGS OF THE AMERICAN ACADEMY OF ARTS AND SCIENCES | USA |
| 97 | PROCEEDINGS OF THE INSTITUTE OF RADIO ENGINEERS | USA |
| 98 | PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA | USA |
| 99 | PSYCHOLOGICAL BULLETIN | USA |
| 100 | PSYCHOLOGICAL REVIEW | USA |
| 101 | PUBLIC HEALTH REPORTS | USA |
| 102 | PUBLICATIONS OF THE AMERICAN STATISTICAL ASSOCIATION | USA |
| 103 | QUARTERLY PUBLICATIONS OF THE AMERICAN STATISTICAL ASSOCIATION | USA |
| 104 | QUARTERLY REVIEW OF BIOLOGY | USA |
| 105 | REVIEWS OF MODERN PHYSICS | USA |
| 106 | SCIENCE | USA |
| 107 | STAIN TECHNOLOGY | USA |
| 108 | TRANSACTIONS OF THE AMERICAN INSTITUTE OF CHEMICAL ENGINEERS | USA |
| 109 | TRANSACTIONS OF THE AMERICAN INSTITUTE OF MINING AND METALLURGICAL ENGINEERS | USA |
| 110 | TRANSACTIONS OF THE AMERICAN MATHEMATICAL SOCIETY | USA |
| 111 | TRANSACTIONS-AMERICAN GEOPHYSICAL UNION | USA |
| 112 | AMERICAN JOURNAL OF MENTAL DEFICIENCY | USA |
| 113 | AMERICAN JOURNAL OF ORTHOPSYCHIATRY | USA |
| 114 | ANNALS OF MATHEMATICAL STATISTICS | USA |
| 115 | BACTERIOLOGICAL REVIEWS | USA |
| 116 | DISEASES OF THE NERVOUS SYSTEM | USA |
| 117 | ECONOMETRICA | USA |
| 118 | INDUSTRIAL AND ENGINEERING CHEMISTRY-ANALYTICAL EDITION | USA |
| 119 | JOURNAL OF APPLIED PHYSICS | USA |

Table D5: Country Coding of Journals (3/6)

