Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Research Policy 33 (2004) 287–321
research policy
www.elsevier.com/locate/econbase

The relationship between research and innovation in the semiconductor and pharmaceutical industries (1981–1997)

Kwanghui Lim*

NUS Business School, BIZ2-03-10, National University of Singapore, Singapore 117591, Singapore

Received 9 December 1999; received in revised form 7 July 2003; accepted 12 August 2003

Abstract

This paper evaluates the impact of basic and applied research on innovation in two industries. Whereas innovation in the pharmaceutical industry is closely tied to both basic and applied research, innovation in the semiconductor industry depends mainly upon applied research. Surprisingly, many firms perform little basic research, but they produce many innovations. Within each industry, firms pursue different R&D strategies: firms that emphasize basic research absorb more basic scientific knowledge than those that emphasize applied research. These findings suggest that future research must carefully consider how industry context and the composition of R&D mediate the relationship between research and innovation.

Keywords: Semiconductors; Pharmaceuticals; Innovation; Basic research; Absorptive capacity

1. Introduction

Technological change is the basis of economic growth (Solow, 1957; Romer, 1990) and many technological innovations depend upon scientific progress. Thus, the relationship between scientific research and innovation is an important one. Several theories suggest that scientific research is linked to innovation not just at the societal level, but also within firms. A firm's R&D efforts may directly improve its ability to produce innovations (Griliches, 1980) or indirectly help the firm to absorb outside knowledge (Cohen and Levinthal, 1990). A number of empirical studies support these theories (see Section 2). However, much of the evidence comes from the pharmaceutical and biotechnology industries; it remains uncertain whether research and innovation are as closely related in other industries. Furthermore, we do not yet know whether this relationship differs for basic than for applied research.

Assuming that a firm's internal research enhances its ability to innovate and to absorb external knowledge, we should expect research activity and innovation to be correlated at the level of the firm. I explore the strength of this relationship for the semiconductor and pharmaceutical industries, during the period 1981–1997. Both of these industries are highly science-intensive and exhibit high levels of spillovers, so we should expect a strong link between scientific research and innovation.

For each industry, I separately examine the relationship of basic and applied research to innovation. I adopt the NSF definition of "basic" research as that which seeks to understand a phenomena without specific applications in mind, while "applied" research seeks to produce knowledge for a specific

* Tel.: +65-6874-1342.
E-mail address: k@kwanghui.com (K. Lim).

0048-7333/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.respol.2003.08.001

end-use (NSF, 2000).[1] In reality, the distinction between basic and applied research is often ambiguous (Stokes, 1997). Nonetheless, this serves as a useful working definition. In the semiconductor industry, basic research includes efforts to understand solid-state physics, quantum mechanics and basic chemistry, while applied research aims towards improving manufacturing techniques and processes. In the pharmaceutical industry, basic research includes attempts to reveal the mechanisms and processes of disease, while applied research includes clinical trials, dosage testing and information about product labeling. This is in line with the definitions used by the Pharmaceutical Research and Manufacturers of America (PhRMA Profile 2001, pp. 12–13) and by academic scholars (e.g. Cockburn et al., 1999).

Regression analysis is used to explore the relationship between research and innovation. Each firm's innovation output is measured using the number of U.S. patents awarded to that firm, while research is measured using the number of articles it publishes in scientific journals (measurement issues are discussed below). These journals are divided into "basic" and "applied" categories to give us an estimate of how much basic and applied research is performed by each firm, albeit an imperfect one. Four different ways of classifying "basic" versus "applied" journals are used, so as to account for different interpretations for what constitutes "basic" and "applied" research.

The regression results show that in both industries, applied research is strongly correlated with innovation. In the pharmaceutical industry, a positive coefficient is found between basic research and innovation, which is consistent with prior research. However, in the semiconductor industry, basic research is *negatively* correlated with innovation. This is surprising if we believe that basic research improves a firm's innovativeness and helps it to absorb external knowledge.

Further analysis reveals that in the semiconductor industry, many firms produce very little basic research, but they produce a large number of innovations. This observation is empirically tested using a novel application of the Ellison–Glaeser index, which measures the extent to which one variable is more concentrated than another (see Appendix B). In the pharmaceutical industry, the E–G index reveals no surprise: pharmaceutical firms that produce more basic research also produce more innovation, so the concentration of basic research is similar to that of innovation. However, in the semiconductor industry, basic research is highly concentrated, while innovation is widespread across many firms. By itself, the fact that basic research is concentrated in the semiconductor industry is unsurprising: the scientists at the IBM Watson Laboratories and AT&T Bell Laboratories are famous for their basic scientific work, and a number have even won the Nobel Prize. The surprise is that other semiconductor firms produce a large number of innovations, even though they perform very little basic research. These results persist when citations counts are used to adjust for the quality of patents.

There are many possible explanations for why basic research is highly concentrated relative to innovation in the semiconductor industry. My interpretation is that semiconductor innovations depend primarily upon *applied* research. Hence, even firms that perform very little basic research can innovate because their applied research enables them to generate patents and to absorb applied knowledge from external sources. Circumstantial evidence supports this interpretation. The patents awarded to semiconductor firms make many citations to applied research journals, but they make very few citations to basic research journals. This suggests a greater dependence on applied research than on basic research. In contrast, the patents of pharmaceutical firms cite basic research journals even more than they do applied research journals.

The analysis of patent-to-science citations reveals another exciting insight. In both industries, the intensity of basic research varies greatly. Moreover, firms with a higher intensity of in-house basic research are more likely to cite basic scientific journals than those with a lower intensity of in-house basic research. This suggests that different kinds of absorptive capacity exist, and that a firm's absorptive capacity depends upon the composition of its R&D: firms that emphasize basic research (IBM, AT&T, Merck) absorb more basic scientific knowledge than firms that emphasize applied research (e.g. Motorola, Toshiba, Elan Pharmaceuticals).

[1] Others have defined "basic" research along the dimensions of originality, autonomy, length of time between discovery and use, motivation of researcher, institutional affiliation of researcher, and source of funding (see Stokes, 1997, p. 7, for a review).

This paper makes several contributions to the literature on innovation. Firstly, it shows that the relationship between research and innovation is contingent upon the industry and type of research involved. This study reinforces prior findings of a strong link between basic research and innovation in the pharmaceutical industry, while showing that the same does not apply in the semiconductor industry. Secondly, it suggests that absorptive capacity is a subtle concept, and that different types of absorptive capacity may exist. Basic research helps a firm absorb external basic research, while applied knowledge helps a firm to absorb applied knowledge. Thirdly, this paper makes several methodological contributions. It presents the first estimates for the concentration of basic and applied research relative to innovation using a comparable methodology across industries. The fact that basic research is concentrated in the semiconductor industry is itself uninteresting, but we learn something new by coupling this fact with the realization that innovation is widespread. The Ellison–Glaeser index offers a formal model for making such comparisons, controlling for several competing explanations. Finally, this paper provides statistical evidence that publication counts are a valid measure of a firm's research output, and explores ways of decomposing this research into basic and applied categories using different journal classification schemes.

In the next section, I discuss the relationship between a firm's internal research and its level of innovation. Section 3 describes the dataset, variables used, and measurement issues. Section 4 presents the regression analysis for exploring the extent to which innovation is closely related to basic and applied research. Section 5 delves deeper into the analysis, using the Ellison–Glaeser Index to test if either basic or applied research is more highly concentrated than innovation in each industry. Section 6 discusses possible interpretations of the results, along with the limitations of this study. Section 7 concludes.

2. Scientific research and innovation

The section begins with a review of prior studies exploring the relationship between research and innovation, and then examines the nature of research in the context of the pharmaceutical and semiconductor industries.

Among the many factors that influence a firm's performance, perhaps the most crucial is its ability to innovate. Empirical evidence suggests that a firm's research effort, when combined with conventional inputs, has a profound influence on its productivity (e.g. Griliches, 1980; Hall, 1996). As shown in Fig. 1, a firm's research activities can contribute directly to its innovative capability, or it can contribute indirectly by enhancing the firm's absorptive capacity, which is the ability to "recognize the value of external information, assimilate that information, and then apply it to commercial ends" (Cohen and Levinthal, 1990). Due to the cumulative nature of scientific progress, innovation may lead to new research (Scotchmer, 1991), as is represented by the dotted line in Fig. 1. When such feedback occurs, it reinforces the link between research and innovation.

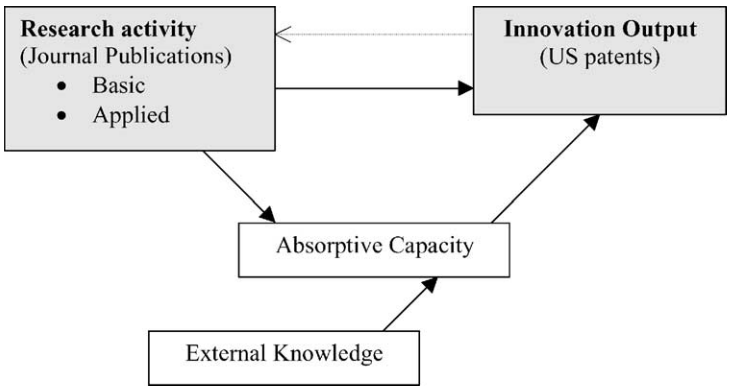


Fig. 1. Basic research, applied research and innovation.

The benefits of research to a firm's innovative capacity are derived not only from applied research, but from basic research as well (Mansfield, 1981; Griliches, 1986). According to Rosenberg (1990), a firm that performs basic research may benefit from first-mover advantage, unexpected innovations arising from the research, credibility in contests for government contracts, an improved ability to select areas of applied research, and an improved ability to evaluate the outcome of applied research.

Despite the potential benefits, firms may be reluctant to invest in research. This is because the knowledge produced by R&D is like a public good, which spills easily from the innovating firm to other companies that can free ride on its efforts (Nelson, 1959; Arrow, 1962). There are two kinds of spillovers, the first being the mispricing of factor inputs, and the second being the inability of firms to fully appropriate the economic benefits of the knowledge they produce (Mohnen, 1994; Griliches, 1992). Mispricing merely creates measurement problems, while weak appropriability is critical because it alters the incentives of firms to invest in research (Griliches, 1992). All else being equal, the benefits from basic research are harder to appropriate than applied research.[2] Therefore, firms may be more reluctant to invest in basic than in applied research.

Several researchers have questioned whether knowledge spillovers occur as easily as portrayed by Nelson and Arrow. Cohen and Levinthal (1989, 1990) argue that especially when learning is difficult, firms may need to invest in R&D in order to effectively absorb external knowledge. This prediction is broadly consistent with empirical data. Jaffe (1986, p. 993) found that the interaction between a firm's R&D expenditure and spillovers is strongly correlated with the firm's performance.[3] Other studies have found evidence that performing research helps a firm to absorb external knowledge, including Gambardella (1992), Henderson and Cockburn (1996), Arora and Gambardella (1994), Lane and Lubatkin (1999), and Zucker and Darby (1995). The need to develop absorptive capacity may be quite important because a great deal of knowledge exists outside the firm, especially given the crucial role that universities play in creating new knowledge.

Unfortunately, the literature on absorptive capacity does not deal satisfactorily with how much of basic and applied research are needed to absorb spillovers. This is an important question because investing in basic and applied research is costly and often irreversible. Existing theories assume that some basic research is necessary, based on the notion that knowledge is tacit and that a firm has to be involved in an activity to understand or exploit relevant knowledge (Nonaka, 1994; Leonard-Barton, 1995, Chapter 6). Cohen and Levinthal (1990) speculate that "firms may conduct basic research less for particular results than to be able to provide themselves with the general background knowledge" that would help them exploit technological advances more effectively (p. 148). Hence, "as a firm's technological progress becomes more closely tied to advances in basic science (as has been the case in pharmaceuticals), a firm will increase its basic research, whatever its degree of product-market diversification". It remains unclear how important basic and applied research are for absorbing external knowledge. In the remainder of this section, I examine the role of research in the pharmaceutical and semiconductor industries. The view that emerges is one in which the roles of basic and applied research are heavily shaped by firm-level characteristics (economies of scale and scope) and the conditions within each industry (the strength of intellectual property protection, the role of universities and consortia, and the level of technological maturity).

2.1. Pharmaceuticals and biotechnology

Prior research shows that a strong link exists between basic research and innovation in the pharmaceutical and biotechnology industries. According to Gambardella (1992), the more basic research a pharmaceutical firm performs, the more patents it produces.[4] He points out "the winning models of the

[2] A firm's appropriability from innovation is also affected by its size and access to complementary assets (Levin et al., 1987; Teece, 1987).

[3] In contrast, Levin and Reiss (1988) do not assume that a firm's research interacts with knowledge spillovers to increase productivity. They propose that if a firm's own R&D and that of its rivals are strategic complements, an increase in spillovers might actually increase each firm's R&D expenditure.

[4] However, it is inconclusive from this study whether research-intensive firms produced more patents because they are better at capturing spillovers or because they have higher research productivity.

U.S. pharmaceutical industry during the 1980s were firms like Merck, which organized their internal research like academic departments" (p. 404). Cockburn and Henderson (1998) found that drug discovery firms with a strong research orientation produced a greater number of important patents. According to them, successful firms decentralized decision-making on the allocation of R&D resources and promoted scientists based on their publications in the open literature.

For firms that develop new drugs, investing in basic research brings tangible benefits. Firstly, the nature of knowledge is unique in the Pharmaceutical Industry. Unlike in many other fields, medical research seeks a fundamental understanding of phenomena and yet is motivated by practical goals (Stokes, 1997). Secondly, appropriability is very high in the pharmaceutical industry. Patent protection is extremely effective (Levin et al., 1987) and the rate of imitation is slower for ethical drugs than for other products (Mansfield et al., 1981). Thirdly, the industry has high entry barriers because buyers have low bargaining power, because drug development is a complex process requiring specialized knowledge, and because extensive human trials are needed to satisfy regulatory requirements (Pisano, 1997, pp. 55–57). The link between drug discovery and basic science has increased over time (Pisano, 1997; Cockburn et al., 1999). Firms have moved away from randomly screening a large number of potentially useful compounds, towards a more systematic approach called "rational drug design". This involves exploiting knowledge about the biochemical mechanisms causing a disease to identify or develop chemicals that inhibit such mechanisms (Pisano, 1997, p. 64). Firms are beginning to use biotechnology as a process technology as well as a research tool to improve their search for new drugs (Cockburn et al., 1999, pp. 379–381).

In addition to the direct benefits, basic research also helps pharmaceutical and biotechnology firms to absorb external knowledge. Encouraging such research activity allows a firm to hire high-quality researchers and to develop strong social networks, so that they are "actively connected to the wider scientific community" (Cockburn and Henderson, 1998, p. 158). The ability to hire highly talented researchers is underscored by research on "stars", or scientists who produce many papers, are highly cited, and collaborate heavily in public science. Employing star scientists has a large and positive impact on the research productivity of biotechnology firms (Zucker and Darby, 1995).

As a result of the direct and indirect benefits mentioned above, pharmaceutical companies engage heavily in basic research. The research produced by these firms is of tremendously high quality and is cited as widely as that from NIH-supported medical schools (Koenig, 1982).

