International Knowledge Flows:
Evidence from an Inventor-Firm Matched Data Set

Jinyoung Kim[†], Sangjoon John Lee[††], and Gerald Marschke[†††]

[†]Korea University
[††]Alfred University
[†††]University at Albany, NBER and IZA

April 2008

Abstract

We describe the construction of a panel data set from the U.S. patent data that contains measures of inventors' life-cycle R&D productivity—patents and patent citations. We match the data set to information on the U.S. pharmaceutical and semiconductor firms for whom they work. In this chapter we use these data to examine the role of research personnel as a pathway for the diffusion of ideas from foreign countries to U.S. innovators. In particular, we find in recent years an increase in the extent that U.S. innovating firms collaborate with or employ researchers with foreign experience. This increase appears to work primarily through an increase in U.S. firms' employment of foreign-residing researchers; the fraction of research-active U.S. residents with foreign research experience appears to be falling, suggesting that U.S. pharmaceutical and semiconductor firms are increasingly locating operations in foreign countries to employ such researchers, as opposed to such researchers immigrating to the U.S. to work. In addition, we investigate which U.S. firms conducting R&D build upon innovations originating abroad. We find that employing or collaborating with researchers who have research experience abroad seems to facilitate the use of output of non-U.S. R&D. We also find that in the semiconductor industry smaller and older firms, and in the pharmaceutical industry, younger firms are more likely to access foreign R&D output.

JEL Classification: J62, O31, O33
Key words: Patents; Innovation; Knowledge spillovers

1. Introduction

This chapter uses U.S. patent records to examine the nature and extent of knowledge spillovers from outside of the U.S. to U.S. industry. Because of their implications for economic development and science and technology policy, knowledge spillovers within a country or across borders have received considerable attention in the literature (e.g., Griliches, 1992). Knowledge spillovers between innovating firms on opposite sides of a national boundary can occur via arm's length communication (scholarly publications, the material published in patent applications and the like) or through person-to-person contacts in informal settings.[1] Knowledge spillovers across countries may accompany the migration of workers who relocate across international borders or collaborations between workers across borders, which is a focus of this chapter.

We examine whether international migration of researchers or the international location of subsidiaries by U.S. firms facilitates knowledge transfers across borders. Understanding how knowledge spillovers across countries work is of interest because of the role spillovers may play in economic growth and because of their implications for science and technology policy (Freeman, 2005; Regets, 2007). Knowledge spillovers from the U.S. and Europe may be an important factor for the impressive growth rates enjoyed in countries such as South Korea and Taiwan (Hu and Jaffe, 2003). Understanding the consequences of the immigration of scientists and researchers to the U.S. for U.S. R&D productivity, for wages and job prospects of native workers, and for national security has important implications for policy-making in the immigration, labor market, and education arenas.

---

[1] See Cohen, Nelson and Walsh (2002) on various means by which innovating firms access know-how developed externally. See Agrawal, Cockburn, and McHale (2003) for evidence of the importance of social networks in promoting diffusion. Von Hippel (1988) documents how direct informal contacts between researchers affect knowledge spillovers.

Studies in both the economics and sociology of innovation literatures argue that new technologies are frequently "tacit" and difficult to transmit to the uninitiated via spoken or written communication (Polyani, 1958, 1966). Often the most efficient means of transmission across organizational boundaries for tacit knowledge is via person-to-person contact involving a transfer or exchange of personnel. Recent findings that technological diffusion appears to be geographically limited (e.g., Jaffe, 1989; Jaffe, Trajtenberg, and Henderson, 1993; Audretsch and Feldman, 1996; Zucker, Darby, and Brewer, 1998; and Mowery and Ziedonis, 2001, Branstetter, 2001; Keller, 2002; Thompson and Fox-Kean, 2005) are often interpreted as evidence of the tacitness of knowledge (e.g., Feldman, 1994).