| | Journal Name | Country |
|---|---|---|
| 120 | JOURNAL OF CHEMICAL PHYSICS | USA |
| 121 | JOURNAL OF CONSULTING PSYCHOLOGY | USA |
| 122 | JOURNAL OF MARINE RESEARCH | USA |
| 123 | JOURNAL OF NEUROPHYSIOLOGY | USA |
| 124 | JOURNAL OF ORGANIC CHEMISTRY | USA |
| 125 | JOURNAL OF PEDIATRICS | USA |
| 126 | JOURNAL OF RESEARCH OF THE NATIONAL BUREAU OF STANDARDS | USA |
| 127 | JOURNAL OF THE NATIONAL CANCER INSTITUTE | USA |
| 128 | PSYCHOMETRIKA | USA |
| 129 | PSYCHOSOMATIC MEDICINE | USA |
| 130 | QUARTERLY JOURNAL OF STUDIES ON ALCOHOL | USA |
| 131 | REVIEW OF SCIENTIFIC INSTRUMENTS | USA |
| 132 | SOCIOMETRY | USA |
| 133 | ANNALS OF EUGENICS | UNKNOWN |
| 134 | ARCHIVES OF DERMATOLOGY AND SYPHILOLOGY | UNKNOWN |
| 135 | CONTRIBUTIONS TO EMBRYOLOGY | UNKNOWN |
| 136 | GENETIC PSYCHOLOGY MONOGRAPHS | UNKNOWN |
| 137 | HEART-A JOURNAL FOR THE STUDY OF THE CIRCULATION | UNKNOWN |
| 138 | JOURNAL OF AGRICULTURAL RESEARCH | UNKNOWN |
| 139 | JOURNAL OF EXPERIMENTAL ZOOLOGY | UNKNOWN |
| 140 | JOURNAL OF HYGIENE | UNKNOWN |
| 141 | JOURNAL OF THE INSTITUTE OF METALS | UNKNOWN |
| 142 | MEDICINE | UNKNOWN |
| 143 | MENTAL HYGIENE | UNKNOWN |
| 144 | PROCEEDINGS OF THE SOCIETY FOR EXPERIMENTAL BIOLOGY AND MEDICINE | UNKNOWN |
| 145 | PSYCHOLOGICAL MONOGRAPHS | UNKNOWN |
| 146 | SOIL SCIENCE | UNKNOWN |
| 147 | ACTA PHYSICOCHIMICA URSS | UNKNOWN |
| 148 | CHARACTER AND PERSONALITY | UNKNOWN |
| 149 | JOURNAL OF CELLULAR AND COMPARATIVE PHYSIOLOGY | UNKNOWN |
| 150 | JOURNAL OF OTOLARYNGOLOGY | UNKNOWN |
| 151 | PHYSICS-A JOURNAL OF GENERAL AND APPLIED PHYSICS | UNKNOWN |
| 152 | PSYCHIATRY | UNKNOWN |
| 153 | SURGERY | UNKNOWN |
| 154 | HELVETICA CHIMICA ACTA | SWITZERLAND |
| 155 | ACTA MEDICA SCANDINAVICA | SWEDEN |
| 156 | HEREDITAS | SWEDEN |
| 157 | ZOOLOGISKA BIDRAG FRAN UPPSALA | SWEDEN |
| 158 | ACTA MATHEMATICA | SWEDEN |
| 159 | COMPTES RENDUS DE L ACADEMIE DES SCIENCES DE L URSS | RUSSIA |
| 160 | JOURNAL OF PHYSICS-USSR | RUSSIA |
| 161 | ACTA RADIOLOGICA | NORWAY |
| 162 | PHYSICA | NETHERLANDS |
| 163 | PHYSICA B-CONDENSED MATTER | NETHERLANDS |
| 164 | PROCEEDINGS OF THE KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN | NETHERLANDS |
| 165 | PROCEEDINGS OF THE KONINKLIJKE AKADEMIE VAN WETENSCHAPPEN TE AMSTERDAM | NETHERLANDS |
| 166 | RECUEIL DES TRAVAUX CHIMIQUES DES PAYS-BAS | NETHERLANDS |
| 167 | ANNALS OF APPLIED BIOLOGY | GREAT BRITAIN |
| 168 | ANNALS OF BOTANY | GREAT BRITAIN |
| 169 | BIOCHEMICAL JOURNAL | GREAT BRITAIN |
| 170 | BIOLOGICAL REVIEWS AND BIOLOGICAL PROCEEDINGS OF THE CAMBRIDGE PHILOSOPHICAL SOCIETY | GREAT BRITAIN |
| 171 | BIOMETRIKA | GREAT BRITAIN |
| 172 | BMJ-BRITISH MEDICAL JOURNAL | GREAT BRITAIN |
| 173 | BRAIN | GREAT BRITAIN |
| 174 | BRITISH JOURNAL OF DERMATOLOGY | GREAT BRITAIN |
| 175 | BRITISH JOURNAL OF DERMATOLOGY AND SYPHILIS | GREAT BRITAIN |
| 176 | BRITISH JOURNAL OF EXPERIMENTAL BIOLOGY | GREAT BRITAIN |
| 177 | BRITISH JOURNAL OF EXPERIMENTAL PATHOLOGY | GREAT BRITAIN |
| 178 | BRITISH JOURNAL OF MEDICAL PSYCHOLOGY | GREAT BRITAIN |
| 179 | BRITISH JOURNAL OF SURGERY | GREAT BRITAIN |