2.2. Semiconductors

Just like the pharmaceutical industry, the semiconductor industry is highly dependent upon scientific breakthroughs. In order to learn more about the roles of basic and applied research in this industry, I interviewed a dozen researchers and managers at semiconductor firms in 1998 and 1999. From my fieldwork, it is not apparent whether basic research is as closely linked to innovation in this industry as in the pharmaceutical industry. Many of those interviewed were quick to point out that the basic research performed at AT&T's Bell Laboratories is crucial to that firm's ability to commercialize new technologies. Yet others were skeptical that basic research is really necessary. Firstly, the basic research upon which the semiconductor industry is based was developed many decades ago (Macher et al., 1999, p. 273). According to researchers and managers at semiconductor firms, much of the fundamental knowledge required is readily available from books, conferences, journals and equipment suppliers. Secondly, appropriability is weak in the semiconductor industry because patents provide very little protection (Hall and Ziedonis, 2001); this was true for as long as the industry has existed (Flatherty, 1984; Tilton, 1971). The wide diffusion of basic knowledge and low appropriability in this industry result in low entry barriers, as is consistent with the large number of new firms entering the sample (see Section 4).

Semiconductor companies spend vast sums of money on R&D.[5] However, most of this R&D is applied rather than basic, including the search for new materials (e.g. low-k dielectrics, copper interconnects), new process technologies, and better tools and methodologies for designing circuits. In part, this

[5] In the year 2000, U.S. firms invested US$ 14 billion in R&D, representing 14% of sales (Source: Semiconductor Industry Association website).

is because semiconductor products are “systemic” in nature (Hall and Ziedonis, 2001), involving hundreds of interwoven design and process steps. As such, output yields are sensitive to minor changes in design and process parameters. Any breakthrough in basic research would require a great deal of effort to commercialize because it has to be carefully incorporated into the rest of the process. In such an environment, investments in process technologies and tacit, embedded knowledge are better ways of competing than basic research.

While the direct benefits of basic research might be low, surely firms must depend on basic research to overcome new technical challenges? Prior research shows that knowledge spillovers are critically important in the semiconductor industry (Mowery, 1983; Appleyard, 1996). But, unlike firms in the pharmaceutical industry, semiconductor firms may be able to rely on other mechanisms to develop absorptive capacity instead of performing their own basic research. An interesting feature of the semiconductor industry is the strong presence of institutions such as the Semiconductor Research Corporation, International Sematech, MARCO and SELETE. These institutions combine the resources of industry players. They invest those resources in pre-competitive research at universities, convert the output into practical knowledge, and diffuse the results. The presence of such institutions reduces the need for individual firms to invest in their own basic research.

Apart from these institutions, labor mobility is another important mechanism for spillovers to occur in the semiconductor industry (Rosenkopf and Almeida, 2003; Tilton, 1971; Wilson, 1980). Individuals move frequently across organizational boundaries, bringing with them the knowledge embodied in techniques and processes. In fact, the diffusion of basic and applied research through labor mobility is responsible for the very existence of this industry. William Shockley, one of the three inventors of the transistor, left Bell Laboratories to form his own company. Subsequently, key personnel left Shockley’s company to form Fairchild Semiconductors, which spawned Intel Corporation and much of the Silicon Valley (Riordan and Hoddeson, 1997).

In such an environment, it is likely that only large and diversified firms such as IBM and AT&T can really capture the benefits of basic research identified by Rosenberg (1990) such as improved cognition and the ability to exploit unexpected innovations. Intel—highly successful but narrowly focused in the microprocessor market—eschews research laboratories. According to Intel co-founder Gordon Moore, “We don’t have a separate R&D laboratory . . . the development work is done right on the manufacturing floor” (Jelinek and Schoonhoven, 1990, p. 295). Intel spends heavily on development but performs little basic research, exploiting spillovers by investing in research at universities and research consortia (Moore, 1996, p. 170).

The short survey presented here shows that the relationship between research and innovation is a complex one. Outside the pharmaceutical and biotechnology industries, there are few conclusive studies showing a strong link between *basic* research and innovation at the level of the firm. Indeed, the relationship between science and spillovers is highly complex, and varies across time and between technologies (Mowery and Rosenberg, 1989, p. 147). Empirical research is needed to discover whether the lessons learnt in the pharmaceutical and biotechnology industries apply elsewhere. This paper takes a step in that direction, by comparing and contrasting the nature of science and innovation in two industries that the author is familiar with: pharmaceuticals and semiconductors.

3. Data and measures

The sample consists of all major semiconductor and pharmaceutical firms worldwide between 1981 and 1997 (data were available in electronic format for this time period). The construction of this dataset is described in Appendix A, which also contains a list of firms included in the sample. For the semiconductor sample, “fabless” firms are excluded as they focus on designing circuits and do not actually manufacture their own semiconductor chips. The number of semiconductor firms in the sample was 49 in 1981, but it grew to 81 firms in 1997 due to the entry of new competitors. The number of pharmaceutical firms in the sample was 35 in 1981, but shrank to 30 firms in 1997, as several of the firms merged. The contrast between the semiconductor and pharmaceutical industries hints at different competitive dynamics at work.

In the remainder of this section, I discuss how I measured the basic research, applied research and

Table 1
U.S. Patent Classes relevant to each industry

Semiconductors	Pharmaceuticals
156/345: Film deposition	*424*: *Drug, bio-affecting and body treating compositions*
257: *Active solid-state devices* (e.g. *transistors, solid-state diodes*)	435: Chemistry: molecular biology and microbiology
327: Miscellaneous active electrical nonlinear devices, circuits, and systems	436: Chemistry: analytical and immunological testing
330: Amplifiers	*514*: *Drug, bio-affecting and body treating compositions*
331: Oscillators	530: Chemistry: natural resins or derivatives; peptides or proteins; lignins or reaction products thereof
365: Static information storage and retrieval	585: Chemistry of hydrocarbon compounds
438: *Semiconductor device manufacturing: process*	
711: Electrical computers and digital processing systems: memory	

Note. The most important patent classes are shown in italics.

innovation produced by each firm. Measuring these variables is necessary in order to empirically test the relationship among them.

3.1. Patents as a measure of innovation output

Patents are often used to measure innovation output because an idea must be sufficiently novel to qualify for patent protection. There is some basis for using patents this way, since the number of patents awarded to a firm correlates with its R&D investments (Pakes and Griliches, 1984), sales from new products (Comanor and Scherer, 1969) and the timing of new innovations (Basberg, 1982). Patents also strongly correlated with expert rankings of the technological capability of firms (Narin et al., 1987). Nonetheless, the use of patent data for measuring innovation is far from ideal (Griliches, 1990). Many inventions are not patentable because they fail to meet the criteria of being sufficiently novel, useful and non-obvious. Other inventions may not be patented because the inventor relies instead on secrecy or other means of protecting intellectual property (Basberg, 1987). Inventors may also be deterred by the high cost, uncertainty and information disclosure requirements associated with obtaining a patent. Despite these difficulties, patents are widely used to measure innovation output (for recent studies, see Ahuja, 2000; Hall and Ziedonis, 2001; Fleming and Sorenson, 2001). Innovation is inherently difficult to measure and firms are reluctant to divulge sensitive internal metrics to outsiders. Thus patents, however imperfect, are often the best measure available. In this paper, I measure each firm's innovation output as the number of patents it is awarded.

One of the biggest concerns about using patent data is that the economic value of patents is highly skewed (Griliches, 1990; Harhoff et al., 1999). Researchers have begun to use the number of citations that a patent receives as a proxy for its true value (Hall et al., 2000; Fleming and Sorenson, 2001). In line with this approach, citation-weighted counts are used to verify the robustness of my empirical results.

Patent data for each firm in the sample were obtained from the U.S. Patent Office. Firms in both industries are involved in multiple businesses—semiconductor firms also make electronic, computer, and telecommunications products while pharmaceutical companies also produce industrial chemicals, personal-care products and hospital supplies. In order to restrict the analysis to innovations directly related to the semiconductor and pharmaceutical industries, I include only patents that fall within the U.S. Patent Classes listed in Table 1. I call these "relevant patents".[6] The patent classes in Table 1 were chosen with reference to the USPTO Technology Profile Reports (TAF3290P for semiconductors and TAF3250P for pharmaceuticals) and by manually examining several hundred patents in each industry.

Between 1981 and 1997, the semiconductor firms in the sample received 46,436 relevant patents while the pharmaceutical firms obtained 22,916 relevant

[6] Most patents fall into more than one patent class; I include a patent in the sample if it is listed in at least one of the patent classes in Table 1.

Table 2
Descriptive statistics

Variable (per firm per year)	Average	S.D.	Median	Minimum	Maximum
Semiconductors					
No. of relevant patents	41	68	12	0	420
No. of basic research articles (JCRBas = 1)	16	57	0	0	510
No. of applied research articles (JCRBas = 0)	64	134	10	0	918
No. of articles co-authored with academic and public-sector laboratories	26	68	4	0	624
Pharmaceuticals					
No. of relevant patents	40	35	31	0	236
No. of basic research articles (JCRBas = 1)	36	49	16	0	294
No. of applied research articles (JCRBas = 0)	88	92	62	0	560
No. of articles co-authored with academic and public-sector laboratories	67	84	37	0	574

patents. On average, each semiconductor firm received 41 relevant patents per year, while each pharmaceutical firm received 40 patents per year (see Table 2). Relevant patents comprise one-fifth of all patents awarded to the semiconductor and pharmaceutical firms, respectively; many patents do not fall within the relevant patent classes. Fortunately, the reader need not be concerned that my choice of relevant patent classes might bias the results towards certain kinds of products or innovations. The empirical results are robust even if we include a broad range of patent classes other than those in Table 1.[7] In fact, the results persist if we include *every single* patent class in which these firms are awarded patents, hence covering their entire portfolio of innovations (see Section 5).

3.2. Publications as a measure of basic and applied research

I measure each firm's research output as the number of articles it published. One of the most important ways that scientists use to establish their reputations is by publishing their research (Stephan, 1996, p. 1201). As such, publication data have been widely used for measuring the output of scientists at universities and publicly funded research laboratories (Martin and Irvine, 1983; Stephan, 1996).[8] Publication data have also been used to measure the scientific output of firms and track the flow of knowledge (Koenig, 1982; Gambardella, 1992; Hicks, 1996; Appleyard and Kalsow, 1999; Gittelman and Kogut, 2003). The validity of using publication data is discussed later in this section.

Data on the scientific publications of each firm between 1981 and 1997 were obtained from the ISI Science Citation Index (SCI). Only original research articles were included; I eliminated meeting notes, review articles and book reviews. The SCI is an excellent source because it covers a broad range of basic and applied scientific journals. Moreover, it lists up to 255 authors and addresses for each publication,[9] unlike other databases, which only include the institutional affiliation of the first author (e.g. Compendex, INSPEC and Biosis). However, the SCI does not indicate which authors are associated with each address, so it is impossible to weight each published article by the number of authors from each firm. I, therefore, adopt the following convention: if a firm is listed in a published article, I add one to the number of articles

[7] Pharmaceutical firms also receive numerous patents for organic compounds (Patent Classes 532–570), but these were excluded because many organic compounds are unrelated to pharmaceuticals (e.g. industrial solvents, petrochemicals, agrochemicals). Including these patents does not qualitatively change the results.

[8] Many excellent studies using publication data appear in the journal *Scientometrics*.

[9] Source: personal communication with ISI staff. In the sample, each article by semiconductor firms had between 1 and 51 authors with a mean of 1.8. Each article by pharmaceutical firms had between 1 and 243 authors, with a mean of 2.2.

published by that company.[10] For articles with authors from multiple firms, this procedure fails to account for the different levels of effort contributed by each firm. Fortunately, the problem is not severe, as only a small number of articles have authors from multiple firms (of the 90,365 articles in the semiconductor industry, only 1117 had authors from more than one firm; the corresponding number for the pharmaceutical industry was 1679 out of 86,637).

A method is needed to determine how much of a firm's research activity is basic, and how much is applied. I classify each research article as "basic" or "applied" based on the *journal in which the article is published*. This makes the classification scheme tractable because there are many more articles than there are journals. The price paid is the inability to capture heterogeneity among papers within each journal. Fortunately, recent research suggests that the influence of an article is determined much more by the characteristics of the journal it is published in than of the article itself (Van Dalen and Henkens, 2001). Moreover, specialization occurs among journals in most disciplines. Thus, it is usually possible to distinguish an applied journal oriented to solid-state engineers from a basic science journal aimed at quantum physicists.[11]

Four different schemes are used to classify journals into "basic" and "applied" categories. This helps to overcome the shortcomings of relying on only one scheme, and it also allows for different interpretations of what constitutes "basic" research. The schemes are as follows:

- The *JCRBas* scheme: The Science Citation Index publishes a Journal Citation Report, which places journals into "JCR categories". I classified some of these as "basic" and others as "applied" based on the NSF definition of basic research (see page 1). The result is shown in Table 3. For pharmaceuticals, "basic" JCR categories include biochemistry, molecular biology, and genetics; for semiconductors, these categories include pure physics, mathematics, and chemistry. Each article was analyzed and the variable *JCRBas* set to 1 if it was published

Table 3
Classification of journal categories into "basic" and "applied"

JCR journal category	Semiconductor industry	Pharmaceutical industry
All Clinical Medical Journals	-U-	Applied
Biochemistry and Molecular Biology	-U-	Basic
Biology	-U-	Basic
Biophysics	-U-	Basic
Cell Biology	-U-	Basic
Chemistry, Analytical	Applied	Applied
Chemistry, Applied	Applied	Applied
Chemistry, Inorganic and Nuclear	Applied	-U-
Chemistry, Medicinal	-U-	Applied
Chemistry, Organic	-U-	Applied
Chemistry, Physical	Basic	-U-
Chemistry	Basic	-U-
Engineering (Electrical, Chemical and Nuclear)	Applied	-U-
Genetics and Hereditary	-U-	Basic
Maternal Science	Applied	-U-
Mathematics, Applied	Applied	-U-
Mathematics, Miscellaneous	Applied	-U-
Mathematics	Basic	-U-
Medicine, General and Internal	-U-	Applied
Medicine, Research and Experimental	-U-	Basic
Microbiology	-U-	Basic
Multidisciplinary Science	Basic	Basic
Physics, Applied	Applied	-U-
Physics, Atomic, Molecular and Chemistry	Basic	-U-
Physics, Condensate Matter	Basic	-U-
Physics, Mathematical	Basic	-U-
Physics, Miscellaneous	Applied	-U-
Physics, Nuclear	Applied	-U-
Physics, Particles and Fields	Applied	-U-
Physics	Basic	-U-
... Other categories		

Notes. -U- indicates an unrelated field. Categories not shown are either applied or unrelated.

[10] Special care was taken when dealing with university laboratories having similar names as these firms. Wrongly attributing university publications to a company can distort the results because universities are a major locus of basic research. Several laboratories are owned by companies but located on university campuses (e.g. Lilly Laboratory at Indiana University). Others are named after family trusts unrelated to the business (Wellcome, DuPont). Such cases were identified by searching the web pages of universities and companies involved and by contacting them where necessary. In the small number of cases where uncertainty could not be resolved, the observation was dropped.