More direct evidence exists that person-to-person interaction is important for the diffusion of technology. Cohen, Nelson, and Walsh (2002) surveyed R&D managers on the means by which they gather and assimilate new technologies. They find that firms access externally-located technology partly through the hiring of and collaboration with researchers from the outside. Moreover, they find that hiring/collaboration with outside researchers is complementary to other means of accessing externally produced knowledge, such as through informal communications with outsiders and more formal (such as consulting) relationships with outsiders. Almeida and Kogut (1999) find that scientific references that firms cite in their patent applications reflect the employment histories of their inventors, suggesting that ideas in the semiconductor industry are spread by the movement of key engineers among firms, especially within a geographical area.[2] Zucker, Darby, and Armstrong (2001) find evidence of a pay-off to

---

[2] See also the (indirect) evidence of a link between scientific mobility and technological diffusion in Kim and Marschke (2005) and Moen (2005). Kim and Marschke find that firms are more likely to patent in environments where scientists are likely to switch employers, suggesting that workers do transmit technological know-how when they move from one employer to another. Technical knowledge acquired by the scientist that can be transmitted to future employers is a form of general human capital. Thus, scientists would be willing to pay by accepting lower wages to acquire technological knowledge that they can exploit with multiple employers. Moen finds some

firms that seek interactions with outside researchers. They find a positive impact on patent productivity for biotech firms that collaborate with university researchers on research and scholarly publications.

The above literature is at the least suggestive of the importance of the movement of technically trained personnel and extramural collaboration in facilitating knowledge transmission. We argue then that if one can measure the movement of researchers one can get a sense for the direction and magnitude of knowledge transmission. This chapter details the construction of a researcher-based data set and then describes its use in an analysis of the influence of foreign R&D on U.S. innovation. This chapter is part of a larger project that empirically examines issues related to the labor market for scientifically and technically trained personnel.

The first half of the chapter describes the construction of these data. The inventors behind a patented invention, as well as their home addresses, are listed on each U.S. patent, as are the firms to which the patent is assigned and the assignees' nationalities of incorporation. The firm to which the patent is assigned is in most cases the employer of the persons named in the inventor field. We match names in the inventor fields of patents to construct a panel data set of inventors that contains the patents in each year of the inventors' careers. The resulting data set allows us to track researchers geographically over the course of their career. These data afford us a window on the migration of technological human capital across national borders, one possible mechanism by which technology diffuses internationally. Patent applications disclose any previous relevant inventions. Through its citations to previous patents each patent documents the "prior art" upon which the new innovation builds, and because we know each cited patent's assignee type, we know in which sector and country the prior art originated. These

evidence of this: he shows that technical workers in R&D intensive firms in Norway accept lower wages early in their career in exchange for higher wages later.

citations provide an additional window on the pathways of knowledge (for evidence that citations proxy for knowledge flows, see Jaffe, Fogarty, and Banks, 1998; and Duguet and MacGarvie, 2005). In the final stage of constructing our patent-inventor data set we merge in citations made by the patent for each patent to which the inventor is named.

One use to which we wish to put our data is in understanding the factors that influence the innovating firm's accessing of recent innovations developed externally. A focus of this part of the analysis is the pharmaceutical and semiconductor industries, two industries that are especially prolific generators of innovations and patents and produce relatively homogenous outputs based on globally standardized technologies. Thus the last stage of data construction involves carefully matching the inventor data to data on publicly traded firms in these two industries.

After detailing our data construction efforts, we put our data to use investigating the international transmission of technology through scientific labor markets. For each patent assigned to a U.S. firm, we can determine the country of the inventor's residence at the time of patent application, and whether they had ever been named as an inventor on a patent while residing abroad. Inventing in a foreign country can be regarded as evidence of an inventor's exposure to research abroad. We also investigate which U.S. firms in our two industries cite foreign-assigned patents as prior art and thus build upon innovations originating abroad.

Our main findings are the following. We find that there has been an increase in recent years of U.S. innovating firms employing or collaborating with researchers with foreign experience. This increase appears to work primarily through an increase in U.S. firms' employment of foreign-residing researchers; the fraction of research-active U.S. residents with foreign research experience is low and appears to be falling, suggesting that U.S. pharmaceutical

4

and semiconductor firms are going to foreign countries to employ such researchers as opposed to such researchers immigrating to the U.S. to work for U.S. firms. We find, however, that inventors migrating to the U.S. with past foreign experience show the fastest growth in patent productivity and their patents receive more citations, possibly because of either accelerating knowledge spillovers or more selective migration of high-productivity inventors. In addition we investigate the firm-level determinants of accessing non-U.S. technological know-how. We find, for example, that employing or collaborating with researchers with research experience abroad seems to facilitate this access. Also, in the semiconductor industry, smaller and older firms and in the pharmaceutical industry, younger firms are more likely to make use of the output of non-U.S. R&D.