Table D6: Country Coding of Journals (4/6)

| | Journal Name | Country |
|---|---|---|
| 180 | BRITISH MEDICAL JOURNAL | GREAT BRITAIN |
| 181 | JOURNAL OF ANATOMY | GREAT BRITAIN |
| 182 | JOURNAL OF ANATOMY AND PHYSIOLOGY | GREAT BRITAIN |
| 183 | JOURNAL OF ECOLOGY | GREAT BRITAIN |
| 184 | JOURNAL OF GENETICS | GREAT BRITAIN |
| 185 | JOURNAL OF MENTAL SCIENCE | GREAT BRITAIN |
| 186 | JOURNAL OF PATHOLOGY AND BACTERIOLOGY | GREAT BRITAIN |
| 187 | JOURNAL OF PHYSIOLOGY-LONDON | GREAT BRITAIN |
| 188 | JOURNAL OF THE CHEMICAL SOCIETY | GREAT BRITAIN |
| 189 | JOURNAL OF THE ROYAL STATISTICAL SOCIETY | GREAT BRITAIN |
| 190 | LANCET | GREAT BRITAIN |
| 191 | MEMOIRS OF THE ROYAL METROLOGICAL SOCIETY | GREAT BRITAIN |
| 192 | MONTHLY NOTICES OF THE ROYAL ASTRONOMICAL SOCIETY | GREAT BRITAIN |
| 193 | NATURE | GREAT BRITAIN |
| 194 | PHILOSOPHICAL MAGAZINE | GREAT BRITAIN |
| 195 | PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON SERIES A-CONTAINING PAPERS OF A MATHEMATICAL OR PHYSICAL CHARACTER | GREAT BRITAIN |
| 196 | PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON SERIES B-CONTAINING PAPERS OF A BIOLOGICAL CHARACTER | GREAT BRITAIN |
| 197 | PROCEEDINGS OF THE CAMBRIDGE PHILOSOPHICAL SOCIETY | GREAT BRITAIN |
| 198 | PROCEEDINGS OF THE CAMBRIDGE PHILOSOPHICAL SOCIETY-BIOLOGICAL SCIENCES | GREAT BRITAIN |
| 199 | PROCEEDINGS OF THE LONDON MATHEMATICAL SOCIETY | GREAT BRITAIN |
| 200 | PROCEEDINGS OF THE PHYSICAL SOCIETY | GREAT BRITAIN |
| 201 | PROCEEDINGS OF THE PHYSICAL SOCIETY OF LONDON | GREAT BRITAIN |
| 202 | PROCEEDINGS OF THE ROYAL SOCIETY OF LONDON | GREAT BRITAIN |
| 203 | PROCEEDINGS OF THE ROYAL SOCIETY OF LONDON SERIES A-CONTAINING PAPERS OF A MATHEMATICAL AND PHYSICAL CHARACTER | GREAT BRITAIN |
| 204 | PROCEEDINGS OF THE ROYAL SOCIETY OF LONDON SERIES B-CONTAINING PAPERS OF A BIOLOGICAL CHARACTER | GREAT BRITAIN |