[11] Some journals are highly multidisciplinary and cover both basic and applied research, including *Nature*, and *Science*. Multidisciplinary journals are classified as "basic" because this adds a conservative bias against finding that basic research is concentrated.

Table 4
Journals with the highest number of publications (1981–1997) by companies in the sample

Basic journals (JCRBas = 1)	No. of publications	Applied journals	No. of publications
Semiconductors			
Phys. Rev. B	2956	Appl. Phys. Lett.	6432
Phys. Rev. Lett.	2336	J. Appl. Phys.	4961
Surface Sci.	937	Electronics Lett.	3695
Physica C	834	IEEE Trans. Magnetics	2387
J. Chem. Phys.	776	J. Vacuum Sci. Technol. B	2201
Inst. Phys. Conf. Series	743	IEEE Trans. Electron Devices	2114
Solid State Commun.	621	Jpn. J. Appl. Phys. 1	2100
Chem. Phys. Lett.	408	J. Electrochem. Soc.	1824
J. de Physique	395	Jpn. J. Appl. Phys. 2	1785
J. Phys. Chem.	393	J. Crystal Growth	1744
Physica B	384	IEEE J. Solid State Circuits	1617
Science	280	IEEE Photonics Technol. Lett.	1369
J. Am. Chem. Soc.	277	J. Lightwave Technol.	1255
Phys. Rev. A	265	J. Vacuum Sci. Technol. A	1218
Nature	231	IEEE Electron Device Lett.	1092
Total	11836		35794
Pharmaceuticals			
J. Biol. Chem.	1925	Tetrahedron Lett. (see note)	2255
Bioorg. Med. Chem. Lett.	1285	J. Med. Chem.	2017
Proc. Natl. Acad. Sci. U.S.A.	1080	J. Pharmacol. Exp. Therapeut.	1292
Biochem. Biophys. Res. Commun.	1053	J. Antibiotics	1118
Biochemistry	890	Antimicrob. Agents Chemother.	994
Ann. New York Acad. Sci.	588	Eur. J. Pharmacol.	960
FEBS Lett.	534	J. Cardiovasc. Pharmacol.	796
Biochem. Pharmacol.	522	J. Immunol.	782
J. Am. Chem. Soc.	510	J. Chromatogr.	615
Life Sci.	498	Tetrahedron (see note)	589
Biochem. J.	435	J. Pharmaceut. Sci.	576
Curr. Ther. Res.	429	Pharmaceut. Res.	568
J. Chem. Soc.	415	J. Antimicrob. Chemother.	541
Helvetica Chim. Acta	412	Int. J. Pharmaceut.	518
Nature	387	Cancer Res.	497
Total	10963		14118

Note. The empirical results are the same using the CHI classification scheme (Appendix C), under which *Tetrahedron* and *Tetrahedron Letters* are “basic” journals.

in a journal within a basic JCR category. Table 4 shows the top basic and applied journals in each industry based on this scheme.

- The *CHIbas* scheme: a journal is “basic” if it has a high *basic science score* as assigned by CHI Research (a private company that specializes in evaluating the quality of intellectual property).
- The *HiIPF* scheme: a journal is “basic” if it is *highly cited* by published articles. One view of applied research is that it builds upon the foundation laid by basic scientific breakthroughs. Assuming that authors cite the knowledge that they rely upon, it is possible that basic research articles will end up more heavily cited than applied ones.
- The *HiAcad* scheme: a journal is considered “basic” if it has a large number of academic authors.

The main text of this paper reports the results using the *JCRBas* scheme. Appendix C further describes the other schemes, and shows that the empirical results are essentially unchanged if we use any of the three alternative schemes.

Table 2 presents descriptive statistics. The average pharmaceutical firm is more publication-intensive than a semiconductor firm, with more basic publications, applied publications, and articles co-authored with academic and public sector laboratories. It also has a larger proportion of its publications in basic research journals (29% for pharmaceuticals versus 20% for semiconductors).[12] The standard deviations are higher for semiconductors than for pharmaceuticals, indicating greater heterogeneity among semiconductor firms.

3.2.1. The validity of using publication data

There are several limitations of using publication data. Not all research is submitted for publication. And even if a research article is submitted to a journal, it might end up being rejected. The use of publication data to measure scientific output is probably more questionable for firms than for academic institutions: firms have less of an incentive to publish their work as this involves disclosing potentially valuable information and because patents cannot be awarded for ideas that have already been published. These concerns should be treated seriously because in contrast to patent data, little work has been done to validate the use of publication data to measure private-sector research output.

An alternative way to measure the research activity of each firm would be to use data on R&D expenditure. However, the *composition* of such data into basic versus applied categories is generally not available at the business unit level.[13] Several researchers have painstakingly obtained such data through interviews and surveys (e.g. Mansfield, 1981; Ernst, 1998). However, the firms I approached were generally unwilling to share such sensitive information. In any case, each firm and industry has a different way of accounting for "basic" versus "applied" research expenditure (Mansfield, 1981). In contrast, publication data are publicly available, both for publicly traded companies and privately owned ones, and thus verifiable. As well, papers that are submitted to the same journal are put through the same peer-review process, so publications may be more comparable across firms than expenditure data.

In order to explore the validity of publication output as a measure of research activity, I regressed each firm's publication output against R&D expenditure and sales (to control for firm size). Data for this analysis were obtained from *CompuStat* for the period 1981–1997. R&D expenditure and sales figures were available only for 41 semiconductor and 13 pharmaceutical firms. Each regression was performed with and without firm fixed-effects and in each case, the coefficient for R&D expenditure was positive and significant. The results of the fixed effect regressions are shown below, with standard errors in the parentheses:

$$\text{Semiconductors : } \ln(\text{publications}) = \underset{(0.5)}{1.9} + \underset{(0.11)}{0.95}\ln(\text{R\&D}) - \underset{(0.11)}{0.55}\ln(\text{Sales})$$

$$\text{Pharmaceuticals : } \ln(\text{publications}) = \underset{(0.7)}{-1.1} + \underset{(0.10)}{0.49}\ln(\text{R\&D}) + \underset{(0.15)}{0.34}\ln(\text{Sales})$$

Thus, there is a strong correlation between publications and R&D expenditure, even after controlling for the size of the firm and firm heterogeneity. The fit of the data were also high, with an R-squared of 0.90 for semiconductors and 0.88 for pharmaceuticals. The positive and significant coefficients estimated for R&D expenditure lends credence to the use of publications as a measure of R&D activity within the firm.

4. Regression analysis

Regression analysis is used to explore the extent to which basic research, applied research and innovation are related. The following reduced-form equation is estimated:

$$\text{Innov}_{jt} = \alpha + \beta_1 \text{BasRes}_{jt} + \beta_2 \text{AppRes}_{jt} + \beta_3 t + \varepsilon_t \tag{1}$$

where Innov_{jt}: firm j's innovation output in year t; BasRes_{jt}: firm j's basic research activity in year t; AppRes_{jt}: firm j's applied research activity in year t.

As discussed in Section 2, a firm's basic and applied research should improve its ability to innovate, both

[12] The semiconductor firms in the sample published 72,449 relevant applied research articles and 17,916 basic articles. The pharmaceutical firms published 61,458 applied and 25,179 basic research articles.

[13] The NSF provides data on R&D expenditures at the level of each industry, but not the firm (see National Science Foundation, 1998).

Table 5
Pair-wise correlation coefficients

Variable (per firm per year)	Patents	Basic research	Applied research	Applied research (5-year lag)	Applied research (5-year lag)	Year
Semiconductors						
No. of relevant patents	1.00					
No. of basic research articles (JCRBas = 1)	0.38	1.00				
No. of applied research articles (JCRBas = 0)	0.48	0.87	1.00			
No. of basic research articles (5-year lag)	0.37	0.85	0.75	1.00		
No. of applied research articles (5-year lag)	0.47	0.75	0.88	0.87	1.00	
Year	0.24	0.01	0.01	0.06	0.07	1.00
Pharmaceuticals						
No. of relevant patents	1.00					
No. of basic research articles (JCRBas = 1)	0.54	1.00				
No. of applied research articles (JCRBas = 0)	0.57	0.92	1.00			
No. of basic research articles (5-year lag)	0.47	0.70	0.60	1.00		
No. of applied research articles (5-year lag)	0.52	0.63	0.64	0.89	1.00	
Year	0.35	0.32	0.32	0.26	0.26	1.00

because of the productivity benefits and as it increases the firm's absorptive capacity. Hence, we should expect the estimated values of β_1 and β_2 to be positive and significant. The magnitude of β_1 shows the strength of the relationship between basic research and innovation, while β_2 shows the strength of the relationship between applied research and innovation. The estimation of these coefficients separately is an important contribution of this paper. An additional term, $\beta_3 t$, is used to control for time trends.

The model is estimated using several approaches. Ordinary Least Squares (OLS) is used to estimate the baseline model in Eq. (1). A fixed-effects model is also estimated, in order to take into account firm heterogeneity. This is important because many factors may vary across firms (profitability, product lines, managerial practices towards research, etc.); a fixed-effects model helps to control for these differences. Poisson and negative binomial specifications are also estimated, so as to account for the fact that patent counts are non-negative integers (Hausman et al., 1984). The use of Poisson models has become standard practice for analyzing patent data (e.g. Hall and Ziedonis, 2001; Ahuja, 2000). In order to account for time lags between research and innovation, the regressions were also performed using lagged independent variables.

Table 5 presents pair-wise correlation coefficients. Due to the large sample size, all the correlations are statistically significant at 5%. The table shows that in both industries, the number of patents received by a firm correlates with its basic and applied research publications. The variable for "lagged basic research" is highly correlated with its unlagged value, so these variables are not used within the same regression (the same situation arises for applied research). There is also a high correlation between basic and applied research, raising concerns about multicollinearity. However, multicollinearity does not bias the estimates of β_1 and β_2, but only affects the precision with which standard errors may be estimated. In this study, the problem is partly mitigated by the large sample size, which improves the precision of the estimated standard errors ($N = 1129$ for semiconductors; $N = 571$ for pharmaceuticals).

Tables 6 and 7 show the regression results for the semiconductor and pharmaceutical industries respectively. For the semiconductor industry, the coefficient for applied research is positive and significant (for the OLS model, $\beta_2 = 0.31$). The estimated coefficient for basic research is *negative* and significant (for OLS, $\beta_1 = -0.19$). This is quite surprising and will be further explored in the next section. To check for multicollinearity, I computed the Variance Inflation Factor (VIF) for each variable. As a rule of thumb, multicollinearity is suspected if VIF values above 10 are observed.[14] For semiconductors, the VIF values for basic and applied research were both only around 3.9,

[14] See www.princeton.edu/~slynch/OLDCLASSES/Multicollinearity.doc.

Table 6
Semiconductor firms (dependent variable is the number of patents per firm per year, 1981–1997)

Explanatory variable	OLS	OLS with firm fixed effects	Poisson	Negative binomial	Negative binomial (5-year lag)
No. of basic research articles (JCRBas = 1)	−0.19* (0.06)	−0.98* (0.11)	−0.003* (0.0009)	−0.012* (0.001)	
No. of applied research articles (JCRBas = 0)	0.31* (0.02)	0.57* (0.05)	0.004* (0.00004)	0.012* (0.0008)	
No. of basic research articles (5-year lag)					−0.014* (0.001)
No. of applied research articles (5-year lag)					0.012* (0.001)
Year	3.4* (0.4)	4.0* (0.24)	0.09* (0.001)	0.08* (0.008)	0.08* (0.01)
Firm fixed effects		Significant			
Regression statistics					
Adj. R-squared	0.29	0.74			
Log-likelihood			−31497	−4856	−3950
Overdispersion for negative binomial (α)				1.4*	1.5*
N	1129	1129	1129	1129	874

No. of observations: 1129. Standard errors are shown in parentheses.
* Significant at the 5% level.

so the large sample allowed precise estimates to be made, despite collinearity.

For the pharmaceutical industry, the estimates for basic and applied research are both positive (for the OLS model, $\beta_1 = 0.08$ and $\beta_2 = 0.17$). The estimates for applied research were statistically significant, but not those for basic research. It is difficult to tell whether the insignificance of β_1 is due to

Table 7
Pharmaceutical firms (dependent variable is the number of patents per firm per year, 1981–1997)

Explanatory variable	OLS	OLS with firm fixed effects	Poisson	Negative binomial	Negative binomial (5-year lag)
No. of basic research articles (JCRBas = 1)	0.08 (0.06)	0.00 (0.05)	0.0008* (0.0003)	−0.011 (0.015)	
No. of applied research articles (JCRBas = 0)	0.17* (0.03)	0.15* (0.03)	0.003* (0.0002)	0.005* (0.0009)	
No. of basic research articles (5-year lag)					−0.013 (0.020)
No. of applied research articles (5-year lag)					0.005* (0.001)
Year	0.7 (0.3)	1.2* (0.2)	0.02* (0.001)	0.02* (0.007)	0.05* (0.01)
Firm fixed effects		Significant			
Regression statistics					
Adj. R-squared	0.34	0.78			
Log-likelihood			−6610	−2582	−1839
Overdispersion for negative binomial (α)				0.54*	0.52*
N	571	571	571	571	395

Standard errors are shown in parentheses.
* Significant at the 5% level.

multicollinearity or other factors. The VIF values for basic and applied research are 5.9 and 6.2, respectively, which are not particularly low or high.

Table 7 shows that the results are qualitatively similar using other specifications: OLS with firm fixed-effects, Poisson and negative binomial (columns 3 through 5). The rightmost columns of Tables 6 and 7 present the results of using negative binomial regression with 5-year lags. As can be seen from these tables, the use of lagged variables does not significantly change the results. Other lag structures were experimented with, and the tables are available upon request. Lags of 1, 2 and 5 years do not qualitatively change the results. To summarize, the regression analysis reveals a strong positive relationship between applied research and innovation in both industries. The magnitude of β_2 is greater than β_1, suggesting that applied research has a stronger link to innovation than basic research. In the pharmaceutical industry, there is a positive relationship between basic research and innovation, but it is not statistically significant in most of the regression models. In the semiconductor industry, the relationship between basic research and innovation is found to be negative and significant, which is surprising. Admittedly, these simple regressions do not take into account possible endogeneity and serial correlation. Nonetheless, they do present interesting findings and suggest that further investigation is necessary.

5. Ellison–Glaeser index of relative concentration

Scatter plots of patents versus publications provide some intuition for the results found above. Fig. 2 shows the number of patents plotted against basic research publications for each firm in the semiconductor industry. As expected, IBM and AT&T each produce a large number of basic research articles, as well as a large number of patents. However, a surprising number of other companies appear close to the vertical axis: firms like Motorola, Toshiba, Fujitsu and Texas Instruments publish few basic research articles but produce many patents. This is surprising if one were to imagine a strong link between basic research and innovation. Perhaps most striking is the case of Motorola, which published *only* 54 relevant basic research articles between 1981 and 1997, but produced almost as many patents as IBM (which published almost 6000 relevant basic research articles). Fig. 3 shows the number of patents plotted against basic research publications for the pharmaceutical industry. Unlike in Fig. 2, the points in Fig. 3 are generally scattered along a diagonal, so the relationship between innovation and basic research seems stronger for the pharmaceutical industry than for the semiconductor industry.