The chapter is organized as follows. Sections 2 and 3 describe the sources for the data construction and the construction itself. Section 4 details some descriptive statistics of the data set. Section 5 describes our analysis on the influence of foreign R&D on U.S. innovation. Section 6 concludes.

2. Data Sources

The data set we have created contains measures—patents and patent citations—of the R&D productivity of individual researchers between 1975 and 1998. For patents assigned to publicly traded firms in the U.S. pharmaceutical (Primary SIC code 2834) or semiconductor industry (Primary SIC code 3674), these data also contain information on the patents' assignees (e.g., firm size and R&D expenditures). Budgetary and time constraints limited the number of industries that we could include in our analysis. The pharmaceutical and semiconductor

industries were selected because they are especially prolific generators of innovations[3] and their products are relatively homogeneous[4] compared to those of other industries.

The data for this study come from five sources: (1) Patent Bibliographic data (Patents BIB) released by the U.S. Patent and Trademark Office (USPTO), which contains bibliographic information on all U.S. utility patents issued from 1969 to 2002; (2) the Compact D/SEC database from 1989 to 1997, which contains firm information taken primarily from 10-K reports filed with the Securities and Exchange Commission; (3) the Standard & Poor's Annual Guide to Stocks-Directory of Obsolete Securities, which includes a history of firm name changes, and of mergers and acquisitions; (4) the Thomas Register, Mergent, and Corptech data which report a firm's founding year, and finally (5) the NBER Patent-Citations data collected by Hall, Jaffe and Trajtenberg (2001), which contain all citations made by patents granted in 1975-1999. These data sources are described in detail below and the variables used in our study from each data source are in Table 1.

2.1. Patent Bibliographic data (Patents BIB)

Patents BIB is one of the Cassis Series of optical disc products released by the USPTO. Patents BIB contains bibliographic information for U.S. utility patents issued since January 1969. The information includes the patent ID number, dates of the patent's application and granting, patent assignee, and geographic information on all inventors involved. The original optical disc we use covers patents issued between 1969 and 2002, and contains over 3 million U.S. patents

---

[3] Based on NBER-Case Western University data of U.S. patents from 1963 to 1999, 15.3% of industry patents were granted to the firms in the pharmaceutical industry and 14.8% were granted to those in the semiconductor industry.
[4] In a cross-sectional analysis involving multiple industries, differing technologies and patent propensities make interpretation of results difficult. By limiting analysis to the patents and their inventors in a specific industry, we resolve heterogeneity in the propensity to patent across industries, thus making comparisons of patents and citations more meaningful.

granted. We use only the patents granted after January 1975 because detailed geographic information for all inventors is available in Patents BIB only for patents granted after that date. Most foreign innovating firms, especially those in Western Europe and in Japan apply for patents in the U.S. in addition to their home countries so that U.S. patent data reflect nearly the universe of patented innovations. Over this period the USPTO granted 2,493,610 patents (U.S. Patent No. 3,858,241 through 6,351,850), which together list 5,105,754 inventors (an average of 2.05 inventors listed per patent).

2.2. Compact D/SEC

The Compact D/SEC contains about 12,000 firms that have at least $5 million in assets and at least 500 shareholders of one class of stock of U.S. companies traded on the American Stock Exchange, the NASDAQ, the New York Stock Exchange, or the Over-the-Counter equities market. The data set provides financial and other information obtained from Annual Reports, 10-K and 20-F filings, and Proxy Statements for those companies. Most of the companies included are American. Company records include directory information, primary and secondary SIC codes, brief business descriptions, names of subsidiaries, names of top executives, ownership data, financial data, and excerpts from annual reports and other SEC reports.

2.3. Standard & Poor's Annual Guide to Stocks (S&P)

The firm-level information from the Compact D/SEC data cannot be directly matched to assignees in the Patents BIB data because parent firms patent sometimes under their own names and other times under the names of their subsidiaries. Mergers and acquisitions at both the parent firm and subsidiary levels and name changes further complicate linking the patent to firm-

level data. To track the ownership of firms over the entire period of our study, we use the information in the Standard & Poor's Annual Guide to Stocks. The S&P data provide histories of firm ownership changes due to mergers and acquisitions, bankruptcy, dissolution, and name changes, updated through December 2002.