| 205 | PROCEEDINGS OF THE ZOOLOGICAL SOCIETY OF LONDON | GREAT BRITAIN |
| 206 | QUARTERLY JOURNAL OF EXPERIMENTAL PHYSIOLOGY | GREAT BRITAIN |
| 207 | QUARTERLY JOURNAL OF MEDICINE | GREAT BRITAIN |
| 208 | QUARTERLY JOURNAL OF MICROSCOPICAL SCIENCE | GREAT BRITAIN |
| 209 | QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY | GREAT BRITAIN |
| 210 | SURGERY GYNECOLOGY & OBSTETRICS | GREAT BRITAIN |
| 211 | TRANSACTIONS OF THE FARADAY SOCIETY | GREAT BRITAIN |
| 212 | BIOLOGICAL REVIEWS OF THE CAMBRIDGE PHILOSOPHICAL SOCIETY | GREAT BRITAIN |
| 213 | BRITISH HEART JOURNAL | GREAT BRITAIN |
| 214 | CLINICAL SCIENCE | GREAT BRITAIN |
| 215 | JOURNAL OF ANIMAL ECOLOGY | GREAT BRITAIN |
| 216 | JOURNAL OF EXPERIMENTAL BIOLOGY | GREAT BRITAIN |
| 217 | JOURNAL OF NEUROLOGY AND PSYCHIATRY | GREAT BRITAIN |
| 218 | PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES | GREAT BRITAIN |
| 219 | PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON SERIES A-MATHEMATICAL AND PHYSICAL SCIENCES | GREAT BRITAIN |
| 220 | PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON SERIES B-BIOLOGICAL SCIENCES | GREAT BRITAIN |
| 221 | PROCEEDINGS OF THE ROYAL SOCIETY OF LONDON SERIES A-MATHEMATICAL AND PHYSICAL SCIENCES | GREAT BRITAIN |
| 222 | PROCEEDINGS OF THE ROYAL SOCIETY SERIES B-BIOLOGICAL SCIENCES | GREAT BRITAIN |
| 223 | PROCEEDINGS OF THE ZOOLOGICAL SOCIETY OF LONDON SERIES A-GENERAL AND EXPERIMENTAL | GREAT BRITAIN |
| 224 | PROCEEDINGS OF THE ZOOLOGICAL SOCIETY OF LONDON SERIES B-SYSTEMATIC AND MORPHOLOGICAL | GREAT BRITAIN |
| 225 | QUARTERLY JOURNAL OF EXPERIMENTAL PHYSIOLOGY AND COGNATE MEDICAL SCIENCES | GREAT BRITAIN |
| 226 | JOURNAL OF AGRICULTURAL SCIENCE | GREAT BRITAIN |
| 227 | BIOCHEMISCHE ZEITSCHRIFT | GERMANY |
| 228 | HOPPE-SEYLERS ZEITSCHRIFT FUR PHYSIOLOGISCHE CHEMIE | GERMANY |
| 229 | KOLLOID-ZEITSCHRIFT | GERMANY |