Figs. 4 and 5 are the scatter plots for applied research. They show the number of patents per firm plotted against the number of applied research articles per firm. In the pharmaceutical industry, there appears to be a strong positive relationship between applied research and innovation (Fig. 5). For semiconductors (Fig. 4), a positive relationship generally exists, but the data points are not as closely correlated as in Fig. 5.

An alternative way of stating the results for the semiconductor industry is to say that basic research activity is highly concentrated in a small number of firms (IBM, AT&T), while innovation is widespread. This is a statement about the concentration of one variable (basic research) *relative* to another (innovation). Additional analysis shows that in any given year, the four semiconductor firms with the largest number of basic research publications together account for 60–80% of basic research publications in the industry, but they produce only 30% of the patents in that industry. By comparison, the top four pharmaceutical firms account for 30–50% of basic research and roughly the same proportion of patents.

The intuition obtained from the scatter plots can be statistically tested using a technique developed by Ellison and Glaeser (1997). The E–G index tells us whether one variable is more concentrated than another. Appendix B introduces the formal model, shows how it is applied in this paper, and describes why it is better to use the E–G index than to simply compare

Table 8
Concentration of publications relative to patents (1981–1997)

E–G gamma	Semiconductors	Pharmaceuticals
γ_B: Concentration of *basic* research publications relative to patents	0.13	0.02
γ_A: Concentration of *applied* research publications relative to patents	0.04	0.01

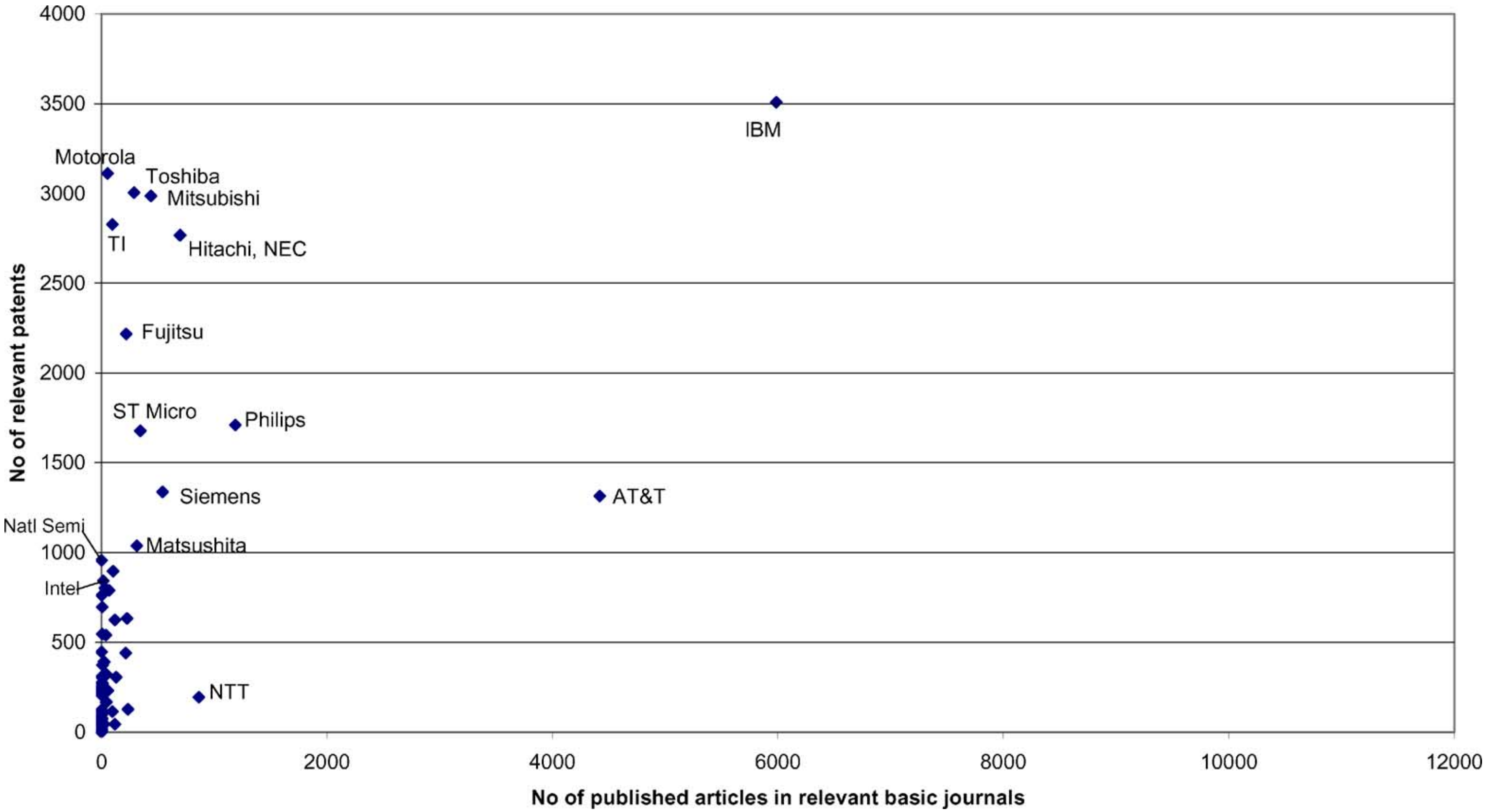


Fig. 2. Semiconductor firms: relevant patents vs. basic research publications, 1981–1997.

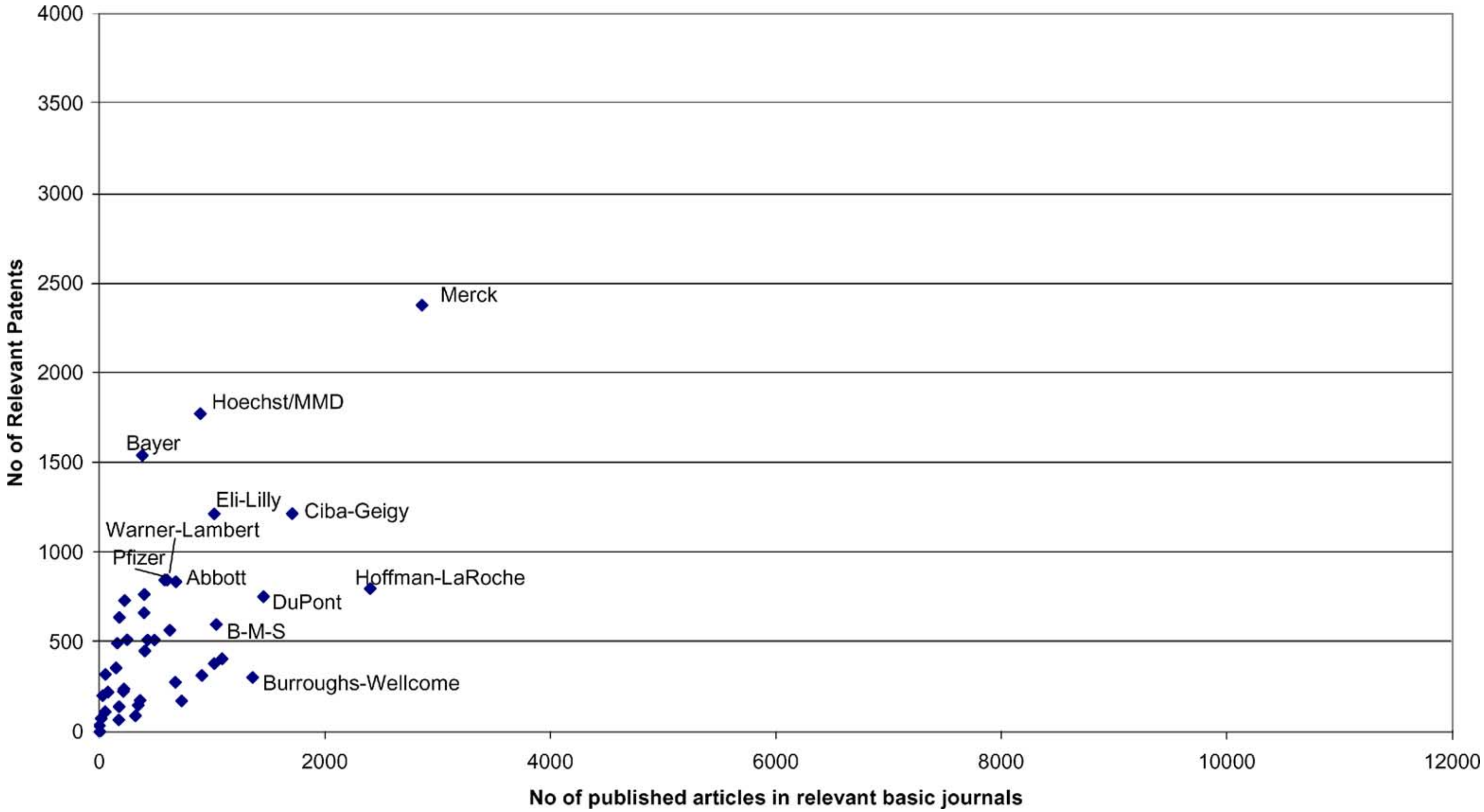


Fig. 3. Pharmaceutical firms: relevant patents vs. basic research publications, 1981–1997.

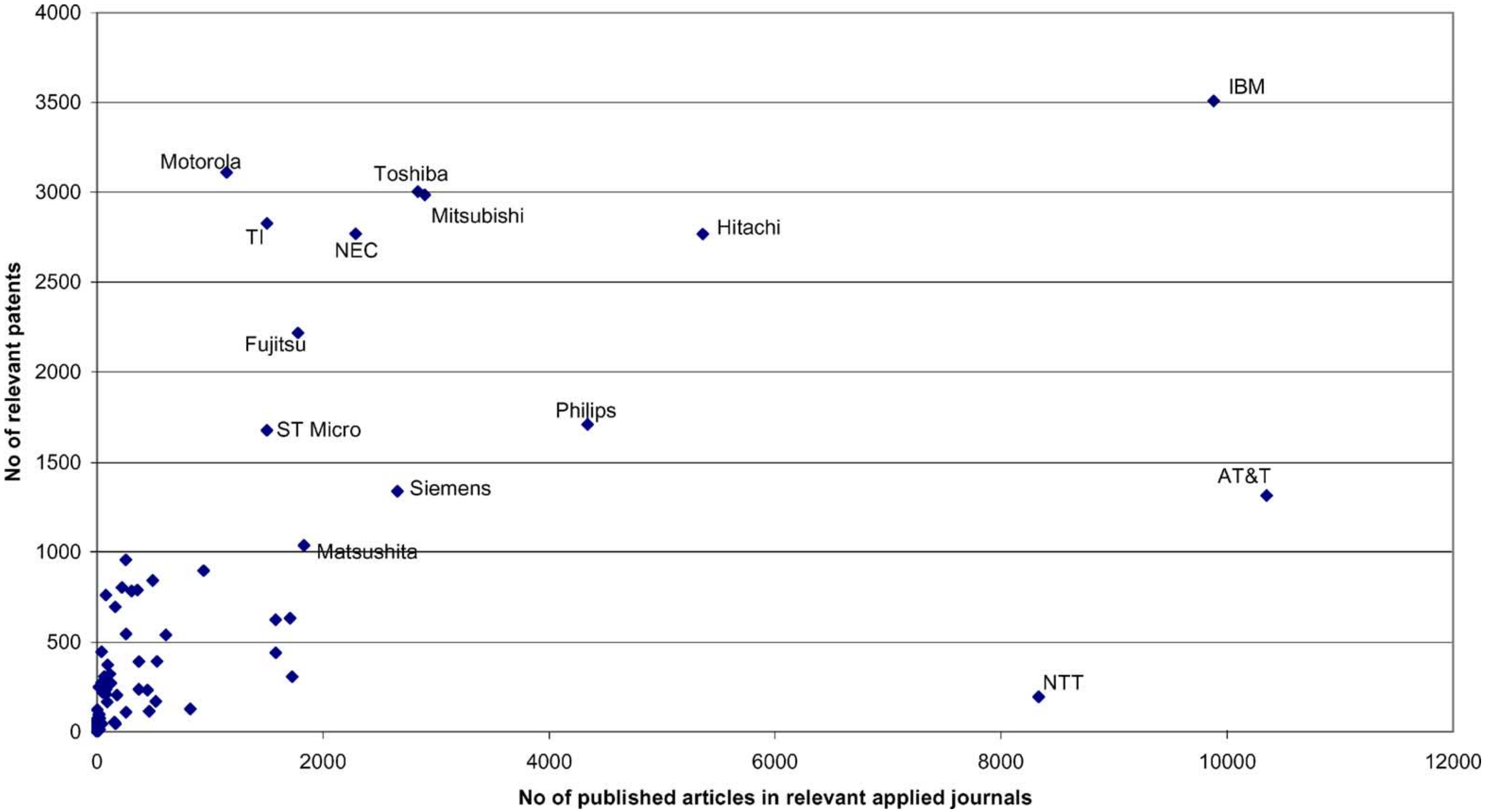


Fig. 4. Semiconductor firms: relevant patents vs. applied research publications, 1981–1997.