2.4. NBER Patent-Citations

Patent applicants are legally obligated to disclose any knowledge they have of previous relevant inventions. Citations are of two kinds: to science (or prior science publications) and to technology (or previous patents). The patent examiner may add to the application relevant citations omitted by the applicant. Thus, through the patent citations each patent documents the "prior art" upon which the new innovation builds. Through the citations we can trace knowledge flow, measure the "closeness" of technological innovations, and measure an innovation's impact.

The data collected by Hall, Jaffe and Trajtenberg (2001) contain all citations made and received by patents granted between 1975 and 1999. Their data contain a total of 16,522,438 citation records; the mean number of citations received by a patent is 5.07, ranging from a minimum of 1 and a maximum of 779, respectively. The number of patents granted to the firms identified in the pharmaceutical and semiconductor industries between 1975 and 1999 is 244,158. The mean citations received by a patent in these two industries is 8.13, ranging from a minimum of 1 and a maximum of 631.

3. Data Set Construction Process

This section discusses key issues that arise in assembling our data set from these five sources. The assembly requires three steps. First, we create an inventor identifier in Patents BIB

because of the non-uniqueness of inventors' names. The primary challenge in this step is identifying who is who among inventors with same or similar names. Second, we identify each firm's ownership structure of subsidiaries and their name changes over the data period to construct firm-level data, using the Compact D/SEC and the S&P data. In the final step, we combine the inventor data and the firm data and then add the patent citation data where each citing patent that was granted between 1975 and 1999 is matched to all patents cited by the patent.

3.1. Identifying the same inventor among 'same/similar' names

Over 5.1 million inventor names are contained in the U.S. patent data from January 1975 through February 2002. Each inventor name record includes the last name, first name, middle name and suffix (Jr., Sr., etc.) of the inventor, as well as his/her city, state, and country of residence at the time of the granting of the patent.

Identifying the same inventor in different records with same or similar names (for example, John Maynard Keynes, John M. Keynes, John Keynes and John Keyens) is not an easy task. Our matching method uses as much information in the patent data as possible to increase the number of names matches without losing matching accuracy. Our name-matching methodology is similar to that in Trajtenberg, Shiff, and Melamed (2006).

To start, we treat each entry that appears in the inventor name field of every patent in the Patents BIB data as a unique inventor. Given N number of names in this name pool, we pair each name with all other names, which generates N(N-1)/2 number of unique pairs. The 5.1 million names in the Patents BIB data (2.05 inventors per patent) thus produce 13 trillion unique pairs. For each pair, we consider the two names as belonging to the same inventor if the

SOUNDEX codes of their last names and their full first names are the same, and at least one of the following three conditions is met: (1) the full addresses for the pair of names are the same; (2) one name from the pair is an inventor of a patent that is cited by another patent whose inventors include the other name from the pair; or (3) the two names from the pair share the same co-inventor. In implementing the second and third conditions, we make comparisons based on whether the first and last names are spelled identically. After our name matching procedure is completed, we go back and check that these conditions are still valid based on the inventor identifier constructed by the matching procedure. If not, we repeat the name matching process to create a new inventor identifier.

SOUNDEX is a coded index for last names based on the way a last name sounds in English rather than the way it is spelled. Last names that sound the same, but are spelled differently, like SMITH and SMYTH, have the same SOUNDEX code. We use the SOUNDEX coding method to expand the list of similar last names to overcome the potential for misspellings and inconsistent foreign name translations to English; misspellings are common in the USPTO data as are names of non-Western European origin (see Appendix A for the detailed SOUNDEX coding method).

We also consider a pair of names as a match if two have the same full last and first names as spelled in the Patent BIB data, and at least one of the following conditions is met: (1) the two have the same zip code; (2) they have the same full middle name; or (3) they reside in the same MSA area. As an additional step beyond the aforementioned pair-wise comparisons, we treat a pair of inventors as mismatched if the middle name initials of the pair are different.

Table 2 illustrates our name-matching procedure. Inventors 001 and 002 in Table 2 have the same last and first names, and share the same co-inventor. Thus, the two records in this pair

are treated as the same inventor. Inventors 002 and 003 do not have the same full middle name but share the same zip code, and thus the two inventors are treated as the same inventor. Although inventors 002 and 005 share the same zip code, the middle name initials are different. Therefore, the pair is not considered a match (they would not be considered a match by our algorithm even if their street addresses were identical, possibly a case of a parent and a child).