Table D7: Country Coding of Journals (5/6)

| | Journal Name | Country |
|---|---|---|
| 230 | MATHEMATISCHE ZEITSCHRIFT | GERMANY |
| 231 | PHYSIKALISCHE ZEITSCHRIFT | GERMANY |
| 232 | ZEITSCHRIFT DES VEREINES DEUTSCHER INGENIEURE | GERMANY |
| 233 | ZEITSCHRIFT FUR ANGEWANDTE MATHEMATIK UND MECHANIK | GERMANY |
| 234 | ZEITSCHRIFT FUR ANORGANISCHE CHEMIE | GERMANY |
| 235 | ZEITSCHRIFT FUR ANORGANISCHE UND ALLGEMEINE CHEMIE | GERMANY |
| 236 | ZEITSCHRIFT FUR BIOLOGIE | GERMANY |
| 237 | ZEITSCHRIFT FUR DIE GESAMTE NEUROLOGIE UND PSYCHIATRIE | GERMANY |
| 238 | ZEITSCHRIFT FUR ELEKTROCHEMIE | GERMANY |
| 239 | ZEITSCHRIFT FUR ELEKTROCHEMIE UND ANGEWANDTE PHYSIKALISCHE CHEMIE | GERMANY |
| 240 | ZEITSCHRIFT FUR KRISTALLOGRAPHIE | GERMANY |
| 241 | ZEITSCHRIFT FUR KRYSTALLOGRAPHIE UND MINERALOGIE | GERMANY |
| 242 | ZEITSCHRIFT FUR PHYSIK | GERMANY |
| 243 | ZEITSCHRIFT FUR PHYSIKALISCHE CHEMIE–STOCHIOMETRIE UND VERWANDTSCHAFTSLEHRE | GERMANY |
| 244 | ZEITSCHRIFT FUR PHYSIKALISCHE CHEMIE-ABTEILUNG A-CHEMISCHE THERMODYNAMIK KINETIK ELEKTROCHEMIE EIGENSCHAFTSLEHRE | GERMANY |
| 245 | ZEITSCHRIFT FUR PHYSIKALISCHE CHEMIE-ABTEILUNG B-CHEMIE DER ELEMENTARPROZESSE AUFBAU DER MATERIE | GERMANY |
| 246 | ZEITSCHRIFT FUR PHYSIKALISCHE CHEMIE-STOCHIOMETRIE UND VERWANDTSCHAFTSLEHRE | GERMANY |
| 247 | ZEITSCHRIFT FUR WISSENSCHAFTLICHE ZOOLOGIE | GERMANY |
| 248 | ANGEWANDTE CHEMIE | GERMANY |
| 249 | ANNALEN DER PHYSIK | GERMANY |
| 250 | ARCHIV FUR DERMATOLOGIE UND SYPHILIS | GERMANY |
| 251 | ARCHIV FUR DIE GESAMTE PHYSIOLOGIE DES MENSCHEN UND DER TIERE | GERMANY |
| 252 | ARCHIV FUR ENTWICKLUNGSMECHANIK DER ORGANISMEN | GERMANY |
| 253 | ARCHIV FUR EXPERIMENTELLE PATHOLOGIE UND PHARMAKOLOGIE | GERMANY |
| 254 | ARCHIV FUR EXPERIMENTELLE ZELLFORSCHUNG | GERMANY |
| 255 | ARCHIV FUR MIKROSKOPISCHE ANATOMIE | GERMANY |
| 256 | ARCHIV FUR MIKROSKOPISCHE ANATOMIE UND ENTWICKLUNGSGESCHICHTE | GERMANY |
| 257 | ARCHIV FUR MIKROSKOPISCHE ANATOMIE UND ENTWICKLUNGSMECHANIK | GERMANY |
| 258 | ARCHIV FUR PATHOLOGISCHE ANATOMIE UND PHYSIOLOGIE UND FUR KLINISCHE MEDICIN | GERMANY |
| 259 | ARCHIV FUR PSYCHIATRIE UND NERVENKRANKHEITEN | GERMANY |
| 260 | BEITRAGE ZUR PATHOLOGISCHEN ANATOMIE UND ZUR ALLGEMEINEN PATHOLOGIE | GERMANY |
| 261 | BERICHTE DER DEUTSCHEN CHEMISCHEN GESELLSCHAFT | GERMANY |
| 262 | DERMATOLOGISCHE WOCHENSCHRIFT | GERMANY |
| 263 | DEUTSCHE MEDIZINISCHE WOCHENSCHRIFT | GERMANY |
| 264 | FORTSCHRITTE DER NEUROLOGIE UND PSYCHIATRIE UND IHRER GRENZGEBIETE | GERMANY |
| 265 | JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK | GERMANY |
| 266 | JOURNAL FUR PRAKTISCHE CHEMIE-LEIPZIG | GERMANY |
| 267 | JOURNAL FUR PSYCHOLOGIE UND NEUROLOGIE | GERMANY |
| 268 | JUSTUS LIEBIGS ANNALEN DER CHEMIE | GERMANY |
| 269 | MATHEMATISCHE ANNALEN | GERMANY |
| 270 | NATURWISSENSCHAFTEN | GERMANY |
| 271 | NAUNYN-SCHMIEDEBERGS ARCHIV FUR EXPERIMENTELLE PATHOLOGIE UND PHARMAKOLOGIE | GERMANY |
| 272 | NERVENARZT | GERMANY |
| 273 | PFLUGERS ARCHIV FUR DIE GESAMTE PHYSIOLOGIE DES MENSCHEN UND DER TIERE | GERMANY |
| 274 | PSYCHOLOGISCHE FORSCHUNG | GERMANY |
| 275 | SCHWEIZER ARCHIV FUR NEUROLOGIE UND PSYCHIATRIE | GERMANY |