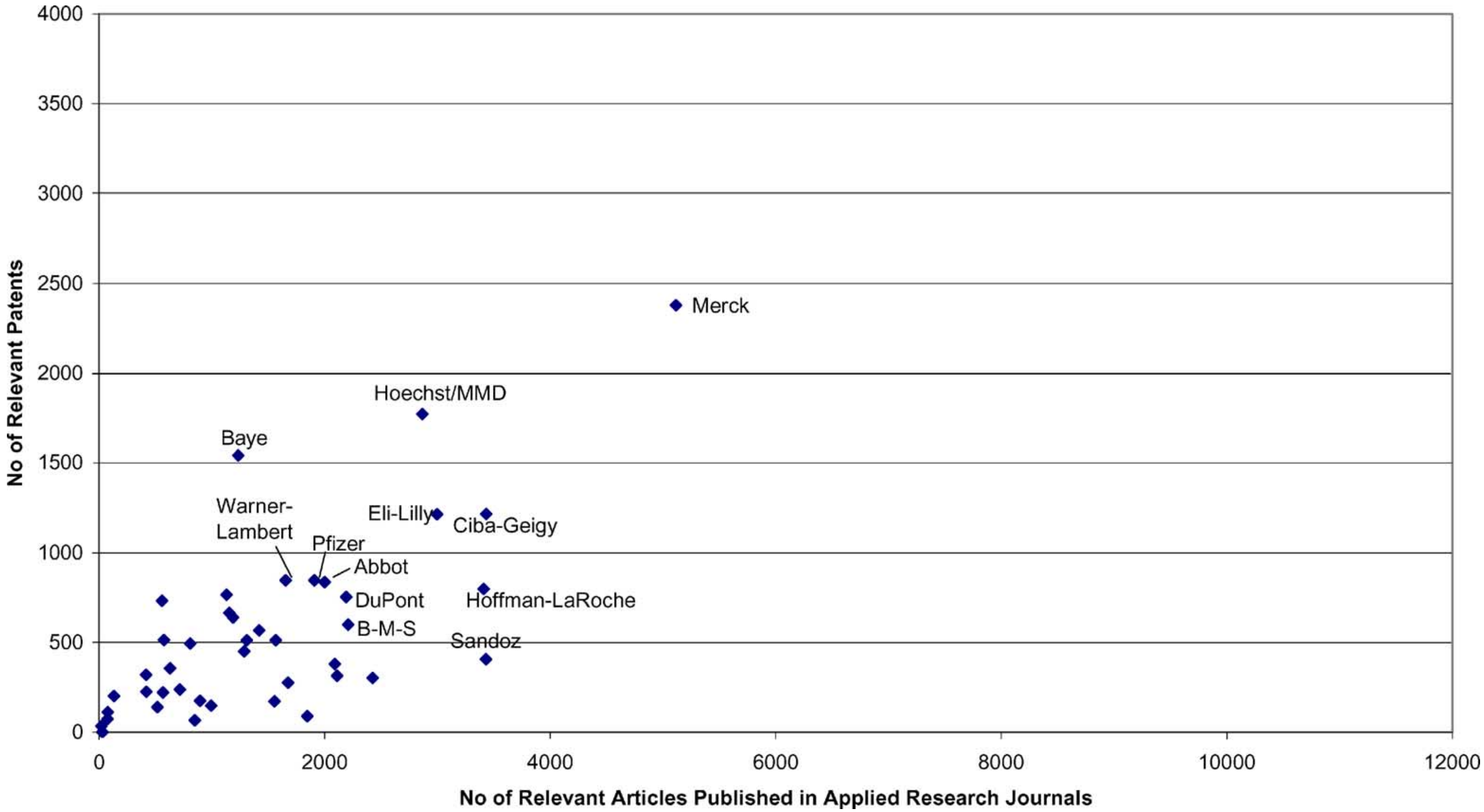


Fig. 5. Pharmaceutical firms: relevant patents vs. applied research articles, 1981–1997.

Table 9
Hypothetical scenarios: gamma of basic research relative to the relevant patents (1981–1997)

Scenario	Semiconductors		Pharmaceuticals	
	γ_B	γ_A	γ_B	γ_A
The top THREE firms publish an equal number of basic or applied research articles; the remaining firms publish none.	0.29	0.32	0.26	0.27
The top FOUR firms publish an equal number of basic or applied research articles; the remaining firms publish none.	0.23	0.22	0.19	0.20
The top FIVE firms publish an equal number of basic or applied research articles; the remaining firms publish none.	0.17	0.17	0.15	0.14
The top SIX firms publish an equal number of basic or applied research articles; the remaining firms publish none.	0.12	0.12	0.13	0.10
The top TEN firms publish an equal number of basic or applied research articles; the remaining firms publish none.	0.06	0.05	0.07	0.06
Research articles are uniformly distributed among all firms in the sample.	0.03	0.03	0.02	0.02

Note. In each case, the number of patents awarded to each firm between 1981 and 1997 was used for computing gamma.

scatter plots or Herfindahl indices. For each industry, I compute a separate index for the concentration of basic research relative to innovation (γ_B), and a similar index for applied research (γ_A). Table 8 presents the values of γ_B and γ_A (a larger value of γ implies a higher concentration of research relative to innovation). For semiconductors, γ_B is around 0.13, while for pharmaceuticals, it is around 0.02. In order to provide meaningful interpretations of these numbers, I calculate the E–G index for several hypothetical scenarios (see Table 9). In each scenario, the number of patents per firm is kept constant at its actual value, while the number of research publications is varied. This changes the concentration of research articles relative to patents, each time producing a different value of gamma.[15] For example, suppose that IBM, AT&T and Philips (the three firms with the largest number of basic research articles in the semiconductor industry) each produces an equal amount of basic research, while all other firms in that industry do not produce any basic research, then $\gamma_B = 0.29$ (see Table 9).

For the semiconductor industry, the actual value for γ_B was computed to be around 0.13, which is similar to the hypothetical scenario in which basic research is performed *only* by the top 5 or 6 firms. For pharmaceuticals, however, the concentration of basic research relative to innovation is around 0.02, which is no different than the scenario in which it is uniformly distributed among all firms. In both industries, applied research is not concentrated relative to innovation: both values for γ_A obtained in Table 8 are quite similar to the scenario in which applied research is uniformly spread out across firms in Table 9. These findings provide rigorous support for the intuition derived earlier from the scatter plots.

5.1. Robustness tests and patent-to-patent citations

One source of concern is the manner in which journals were classified as “basic” and “applied”. Appendix C explores three other classification schemes (including one developed by an company specializing in such work) and shows that the results are robust to different classification schemes.

For the pharmaceutical industry, I performed an additional test to explore the possibility that our definition of “basic” research is overly generous. Perhaps basic research is not found to be concentrated because many applied journals are inadvertently included into the “basic” category. To address this issue, I calculated γ_B using only articles published between 1981 and 1997 in three leading basic scientific journals: *Nature*, *Science*, and *Cell*. These journals are considered “basic” by all four classification schemes mentioned in the preceding paragraph. Even if we consider only these three journals, a value of $\gamma_B = 0.06$ is obtained, which is not particularly high compared to the hypothetical scenarios shown in Table 9. It appears safe to conclude, therefore, that basic

[15] In Eq. (B.2) of Appendix B, x_j is kept constant while s_{lj} is varied, thus yielding values for γ_l.

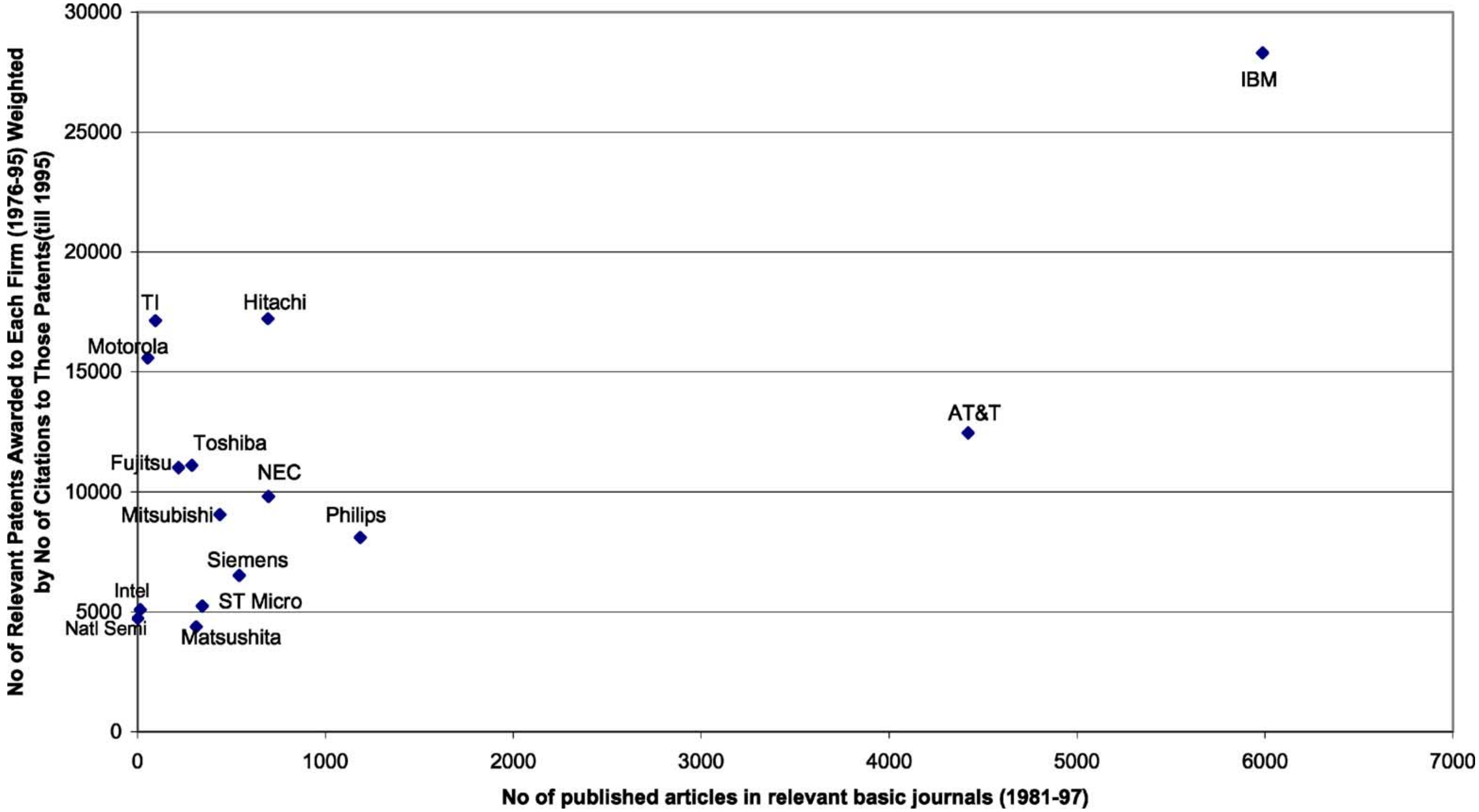


Fig. 6. Semiconductor firms: relevant patents weighted by citations vs. basic publications.

pharmaceutical research is indeed not concentrated relative to innovation.

For the semiconductor industry, there is a legitimate concern that the omission of patent classes from Table 1 may be driving the results. In other words, the top firms may have received “too few” semiconductor patents relative to basic research because their basic research led to patents in other areas. It turns out, though, that the results are extremely robust to the inclusion of a broad range of patent classes. Even if we include patents from *every single patent class* awarded to each firm, the concentration of basic research relative to innovation remains practically unchanged at $\gamma_B = 0.13$.

Another possible explanation for the high concentration of basic research relative to innovation in the semiconductor industry is that patent counts provide an unsatisfactory measure of innovation output. The value of patents is highly skewed (Trajtenberg, 1990; Harhoff et al., 1999), so a firm may possess many patents but they may not be valuable. Perhaps, IBM’s and AT&T’s patents are truly valuable and those of other firms, much less so? Following a stream of research pioneered by Trajtenberg (1990), I address this issue by using patent-to-patent citations. The number of times a patent is cited by other patents provides a quality-adjusted measure of innovation output (Hall et al., 2000). As this analysis is highly data-intensive, it is restricted to the fifteen semiconductor firms in the sample having the largest number of patents.[16] I identify U.S. patents issued to each firm within relevant patent classes between 1976 and 1995. For each of these patents, I count the number of citations made up to and including 1995 by all *other* patents ever awarded by the USPTO.

As shown in Fig. 6, the results appear less stark as those using simple patent counts (shown earlier in Fig. 2). Nonetheless, the earlier findings persist: firms other than AT&T and IBM produce a surprising number of citation-weighted patents, given their level of basic research. For example, Fig. 6 shows that TI, Motorola and Hitachi produce citation-weighted patents in excess of AT&T and over half that of IBM even though their output of basic research is very low compared to those firms. Fujitsu, Toshiba and NEC produce almost as many innovations as AT&T using this measure, even though they produce ten times less basic research than AT&T. For the 15 firms, the E–G index for basic research relative to citation-weighted patents was found to be 0.14. This is less than the value of $\gamma_B = 0.19$ using simple patent counts (Table 9), so there is some truth to the claim that IBM and AT&T produce patents of higher quality than other firms. However, the newly computed E–G index of 0.14 is still pretty high, and corresponds to a scenario where basic research is performed only by the top five firms.

A separate argument can be made that patents are not a good indicator of economic value, and that IBM and AT&T derive economic benefit from their inventions in ways other than patents. For example, there may be economies of scope, or perhaps licensing agreements allow them to capture some of the value of their basic ideas. As a partial way of addressing this concern, I used the cumulative net profit of each firm as an alternative to patent counts. Revenue from licensing and other sources should improve a firm’s net profits. Data from *CompuStat* are available for 43 major U.S. firms in the sample. For these firms, the concentration of basic research relative to relevant patents is $\gamma_B = 0.33$, and that of basic research relative to *cumulative net profits* is $\gamma = 0.31$. This means that basic research is highly concentrated relative to cumulative net profits in the semiconductor industry. The reason is straightforward: among U.S. semiconductor companies during the period examined, IBM and AT&T account for 58% of basic research articles but only 32% of net profits. Apart from them, Intel, Hewlett Packard, and Motorola also captured sizeable portions of industry profits. The use of cumulative net profits does not represent a carefully controlled test, but if basic research does indeed generate benefits to a firm beyond patents, IBM and AT&T should have captured a larger share of the industry profits than they actually did.

6. Discussion

The high concentration of basic research within two companies (IBM and AT&T) reflects the existence of the Watson and Bell laboratories, and the absence of

[16] The firms are: IBM, Motorola, Toshiba, Mitsubishi, Texas Instruments, NEC, Hitachi, Fujitsu, Philips, STMicroelectronics, Siemens, Lucent (AT&T), Matsushita, National Semiconductors and Intel.

equally prominent laboratories at the other semiconductor firms. Given the low levels of basic research performed by those other firms, it is surprising that they produce so many innovations.[17] Many possible explanations exist for this pattern, but my interpretation is that semiconductor firms rely on different R&D strategies: some firms like IBM and AT&T benefit from investing heavily in *both* basic and applied research, while other firms rely primarily on *applied* rather than basic research. I provide evidence to support this explanation in Section 6.1. Other explanations are discussed in Section 6.2. Finally, the limitations of this paper are described in Section 6.3.

6.1. Different R&D strategies and absorptive capacity

I propose a simple explanation for the empirical results: IBM and AT&T produce innovations that rely on basic and applied research, and so they invest in both kinds of research. Other semiconductor firms produce innovations that rely mainly upon *applied* research, thus they do not perform much basic research. In order to provide evidence for this hypothesis, I analyze citations made by each firm's patents to the scientific literature. Such citations can reveal the knowledge base and scientific community upon which a firm depends for its innovations (e.g. see Gittelman and Kogut, 2003). The evidence provided here is suggestive rather than conclusive, and further research is necessary to corroborate my interpretation.

Performing patent-to-science citations is highly laborious, and so I only performed it for the 15 semiconductor firms having the most patents. Between 1981 and 1997, these firms produced 49,121 relevant applied publications, 15,294 relevant basic publications and 32,076 relevant patents. I analyzed the citations contained within these patents to the top 30 basic journals and the top 30 applied journals. Only a small number of journals are included, but this does not pose a problem because they constitute the core journals in the field of semiconductor research. According to experts in bibliography, the core journals within a scientific field account for the lion's share of all citations; this effect is known as Zipf's Law (Garfield, 1980).

My analysis reveals that the patents of these semiconductor firms make fifteen times as many citations to applied research journals as they do to basic research journals (they make only 955 citations to articles in the top 30 basic research journals, but a staggering 15,188 citations to articles in the top 30 applied journals). This suggests a much stronger dependence of these patents on applied research than on basic research. I performed a similar analysis for all of the pharmaceutical firms in the sample, and found the reverse to be true. Contrary to semiconductor firms, the patents of pharmaceutical firms make more citations to basic research journals than they do to applied journals (they make 17,710 citations to the top 30 basic journals, but only 15,187 citations to the top 30 applied journals). Thus, it appears that semiconductor firms rely a great deal more on applied than on basic research, while pharmaceutical firms rely slightly more on basic than on applied research.