Table D8: Country Coding of Journals (6/6)

| | Journal Name | Country |
|---|---|---|
| 276 | SITZUNGSBERICHTE DER KONIGLICH PREUSSISCHEN AKADEMIE DER WISSENSCHAFTEN | GERMANY |
| 277 | SITZUNGSBERICHTE DER PREUSSICHEN AKADEMIE DER WISSENSCHAFTEN PHYSIKALISCH-MATHEMATISCHE KLASSE | GERMANY |
| 278 | SKANDINAVISCHES ARCHIV FUR PHYSIOLOGIE | GERMANY |
| 279 | VIRCHOWS ARCHIV FUR PATHOLOGISCHE ANATOMIE UND PHYSIOLOGIE UND FUR KLINISCHE MEDIZIN | GERMANY |
| 280 | WILHELM ROUX ARCHIV FUR ENTWICKLUNGSMECHANIK DER ORGANISMEN | GERMANY |
| 281 | ANNALES DE CHIMIE ET DE PHYSIQUE | FRANCE |
| 282 | ANNALES DE CHIMIE FRANCE | FRANCE |
| 283 | ANNALES MEDICO-PSYCHOLOGIQUES | FRANCE |
| 284 | ARCHIVES INTERNATIONALES DE PHARMACODYNAMIE ET DE THERAPIE | FRANCE |
| 285 | COMPTES RENDUS DES SEANCES DE LA SOCIETE DE BIOLOGIE ET DE SES FILIALES | FRANCE |
| 286 | COMPTES RENDUS HEBDOMADAIRES DES SEANCES DE L ACADEMIE DES SCIENCES | FRANCE |
| 287 | ENCEPHALE-REVUE DE PSYCHIATRIE CLINIQUE BIOLOGIQUE ET THERAPEUTIQUE | FRANCE |
| 288 | JOURNAL DE PHYSIQUE ET LE RADIUM | FRANCE |
| 289 | REVUE NEUROLOGIQUE | FRANCE |
| 290 | ACTA PSYCHIATRICA ET NEUROLOGICA | EUROPE |
| 291 | CANADIAN MEDICAL ASSOCIATION JOURNAL | CANADA |
| 292 | RECUEIL DES TRAVAUX CHIMIQUES DES PAYS-BAS ET DE LA BELGIQUE | BELGIUM |
| 293 | TRANSACTIONS OF THE OPHTHALMOLOGICAL SOCIETY OF AUSTRALIA | AUSTRALIA |

## D.4 Comparison Between Gap Measures

Table D9 compares the three measures of scientific "strength" (relative backwardness) from WoS and AMS. The "Ratio" columns for each measure present the number of European-affiliated scientists, European-authored papers and citations to European journals by American papers divided by the number of scientists never affiliated with a European institution, the number of American-authored papers and citations to American journals, respectively. Intuitively, these ratios can be thought of as the "gap" or "lag" that exists between European and American institutions (fields with relatively large values are those where the scientific gap between Europe and the U.S. is large). The three measures do not yield identical results. For instance, the AMS ratio is smaller in mean and variance compared to the ones based on Web of Science publications and citations. Given the lack of citations data in civil engineering and agriculture, forestry & fisheries journals, the citations-based gap measure cannot be calculated for these fields. Notwithstanding this, when we compare the AMS measure with the scientific publication-based measure ("Papers (WOS)") by their rankings, 10 out of the 12 fields where both measures have non-missing values differ by no more than three ranks (e.g., Mathematics is ranked 2nd in the AMS-based ratios, and is also ranked 2nd in the WoS-based ranking). The two outliers are basic medical research and agriculture, forestry & fisheries. For medicine, we suspect that there may be an over-representation of practitioners (i.e., practicing physicians). In agriculture, forestry, and fisheries, there may be a measurement error given that it encompasses a wide variety of fields. At the firm level, i.e., when the observations are weighted by the industries or scientific subfields of firms in the sample, the correlation between the two measures (r=0.527) is slightly higher than the correlation at the scientific field level (r=0.508), suggesting that fields with the highest mismatches between AMS and WoS are not very important in the patent classes used by our sample firms (Figure D2).