An even more interesting pattern emerges if we examine the patent-to-science citations for each firm. Figs. 7 and 8 show the relationship between a firm's propensity to cite basic research and the intensity of its internal basic research. I measure the former as the ratio of citations to basic versus applied journals, and I measure the latter as the ratio of relevant publications in basic versus applied journals. On both charts, a positive slope is apparent. I interpret this to mean that firms that focus more on basic research are better able to absorb basic scientific knowledge, while firms that focus more on applied research are better able to absorb applied knowledge. For example, Fig. 7 shows that semiconductor firms such as IBM and AT&T emphasize in-house basic research and are also more likely to cite basic research journals; other firms such as Motorola, Intel and Mitsubishi have a lower intensity of basic research and are also less likely to cite basic research journals. This lends support to the

[17] It is tempting to argue that the results are driven by IBM and AT&T. However, it is precisely the contrast between IBM and ATT on the one hand, and the remaining firms on the other hand, that is interesting. Even if we remove IBM and AT&T from the sample, it is still surprising that the remaining firms like Motorola produce practically no basic research, but obtain a great number of patents. However, dropping AT&T and IBM when calculating the E–G index would not be correct, since it is not possible to show the contrast between them and the other firms if they were removed.

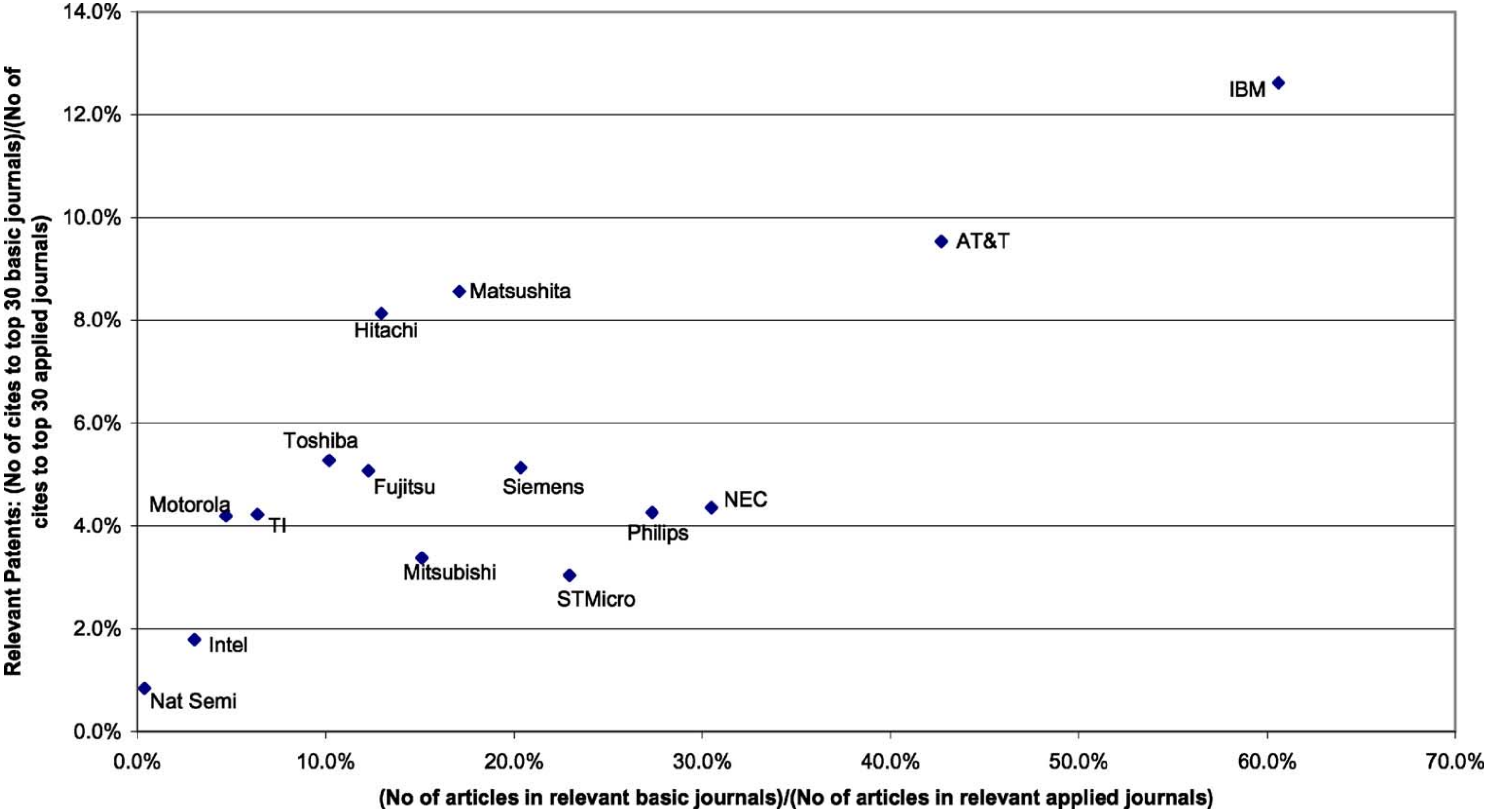
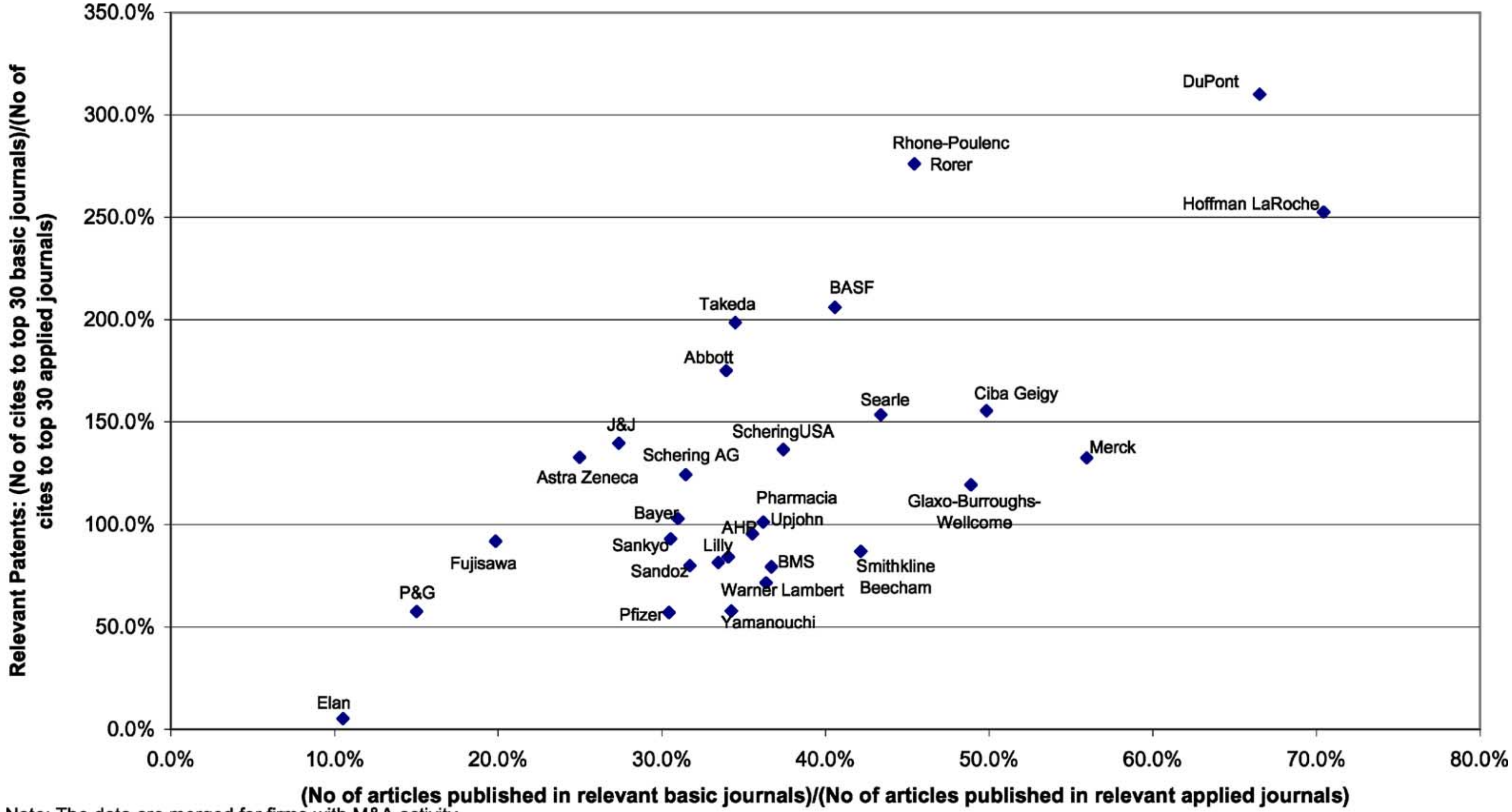


Fig. 7. Semiconductor firms: citations by relevant patents to top scientific journals, 1981–1997.



Note: The data are merged for firms with M&A activity.

Fig. 8. Pharmaceutical firms: citations by relevant patents to top scientific journals, 1981–1997.

concept of absorptive introduced by Cohen and Levinthal (1990). Yet, the story seems more complex than they had imagined. Within each industry, firms vary tremendously in their basic research intensity (they appear over a broad range along the *x*-axis in Figs. 7 and 8). Thus, different strategies appear to exist, with some firms gravitating towards basic science (IBM and AT&T in semiconductors; DuPont and Merck in pharmaceuticals) and other firms favoring applied research (Motorola and Intel in semiconductors; Elan and Fujisawa in pharmaceuticals). This does not mean that basic knowledge is unimportant for firms that focus on applied research. Instead, they probably absorb basic knowledge after it had been codified and translated into applied knowledge through the efforts of universities, research consortia and other firms (Lim, 2000).

A separate idea that emerges from these charts is that different kinds of absorptive capacity may be needed across industries. The vertical axes in Figs. 7 and 8 have vastly different scales, and even IBM and AT&T, the semiconductor firms with the greatest dependence on basic research, have far lower propensities to cite basic research journals than all of the pharmaceutical firms, save one. Hence, the type of absorptive capacity relevant to an industry (along with the mechanisms for developing it) may be sensitive to the competitive and industry-specific context surrounding a firm.

Since semiconductor patents depend primarily upon applied research, why did IBM and AT&T even bother to invest in basic research, which is a costly and uncertain activity? As mentioned in Section 2, IBM and AT&T are probably the only semiconductor firms large and diversified enough during this period to truly gain the benefits of investing in basic research. Furthermore, during the period of study, they held dominant positions in the computer and telecommunications markets respectively. The rents they derived from these markets *may* have given them the financial resources required to invest in basic research. The data provides some indirect evidence of this. Between 1981 and 1997, γ_B declined from 0.26 to 0.07, largely driven a reduction in basic research publications by IBM and AT&T. This decline coincided with financial difficulties faced by both firms. IBM, which led basic research for semiconductors, faced difficulties in the early 1990s and drastically reoriented its R&D organization towards applied science and development.[18] Fig. 9 shows that IBM's output of research articles began to dip shortly after its net income became negative in 1991. It slashed by more than half the number of basic research articles it published from 520 in 1992 to 206 in 1997. Similarly, AT&T faced the end of a long monopoly and reoriented Bell Laboratories, which was eventually spun off as part of Lucent Technologies in 1996.[19] The number of basic research publications it produced declined from its peak of 391 in 1990 to 193 in 1997.

6.2. Other possible explanations

While the explanation presented above is consistent with the data, other explanations exist and should be explored in future research.[20] One possibility is that firms like Motorola, Texas Instruments, and Toshiba are simply more productive than IBM and AT&T at translating their basic research into a large number of patents. Yet, this is hard to believe because of the *low absolute number* of basic research articles published by those firms. Moreover, the highly uncertain nature of basic research should favor IBM and AT&T rather than the other firms, since IBM and AT&T are more likely than them to have the scale and scope needed to translate the results of basic research into patentable inventions.

A second possibility is that IBM and AT&T refrained from patenting their inventions because they relied instead on trade secrets. While possible, this explanation also carries doubt: if secrecy were so important, why then are IBM and AT&T publishing their

[18] "In 1992, when newly appointed chairman Louis Gerstner began downsizing IBM, the research division was included and saw its budget cut by a third, from US$ 6.5 billion in 1992 to a low of US$ 4.3 billion in 1994. It has since bounced back to US$ 6 billion in 1995 as more of its work became product oriented . . . " ("IBM Reconnects Research", *Electronics Business Today*, September 1996). See also "Into the Big Blue Yonder," *Technology Review*, 1999, 102 (4) 46–53; and "R&D Gets Real," *Electronic Business Today*, October 1997.

[19] In "Lucent's Ascend," *BusinessWeek*, February 8, 1999, the firm's CEO describes how he linked Bell Lab's research budget directly to revenue growth.

[20] I thank the reviewers at *Research Policy* for pointing out many of these explanations.

Fig. 9. IBM-net income and number of articles published in relevant journals.

research on such a large scale? Furthermore, patents play an important role as bargaining chips in the semiconductor industry (Hall and Ziedonis, 2001), so it would be unwise for IBM and AT&T to resort to secrecy only to find out that another firm had received patents on similar inventions.

A third explanation arises because patents cannot be issued for ideas that have been published. So, perhaps companies like Motorola chose to patent their ideas rather than to publish them. This explanation is somewhat improbable: the tension between publishing and patenting is more likely to arise for applied than for basic research, since the latter are more likely to result in practical and patentable inventions. Yet, we observe that Motorola, TI and other firms publish extensively in applied research journals. Had they really wanted to suppress publications in order to receive patents, they would have done so for applied rather than basic research. In that case, we should have observed them publishing a large number of basic research articles and few applied articles, instead of the other way around. It is more probable that they failed to publish much basic research simply because they lacked the internal capabilities present at the AT&T Bell Laboratories and IBM. Besides, various approaches exist for resolving the conflict between patenting and publishing, and these are widely used in practice. One method is to delay the publication of a paper until a patent application is awarded; another method involves filing a provisional patent application with the USPTO.

Another possible issue is that the antitrust issues surrounding IBM and AT&T could have led them to publish a greater number of their basic research results, in order to avoid government scrutiny. Or perhaps the social norms within the semiconductor industry are different from those in pharmaceuticals, so that researchers value patents in the former but publications in the latter. These are interesting ideas that should be examined in future research.

6.3. Limitations

A key limitation of this study is that time lags between research and innovation are only superficially explored. Previous work has shown a high variance in the time before basic research comes to fruition (e.g. Adams, 1990). As discussed earlier, introducing dependent variables that are lagged by 1, 2 and 5 years does not significantly change the results. However, more sophisticated lag structures need to be examined.