A direct ranking comparison between the citation-based measure and the other two measures is less feasible given the number of missing values. However, the fact that physics and chemistry have high (gap) scores, whereas clinical and medical sciences have relatively low gap scores, accords with the other measures. A notable outlier in this measure is Electrical engineering, which has a high score (5.80) partly

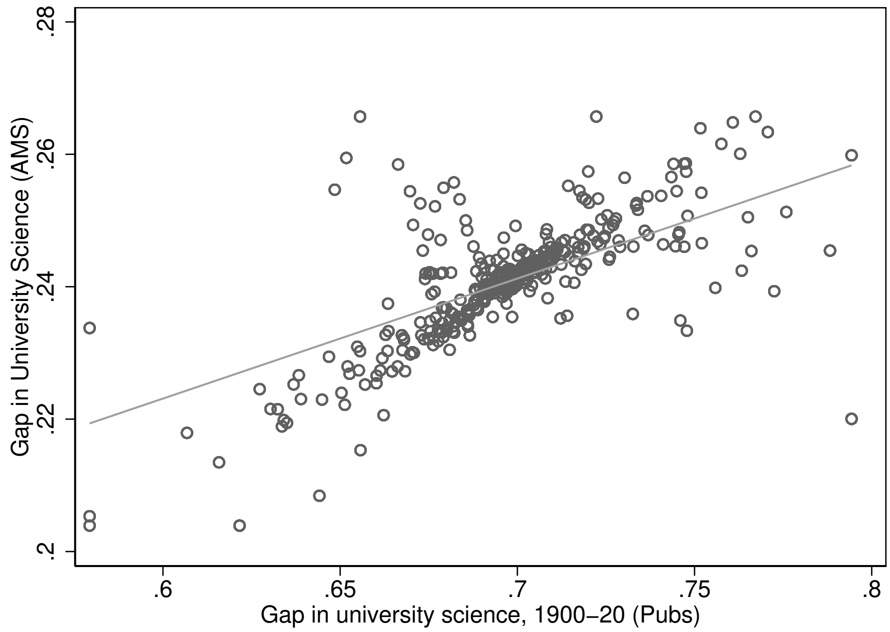
Table D9: Number of Scientists and Papers, Europe vs America

| | Papers (WOS) | | | Scientists (AMS) | | | Citations (WOS) | | |
|---|---|---|---|---|---|---|---|---|---|
| OECD Subfield Equivalent | U.S. | Europe | Ratio | U.S. | Europe | Ratio | U.S. | Europe | Ratio |
| 2.05 Materials engineering | 3,806 | 15,456 | 4.06 | 82 | 41 | 0.50 | 143 | 44 | 0.31 |
| 1.01 Mathematics | 5,334 | 19,556 | 3.67 | 525 | 229 | 0.44 | 134 | 71 | 0.53 |
| 2.07 Environmental engineering | - | - | - | 68 | 27 | 0.40 | - | - | - |
| 3.02 Clinical medicine | 43,007 | 81,883 | 1.90 | 178 | 65 | 0.37 | 6,017 | 2,200 | 0.37 |
| 1.03 Physical sciences and astronomy | 12,802 | 42,719 | 3.34 | 605 | 219 | 0.36 | 197 | 665 | 3.38 |
| 3.03 Health sciences | 5,121 | 9,373 | 1.83 | 63 | 22 | 0.35 | 1,042 | 336 | 0.32 |
| 3.01 Basic medical research | 32,556 | 34,614 | 1.06 | 928 | 324 | 0.35 | 4,845 | 2,721 | 0.56 |
| 1.04 Chemical sciences | 31,330 | 75,596 | 2.41 | 1,189 | 383 | 0.32 | 650 | 656 | 1.01 |
| 2.02 Electrical eng, electronic eng | 125 | 143 | 1.14 | 177 | 54 | 0.31 | 5 | 29 | 5.80 |
| 1.05 Earth and related environmental sciences | 7,996 | 1,189 | 0.15 | 561 | 167 | 0.30 | 369 | 128 | 0.35 |
| 1.06 Biological sciences | 39,262 | 44,261 | 1.13 | 1,482 | 360 | 0.24 | 3,764 | 3,285 | 0.87 |
| 2.03 Mechanical engineering | - | - | - | 134 | 32 | 0.24 | - | - | - |
| 2.01 Civil engineering | 1,010 | 636 | 0.63 | 120 | 19 | 0.16 | - | - | - |
| 4.01 Agriculture, forestry, fisheries | 2,112 | 3,594 | 1.70 | 385 | 60 | 0.16 | - | - | - |