A second limitation is the inclusion of U.S. patents only. Omitting European and Japanese patents may bias the sample. Fortunately, this risk is low because the United States is the largest market for both semiconductor and pharmaceutical products, so important European and Japanese patents are also filed in America.

A third is the use of publication data. While I have attempted to validate its use by showing that it correlates with R&D expenditure, it remains possible that neither is a good measure of research performed by firms. Another issue with publication data is that the SCI is biased towards English-language publications. Nevertheless, the results persist when all non-U.S. firms are dropped from the sample.

Despite the limitations, there emerges a robust empirical regularity: basic research is surprisingly concentrated relative to innovation in the semiconductor industry.

7. Conclusion

This paper's main contribution is a careful measurement of the relationship between basic research, applied research and innovation in two technology-intensive industries. While basic research mirrors the distribution of patents rather closely in the pharmaceutical industry, it is much more concentrated than patents in the semiconductor industry. In both industries, the concentration of applied research closely matches that of innovation output. The results are robust to the use of alternative definitions for "basic" research and the use of citations to adjust for the quality of each patent.

Further research is required to disentangle various possible explanations for why basic research is so highly concentrated relative to innovation in the semiconductor industry. At this stage, I provide circumstantial evidence that innovation in the semiconductor industry depends mainly on applied research. Therefore, firms are not precluded from innovating even if they perform very little basic research. I also suggest that different kinds of absorptive capacity exist, with IBM and AT&T absorbing both applied and basic knowledge and other firms mainly absorbing applied knowledge.

Many questions remain unanswered and point to future research opportunities. Why do IBM and AT&T perform so much basic research? How do the results look like in yet other industries? When is absorbing basic research necessary and when is it not? To answer these questions, we need to better understand the process by which firms produce and absorb knowledge.

Acknowledgements

I am indebted to Scott Stern, Rebecca Henderson and Eric von Hippel for their guidance. I thank Glenn Ellison, Iain Cockburn, Richard Nelson, Ashish Arora, Rosemarie Ziedonis, Julia Liebeskind, Jeff Furman, Peter Murmann, Pierre Azoulay, David Hsu, Jaeyong Song, Sam Kortum, Jorgen Pedersen and two anonymous referees at Research Policy for their comments. I also thank the participants of the CCC Doctoral Colloquium 1999, Academy of Management 2001, NBER Productivity Lunch, Wharton Technology Miniconference 2001 and MIT Innovation Seminar. Adam Jaffe generously provided data on patent-to-patent citations and Diana Hicks on the 'basicness' of each journal. All errors and omissions are my own. This paper would not have been possible without the encouragement of my wife, Seto Wai Ling.

Appendix A. Construction of the dataset

The companies in the sample are listed in the table below. I compiled a list of 297 semiconductor firms from reports by the Integrated Circuit Engineering (ICE) Corporation, Semiconductor Industry Association, *Electronics Business* and other sources. Seventy-three "fabless" semiconductor companies were eliminated because they do not manufacture their own semiconductor chips.[21] Another 135 companies with less than 15 publications per year and less than 100 patents between 1985 and 1995 were also dropped. Each of these firms represents less than 0.2% all publications and 0.1% of all patents during the time period, therefore, contributing insignificantly to any concentration index. Six other companies were dropped because they compete primarily in other lines of business, but happen to operate semiconductor-manufacturing facilities (Honda, Nissan, Ford, Toyota, Nippon Steel, and Kawasaki Steel). Including these companies in the analysis does not change the results because they contribute only a small number of relevant patents and publications ($\gamma_B = 0.12$, $\gamma_A = 0.04$).

For the pharmaceutical sample, the list of firms used by Cockburn and Henderson (1998) was used as a starting point. To this list, I added four firms with significant numbers of relevant publications or patents (BASF, Bayer, Astra-Zeneca, and DuPont). I identified these firms by analyzing patent data from the U.S. Patent Office and publication data from the Science Citation Index.

For each industry, a list of major subsidiaries, mergers, and acquisitions was painstakingly constructed from public sources.[22] This list was used to combine the patents and publications of subsidiaries with the parent company. Majority-owned subsidiaries and acquisitions are considered part of the parent company with effect from the year the transaction is completed. Merged companies are treated as new entities from the year the merger takes effect.

[21] According to the Fabless Semiconductor Association, "Fabless (without fab) refers to the business methodology of outsourcing the manufacturing of silicon wafers, which hundreds of semiconductor companies have adopted. Fabless companies focus on the design, development and marketing of their products and form alliances with silicon wafer manufacturers, or foundries."

[22] Sources include annual reports, company websites, the *Directory of Corporate Affiliations*, *Hoover Company Profiles*, and analyst reports. I thank Celina Lee for helping me to compile these data for the pharmaceutical firms.

Companies included in the sample

Semiconductors		Pharmaceuticals
Acer Labs, Inc.	LSI Logic	Abbott
Actel Corporation	Macronix	American Home Products
Advanced Micro Devices (AMD)	Matsushita Electric Corporation	Astra
Alcatel	Microchip Technology	BASF
Altera	Micron Semiconductor Inc. (subsidiary of Micron Technology)	Bayer
American Microsystems, Inc.	Mitel Semiconductor	Beecham
Analog Devices	Mitsubishi Electric Corporation	Bristol-Myers
Asahi Kasei Microsystems	Mostel-Vitelic	Bristol-Myers-Squibb
AT&T (Lucent)	Motorola, Inc. (Semiconductor Products Sector)	Burroughs-Wellcome
Atmel Corporation	National Semiconductor Corp.	Ciba-Geigy
Brooktree Corporation	NCR Microelectronic Products	DuPont
Burr-Brown	NEC Corporation	ELAN
Canon, Inc.	Newport Wafer Fab Limited	Fujisawa
C-Cube	Northern Telecom (Nortel)	Glaxo
Chartered Semiconductor Manufacturing	Northrop & Northrop Grumman Corp.	Glaxo-Wellcome
Cherry Semiconductor Corp.	NTT (Nippon Telephone & Telegraph)	Hoechst (& Roussel)
Cirrus Logic	Oki Electric industry Co., Ltd.	Hoffman-LaRoche
Cray Computer	Philips	Johnson & Johnson
Cray Research, Inc.	Raytheon Semiconductor Division	Eli Lilly
Cypress Semiconductor, Inc.	Ricoh Co., Ltd.; Electronic Device Division	Marion
Cyrix Corporation	Rockwell International	Marion-Merrell-Dow
Daewoo Electronic Components Co.	Rohm Co., Ltd.	Merck
Dallas Semiconductor	S3 Inc.	Merrell-Dow
Digital Equipment Corporation (DEC-Now Compaq)	Samsung Electronics Company	Procter & Gamble Pharmaceuticals
Ericsson Components A.B.	Sanyo Electric Co.	Pfizer (& Roerig)
ESS Technology	Seiko Epson Corp.	Pharmacia
Fuji Electric Co., Ltd.	Sharp Corporation	Pharmacia & Upjohn
Fujitsu	Siemens	Rhone-Poulenc
GEC-Plessey (acquired by Mitel)	Sony Corporation	Rorer
General Semiconductor	ST Microelectronics (SGS-Thomson)	Sandoz
Grumman (pre 1994)	Symbios	Sankyo
Harris Semiconductor	Taiwan Semiconductor Manufacturing Co., Ltd.	Schering (German)
Hewlett-Packard Company	Tech Semiconductor Singapore Pte. Ltd.	Schering-Plough (USA)
Hitachi	Temic (bought by Vishay; IC division sold to Atmel)	Searle (Monsanto owns it)

Semiconductors		Pharmaceuticals
Honeywell, Inc., Solid State Electronics Center	Texas Instruments	Smithkline
Hughes Aircraft Company (merged with Raytheon)	Toshiba	Smithkline-Beecham
Hyundai Electronic Industries Co., Ltd., Semiconductor Division	United Microelectronics Corporation (UMC)	Squibb
IBM	United Technologies Microelectronics Center	Takeda
Integrated Device Technology (IDT)	VLSI Technology, Inc.	Upjohn
Intel Corporation	Weitek	Warner-Lambert
International Rectifier	Westinghouse, Advanced Technology	Yamanouchi
ITT Semiconductors (ITT Industries)	Winbond	Zeneca
Lattice Semiconductor	Yamaha Corporation	
LG Semicon (Lucky-Goldstar)	Zilog, Inc.	
Linear Technology Corporation		

Appendix B. Ellison–Glaeser index of relative concentration

Ellison and Glaeser (1997) introduced a formal model that allows us to ask whether research activity is concentrated in a small number of firms relative to the amount of innovation they produce. As discussed below, this model controls for several other explanations that might otherwise cause spurious results. The Ellison–Glaeser index tells us whether one variable is much more concentrated than we would expect, using a second variable as a baseline. In their work on the geographic concentration of manufacturing, Ellison and Glaeser examine whether U.S. manufacturing activity is more concentrated geographically than we would expect, given the number of people employed in manufacturing in each region. By analogy, in this paper I examine *whether basic and applied research are more concentrated within a few firms than we would expect, given the number of innovations produced by each firm*. Table B.1 summarizes the variables used in the Ellison–Glaeser model and their corresponding use in Stern and Trajtenberg (1998) and in this paper. For our purposes, the E–G index is preferable to standard concentration indices, such as the four-firm concentration ratio (C_4) and the Herfindahl (H),[23] where

$$H = \sum_i s_i^2, \quad s_i \text{ is firm } i\text{'s share of basic research, applied research or innovation.} \tag{B.1}$$

It is appropriate to use the E–G index instead of the C_4 or the Herfindahl index because research activity and innovation may themselves be concentrated (e.g. because the industry as a whole is concentrated). The important issue is whether research activity is *even more concentrated* than innovation. The E–G index of research relative to innovation output, γ_l, is given by

$$\gamma_l = \frac{\sum_j (s_{lj} - x_j)^2 - \left(1 - \sum_j x_j^2\right)\tilde{H}_l}{\left(1 - \sum_j x_j^2\right)(1 - \tilde{H}_l)}, \tag{B.2}$$

where research area: $l \in$ {basic, applied}; s_{lj}: firm j's share of research (publications) in area l; x_j: firm j's share of innovation (patents); $\tilde{H}_l$: Herfindahl of research papers in area l.

One way to understand this index is to visualize a map containing regions proportional to the size of

[23] The properties of these and other indices are discussed in Curry and George (1983).

Table B.1
Applications of the E–G model

Variable	Ellison–Glaeser (1997)	Stern–Trajtenberg (1998)	This paper
Inputs			
i, j, l	In industry l, each plant i is located in geographic area j	For physician l, each patient i is allocated to drug j	In research area l, each research paper i is published by firm j. Note: $l \in$ {basic, applied}
x_j	Area j's share of total employment	Drug j's market share	Firm j's share of patents
$s_{lj} = \mathrm{Sum}(z_{il}u_{ij})$	Area j's share of employment in industry l	Drug j's share of the prescriptions by physician l	Firm j's share of research paper in area l
z_{il}	Share of firm (plant) i's employment in industry l	Share of patient i's visits seen by physician l	Paper i's share of basic and applied research papers (area l)
$\tilde{H}_l = \sum_i^{N_l} z_{il}^2$	Herfindahl index of plant size in industry l	Herfindahl index of physicians (l) in terms of their patients	Herfindahl index of research papers in area (l)
Output			
γ_l	Excess geographic concentration in industry l	Excess concentration of drug prescription by physician l	Excess concentration of research over innovation in area l

Note. u_{ij} is an indicator variable set to 1 if research paper i is published by firm j.

each firm's innovative output. Now, imagine a person randomly throwing darts—each representing a basic research article—at the map. The E–G index for basic research (γ_B) tells us whether basic research articles are more concentrated than we would expect from this random process. A similar map may be drawn for applied research to compute the index for applied research (γ_A).

The Ellison–Glaeser index overcomes spurious results that may arise if we simply compare Herfindahl indices, C_4 ratios, or scatter-plots. These comparisons are performed at an aggregate level, whereas the Ellison–Glaeser index compares each firm's research and innovation *pair-wise*. This is because γ_l is a function of $(s_{lj}-x_j)^2$. Further, it accounts for the possibility that there may be too few articles in a given research area. For instance, suppose there were only a handful of basic research articles, so that basic research itself is concentrated. The E–G index discounts this by including $\tilde{H}_l$ into Eq. (B.2). Finally, the index accounts for the possibility that innovation may be concentrated in only a few firms, by incorporating $\sum_j x_j^2$ into the equation. Thus, if only two firms produced all the innovations in an industry, $\sum_j x_j^2$ would be high. Conditional on the research level of each firm, this would cause γ_l to be high. The intuition is that it would be surprising to observe other firms performing research since only two are innovating.

A simpler version of the Ellison–Glaeser index can be used if a research area contains many articles. Define N_l to be the total number of scientific publications in research area l by all the firms. Article i's share of research in area l is given by $z_{il} = 1/N_l$. Therefore, $\tilde{H}_l = \sum_i^{N_l} z_{il}^2 = 1/N_l$. If there are many papers in research area l, then $\tilde{H}_l \to 0$, and we obtain a simplified expression (see Stern and Trajtenberg, 1998, footnote 11):

$$\text{As } N_l \to \infty, \quad \gamma_l \to \frac{\sum_j (s_{ij} - x_j)^2}{\left(1 - \sum_j x_j^2\right)}$$

In this case, the concentration of basic and applied research with respect to innovation are given by:

Basic research relative to innovation:

$$\gamma_B \to \frac{\sum_j (s_{Bj} - x_j)^2}{\left(1 - \sum_j x_j^2\right)} \tag{B.3a}$$

and

Applied research relative to innovation:

$$\gamma_A \to \frac{\sum_j (s_{Aj} - x_j)^2}{\left(1 - \sum_j x_j^2\right)} \tag{B.3b}$$

Table C.1
Descriptive analysis of all journals listed in the Science Citation Index

Variable	N	Minimum	Maximum	Average	S.D.	Percentiles		
						25	50	75
IPF85	3421	0	39.7	1.27	1.85	0.41	0.79	1.5
IPF97	3826	0	40.8	1.51	2.44	0.46	0.91	1.7
Acad85	3205	This variable is either 0 or 1		0.79	0.19	0.73	0.84	0.92
Acad90	3043	This variable is either 0 or 1		0.82	0.16	0.77	0.86	0.93
Acad97	3293	This variable is either 0 or 1		0.85	0.14	0.80	0.88	0.93
JCRBas	4901	This variable is either 0 or 1		0.18	0.38	NA	NA	NA
CHIbas	4901	This variable takes values from 0 to 4		2.75	1.15	2	3	4

Note. Numerical suffixes represent the year. For example, IPF85 is the Science Citation Index Impact Factor for a journal in 1985.

These simplified equations will be estimated because the sample contains a large number of basic and applied research publications (these numbers are reported in footnote 12).

Appendix C. Robustness of the results to journal classification schemes

In this appendix, I explore the robustness of the results to alternative ways of classifying journals. Four classification schemes are implemented using the indicator variables *JCRBas*, *CHIbas*, *HiIPF* and *HiAcad*. The first two variables attempt to measure “basic” research as that which seeks a fundamental scientific understanding of phenomena. The third and fourth variables expand beyond basic research to look more generally at highly cited research and the types of research typically performed at academic institutions. Summary statistics of these four variables are shown in Table C.1.