*Notes:* This table presents the number of scientists (from AMS) and citation-weighted articles (from WoS) that have non-missing subject and affiliation fields. The "Ratio" column for the AMS sub-columns divides the number of European-affiliated American scientists by those unaffiliated with Europe. The rows are downward-sorted by this value. The "Ratio" column for WoS sub-columns divides the number of European-affiliated papers (published globally) divided by American-affiliated papers.

due to low overall citations (34 citations in total throughout the 20-year period, compared to chemistry, which made 1,306 total citations).[64] Excluding this outlier, the correlation between the citations-based measure and the publications-based measure is positive (r=0.302) at the scientific field level.

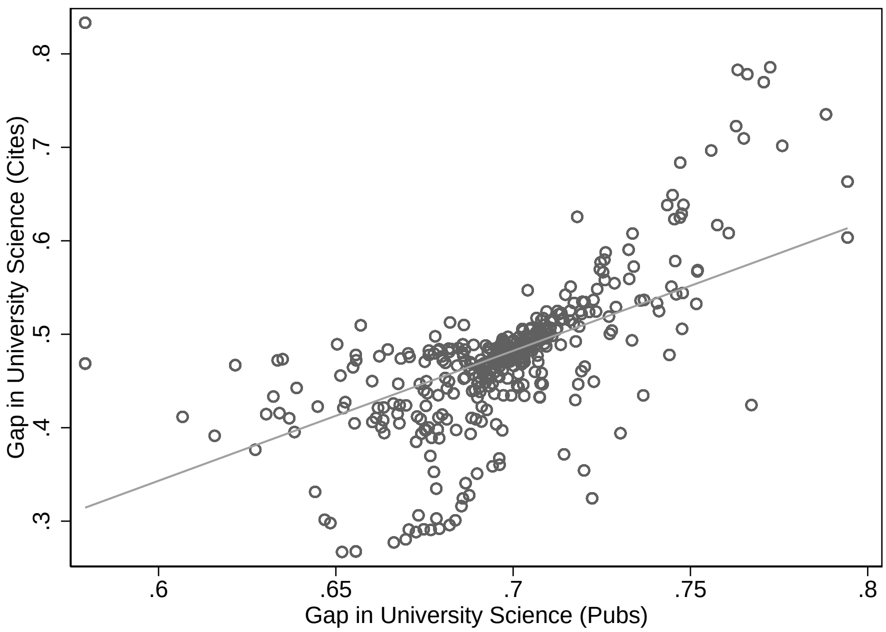
Figure D2: Comparison of Gaps in University Science




*Notes:* This figure compares the two scientific gap measures at the firm level. Higher values represent a larger gap between Europe and the United States. The AMS Scientist affiliation-based gap measure (on the vertical axis) is positively correlated with the publication volume-based gap measure (on the horizontal axis).

[64] It is unclear whether this represents a measurement error. On the one hand, the only electrical engineering journal in print during this period (1900-20) is American ("Proceedings of the Institute of Radio Engineers"") and the only other electrical engineering journal indexed in the SCI before 1940 is the BELL SYSTEM TECHNICAL JOURNAL, which is American. Therefore, electrical engineering can be thought of as an area of American excellence. However, it is also possible that this field still relied on European science in the 1900-20 period, since 21 (62%) of the 34 citations were made to physics journals (1.03).

Figure D3: Comparison of Gaps in University Science




*Notes:* This figure compares the two scientific gap measures at the firm level. Higher values represent a larger gap between Europe and the United States. The journal citation-based gap measure (on the vertical axis) is positively correlated with the publication volume-based gap measure (on the horizontal axis).