The first variable, *JCRBas*, is based on a classification scheme I developed by using the SCI Journal Citation Report to identify “basic” and “applied” journal categories for each industry. The second variable, *CHIbas*, uses the classification scheme developed by CHI Research, Inc. CHI awards each journal a score from zero to four. For the physical sciences, levels 1 through 4 correspond to applied technology, engineering sciences, applied research, and basic research, respectively. For the biomedical sciences, they correspond to clinical observation, clinical mix, clinical investigation and basic science (see Hicks, 1996, for more details). In this thesis, I define a research article as “basic” if it is published in a journal with $CHIbas = 4$.

The third variable, *HiIPF*, identifies research articles that are published in highly cited journals. The *HiIPF* variable is based on each journal’s SCI Impact Factor (*IPF*), which is published in the Journal Citation Reports that accompany each year’s edition of the *Science Citation Index*.[24] The Impact Factor for journal k in year y is given by:

$$\mathrm{IPF}_{ky} = \frac{\text{No. of citations in year } y \text{ to articles published in journal } k \text{ in years } (y-1) \text{ and } (y-2)}{\text{No. of articles in journals } k \text{ in years } (y-1) \text{ and } (y-2)} \quad \text{(C.1)}$$

As shown in Table C.1, the distribution of *IPF* is highly skewed. A small number of journals have high impact scores, reaching up to 40.8. However, 75% of the journals in a given year have impact scores less than 1.7. This is consistent with the bibliographic literature on Zipf’s law and Bradford’s law, which states that only a small set of core journals in a scientific discipline are highly cited (Garfield, 1980). I define a journal to be *highly cited* if it falls within the top 25% of this distribution, specifically those with *IPF* scores exceeding 1.7.

$$\text{In year } yy, \text{ define } \mathrm{HiIPF}_{yy} \equiv \begin{cases} 1 & \text{if } \mathrm{IPF}_{yy} > 1.7 \\ 0 & \text{otherwise} \end{cases} \quad \text{(C.2a)}$$

[24] ISI does not publish the SCI scores for every journal each year, but all the major ones are included.

Table C.2
Correlation for various measures of "basic" research (all journals in the SCI)

	HiIPF85	HiIPF97	HiAcad85	HiAcad90	HiAcad97	JCRBas	CHIbas
HiIPF85	1.00						
HiIPF97	0.68	1.00					
HiAcad85	0.01	0.00	1.00				
HiAcad90	0.01	0.03	0.56*	1.00			
HiAcad97	0.03	0.01	0.49*	0.51*	1.00		
JCRBas	0.14*	0.11*	0.21*	0.20*	0.22*	1.00	
CHIbas	0.18*	0.16*	0.31*	0.27*	0.24*	0.36*	1.00

Note. The results are similar for journals relevant to semiconductors or pharmaceuticals.
* Significant at the 5% level.

The fourth variable, *HiAcad*, identifies research that is similar to academic research. It locates articles that are published in journals with a large percentage of academic authors. Such journals are often heavily circulated among academics, and their editorial boards are dominated by academics. *HiAcad* is based on each journal's *Acad* score, which is the percentage of papers in that journal with one or more academic authors. Each journal's *Acad* score is derived by searching the address fields of every article in the SCI for keywords such as "university", "school" and "ecole".[25]

The construction of the *HiIPF*, *Acad* and *HiAcad* variables is a novel contribution of this paper and may have other applications. As shown in Table C.1, most journals have a high percentage of papers with one or more academic authors. The twenty-fifth percentile of *Acad* is around 80%, so in three-quarters of the journals, at least 80% of the articles include an academic author. I define a journal to be *highly academic* if it is in the top quartile in its *Acad* score. This corresponds to having at least 93% of the papers in the journal written by at least one academic author.

$$\text{In year } yy\text{, define HiAcad}_{yy} \equiv \begin{cases} 1 & \text{if Acad}_{yy} > 93\% \\ 0 & \text{otherwise} \end{cases} \quad \text{(C.2b)}$$

Table C.2 shows the correlation between *HiIPF*, *HiAcad*, *JCRBas* and *CHIbas*. As expected, the correlation between variables is low because they measure different things. What is important is the *high correlation across years* for *HiIPF* and *HiAcad* (as is indicated by the shaded regions in Table C.2). This means that a journal's importance and academic orientation remained fairly stable between 1985 and 1997. I, therefore, use *HiIPF97* and *HiAcad97* to measure basicness. Additional analysis shows that the correlation coefficients remain similar within the subsets of journals relevant to semiconductors and pharmaceuticals.

The concentration of basic research relative to patents was computed using the journal classification schemes described in this appendix. The results are shown in Table C.3. Despite the low correlation between these measures, the results are remarkably consistent. Regardless of the way "basic" research is measured, it is much more concentrated than patents in the semiconductor industry, but not in the pharmaceutical industry. I conclude that the results are robust over a wide range of definitions for basic research.

[25] Not all journals have *Acad* scores for any given year, since some journals changed their names, merged with other journals, or were discontinued.

Table C.3
Gamma for various journal classification schemes

Definition of "basic" research article	γ_B for semiconductors	γ_B for pharmaceuticals
Articles in journal with JCRBas = 1	0.13	0.02
Articles in journal with CHIbas = 1	0.14	0.02
Articles in journal with HiIPF97 = 1	0.13	0.01
Articles in journal with HiAcad97 = 1	0.14	0.02

References

Adams, J.D., 1990. Fundamental stocks of knowledge and productivity growth. Journal of Political Economy 98, 673–702.

Ahuja, G., 2000. Collaboration networks, structural holes and innovation: a longitudinal study. Administrative Science Quarterly 45, 425–455.

Appleyard, M., 1996. How does knowledge flow? Interfirm patterns in the semiconductor industry. Strategic Management Journal 17 (Winter Special Issue), 137–154.

Appleyard, M., Kalsow, G.A., 1999. Knowledge diffusion in the semiconductor industry. Journal of Knowledge Management 3 (4), 288–295.

Arora, A., Gambardella, A., 1994. Evaluating technological information and utilizing it. Journal of Economic Behavior and Organization 24, 91–114.

Arrow, K.J., 1962. Economic welfare and the allocation of resources for invention. In: Nelson, R. (Ed.), The Rate of Direction of Inventive Activity. Princeton University Press, New Jersey.

Basberg, B.L., 1982. Technological change in the Norwegian whaling industry: a case study of the use of patent statistics as a technology indicator. Research Policy 11, 163–171.

Basberg, B.L., 1987. Patents and the measurement of technological change: a survey of the literature. Research Policy 16, 131–141.

Cockburn, I., Henderson, R.M., 1998. Absorptive capacity, co-authoring behavior and the organization of research in drug discovery. Journal of Industrial Economics 46 (2), 157–182.

Cockburn, I., Henderson, R.M., Orsenigo, L., Pisano, G., 1999. Pharmaceuticals and biotechnology. In: Mowery, D. (Ed.), U.S. Industry in 2000: Studies in Competitive Performance. National Academy Press, Washington, DC, pp. 363–398.

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

Cohen, W.M., Levinthal, D., 1990. Absorptive capacity: a new perspective on learning and innovation. Administrative Science Quarterly 35, 128–152.

Comanor, W.S., Scherer, F.M., 1969. Patent statistics as a measure of technical change. Journal of Political Economy 77, 392–398.

Curry, B., George, K.D., 1983. Industrial concentration: a survey. Journal of Industrial Economics 31 (3), 203–255.

Ellison, G., Glaeser, E.L., 1997. Geographic concentration in U.S. manufacturing industries: a dartboard approach. Journal of Political Economy 105 (5), 889–927.

Ernst, H., 1998. Industrial research as a source of important patents. Research Policy 27, 1–15.

Flatherty, M.T., 1984. Field research on the link between technological innovation and growth: evidence from the international semiconductor industry. American Economic Review Papers and Proceedings 74 (2), 67–72.

Fleming, L., Sorenson, O., 2001. Technology as a complex adaptive system: evidence from patent data. Research Policy 30, 1019–1039.

Gambardella, A., 1992. Competitive advantage from in-house scientific research: the US pharmaceutical industry in the 1980s. Research Policy 21, 391–407.

Garfield, E., 1980. Bradford's law and related statistical patterns. Current Contents 19 (May 12th), 5–12.

Gittelman, M., Kogut, B., 2003. Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns. Management Science 49 (4), 366–382.

Griliches, Z., 1980. Returns to research and development expenditures in the private sector. In: Kendrik, J., Vaccara, B. (Eds.), New Developments in Productivity Measurement and Analysis. University of Chicago Press, Chicago, pp. 419–454.

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

Griliches, Z., 1990. Patent statistics as economic indicators: a survey. Journal of Economic Literature 18 (4), 1661–1707.

Griliches, Z., 1992. The search for R&D spillovers. Scandinavian Journal of Economics 94, S29–S47.

Hall, B.H., 1996. The private and social returns to research and development. In: Smith, B., Barfield, C. (Eds.), Technology, R&D and the Economy. Brookings Institution, Washington, DC, pp. 140–183.

Hall, B.H., Ziedonis, R.H., 2001. The patent paradox revisited: an empirical study of patenting in the US semiconductor industry, 1980–1995. RAND Journal of Economics 32 (1), 101–128.

Hall, B.H., Jaffe, A., Trajtenberg, M., 2000. Market value and patent citations: a first look. NBER working paper W7741, Cambridge, MA.

Harhoff, D.F., Narin, F.M., Vopel, K.V., 1999. Citation frequency and the value of patented inventions. Review of Economics and Statistics 81 (3), 511–515.

Hausman, J., Hall, B.H., Griliches, Z., 1984. Econometric models for count data with an application to the patents–R&D relationship. Econometrica 52, 909–938.

Henderson, R.M., Cockburn, I., 1996. Scale, scope and spillovers: the determinants of research productivity in drug discovery. RAND Journal of Economics 27 (1), 32–59.

Hicks, D.M., 1996. A morphology of Japanese and European corporate research networks. Research Policy 25, 359–378.

Jaffe, A.B., 1986. Technological opportunity and spillovers of R&D: evidence from firms' patents profits and market value. American Economic Review 76 (5), 984–1001.

Jelinek, M., Schoonhoven, C.B., 1990. The Innovation Marathon: Lessons from High technology Firms. Basil Blackwell, Cambridge, MA.

Koenig, M.E.D., 1982. A bibliometric analysis of pharmaceutical research. Research Policy 12, 15–36.

Lane, P.J., Lubatkin, M., 1999. Relative absorptive capacity and interorganizational learning. Strategic Management Journal 19, 461–477.

Leonard-Barton, D., 1995. Wellsprings of Knowledge. Harvard Business School Press, Boston.

Levin, R.C., Reiss, P., 1988. Cost-reducing and demand-creating R&D with spillovers. RAND Journal of Economics 19 (4), 538–556.

Levin, R.C., Klevorick, A.K., Nelson, R., Winter, S.G., 1987. Appropriating the returns from industrial research and development. Brookings Papers on Economic Activity 3, pp. 783–820.

Lim, K., 2000. The many faces of absorptive capacity: spillovers of copper interconnect technology for semiconductor chips. MIT Sloan Working Paper 4110.

Macher, J.T., Mowery, D.C., Hodges, D.A., 1999. Semiconductors. In: Mowery, D. (Ed.), U.S. Industry in 2000: Studies in Competitive Performance. National Academy Press, Washington, DC, pp. 245–286.

Mansfield, E., 1981. Composition of R&D expenditures: relationship to size of firm, concentration, and innovative output. Review of Economics and Statistics 63 (4), 610–615.

Mansfield, E., Schwartz, M., Wagner, S., 1981. Imitation costs and patents: an empirical study. The Economic Journal 91 (364), 907–918.

Martin, B.R., Irvine, J., 1983. Assessing basic research. Research Policy 12, 61–90.

Mohnen, P., 1994. The econometric approach to R&D externalities. University of Quebec at Montreal working paper.

Moore, G.E., 1996. Some personal perspectives on research in the semiconductor industry. In: Rosenbloom, R.S., Spencer, W.J. (Eds.), Engines of Innovation. Harvard Business School Press, Boston, pp. 165–174.

Mowery, D.C., 1983. Innovation, market structure, and government policy in the American semiconductor electronics industry. Research Policy 12, 183–197.

Mowery, D.C., Rosenberg, N., 1989. Technology and the Pursuit of Economic Growth. Cambridge University Press, New York.

Narin, F., Noma, E., Perry, R., 1987. Patents as indicators of corporate technological strength. Research Policy 16, 143–155.

National Science Foundation, 1998. Science and Engineering Indicators. National Science Board, Washington, DC.

National Science Foundation, 2000. Federal R&D Funding by Budget Function: Fiscal Years 1999–2001. Arlington, VA.

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

Nonaka, I., 1994. A dynamic theory of organizational knowledge creation. Organization Science 5 (1), 14–37.

Pakes, A., Griliches, Z., 1984. Patents and R&D at the firm level: a first look. In: Griliches, Z. (Ed.), R&D, Patents and Productivity. Chicago University Press, Chicago.

Pisano, G., 1997. The Development Factory. Harvard Business School Press, Boston, MA.

Riordan, M., Hoddeson, L., 1997. Crystal Fire: The Birth of the Information Age. W.W. Norton, New York.

Romer, P.M., 1990. Endogenous technological change. Journal of Political Economy 98 (5), 71–102.

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

Rosenkopf, L., Almeida, P., 2003. Overcoming local search through alliances and mobility. Management Science, in press.

Scotchmer, S., 1991. Standing on the shoulder of giants: cumulative research and patent law. Journal of Economic Perspectives 5 (1), 29–41.

Solow, R.M., 1957. Technical change and the aggregate production function. Review of Economics and Statistics 39 (3), 312–320.

Stephan, P.E., 1996. The economics of science. Journal of Economic Literature 34 (3), 1199–1235.

Stern, S., Trajtenberg, M., 1998. Empirical implications of physician authority in pharmaceutical decision making. NBER Working Paper W6851, Cambridge, MA.

Stokes, D.E., 1997. Pasteur's Quadrant: Basic Science and Technological Innovation. Brookings Institution, Washington, DC.

Teece, D.J., 1987. Profiting from technological innovation: implications for integration, collaboration, licensing and public policy. In: Teece, D.J. (Ed.), The Competitive Challenge. Ballinger, Cambridge, MA, pp. 185–219.

Tilton, J.E., 1971. International Diffusion of Technology: The Case of Semiconductors. Brookings Institution, Washington, DC.

Trajtenberg, M., 1990. A penny for your quotes: patent citations and the value of innovations. RAND Journal of Economics 21 (1), 172–187.

Van Dalen, H.P., Henkens, K., 2001. What makes a scientific article influential? Scientometrics 50 (3), 455–482.

Wilson, R.W., 1980. Innovation, Competition and Government Policy in the Semiconductor Industry. Lexington Books, Lexington, MA.

Zucker, L.G., Darby, M., 1995. Virtuous circles of productivity: star bioscientists and the institutional transformation of industry. NBER working paper W5342, Cambridge, MA.