Imposing transitivity

Transitivity is imposed in the following sense: If name A is matched to name B and name B is matched to name C, name A is then matched to name C. We iterate this process until all possible transitivity matches are completed. After the transitivity procedure, we assign the same inventor ID number for all the names matched. For instance, inventors 001 and 003 are not linked in the initial round of name matching, but they are matched through transitivity because inventors 001 and 002 are matched and inventors 002 and 003 are matched.

Imposing transitivity, however, poses a possibility of name mismatch. Suppose, for example, Adam E. Smith and Adam Smith are matched in one pair, and Adam J. Smith and the same Adam Smith are matched in another pair. According to our transitivity procedure, Adam E. Smith and Adam J. Smith are identified as a match although their middle name initials are different. The number of matches through transitivity suffering from this problem appears to be trivial, however: we find 126 cases where two inventors are matched although their middle names were different out of 2.3 million uniquely identified inventors. Upon further investigation of these cases, we found the mismatches are of three kinds. In the first kind, some middle names in the Patents BIB data are incorrectly coded. For instance, our transitivity procedure matched the names 'Laszlo Andra Szporny' and 'Laszlo Eszter Szporny' which appear to belong to the same inventor according to other information. We found that the middle names attributed to him

11

are the first names of the next co-inventors listed on his patents, suggesting that "Andra" and "Eszter" are not his middle names. In the second kind of mismatch, an inventor with two middle names is coded in the Patents BIB data with one middle name in some cases and with the other middle name in other cases. In the third kind, a mismatch occurs when two inventors with the same last and first name but different middle names appear in the same patent. We corrected by hand instances of the first two kinds of mismatch, but dropped from our data the observations displaying the third kind of mismatch.

Trajtenberg et al. assign scores for each matching criterion and consider a pair matched only if its total score from all matching criteria exceeds a threshold. Because the choice of weights and the score threshold for a match is largely arbitrary, we do not use this scoring method in our data construction. Our method also differs in that we do not use as a matching criterion whether two inventors share the same assignee because name matching based on this criterion might bias our measure of mobility among inventors. Instead we apply the rule that two inventors are not treated as a match if their middle name initials differ. From our experience with the patent data, imposing this rule is effective because the SOUNDEX coding system sometimes so loosely specifies names that apparently different last names are considered a match.

In the end, because of these differences, the number of distinct inventors identified with our procedure is a little higher than the number of distinct inventors reported in Trajtenberg et al. (2006). We identified 1.72 million unique inventors (34%) out of 5.1 million names in the entire patent data while Trajtenberg et al. found 1.6 million distinctive inventors (37%) out of 4.3 million names. Note that our patent database is larger because it includes additional years, 2000-2002.

Matching Accuracy

We evaluated the accuracy of the algorithm using the curricula vitae of a sample of 100 inventors (which we will call our "benchmark sample") obtained from the internet. While the benchmark inventors may not be representative of the inventors in our data in many ways—the benchmark inventors are more likely to have academic ties and are more prolific than the inventors in our data set[5]—they still tell us something useful about the errors produced by our algorithm.

The algorithm may over or undermatch patents and inventors. An inventor is subject to undermatching error if the matching algorithm fails to group all of the inventor's patents under one inventor identification number. The algorithm correctly groups 685 (89%) of the 769 patents that belong to the 100 inventors in the CV sample. In other words, a reassignment of as few as 84 patents would eliminate the undermatching error in the benchmark sample. Yet partly because the benchmark inventors are so prolific, we find that the algorithm has failed to match at least one patent for each of 38 of the 100 benchmark inventors. The fewer patents an inventor has, the smaller the opportunity for the algorithm to misassign one of his patents,[6] suggesting that the undermatching error problem in the actual data is smaller than the undermatching error rate in the benchmark sample, at least in terms of the fraction of inventors affected.[7]

Overmatching occurs when a patent is assigned to the wrong inventor. Overmatching error is found in eight of the 100 benchmark inventors. To these eight inventors, the algorithm

---

[5] The inventors in our CV sample have on average over 7 patents each compared to a little over one patent each among our matched data.

[6] The average number of patents per inventor affected by undermatching error was 9.6 compared to 6.6 for those not affected by undermatching error. The average number of patents per inventor in our data set is less than 2.

[7] Undermatching error appears to be due primarily to the importance that the algorithm places on the inventor's middle name for matching and the inconsistency with which an inventor's middle name is represented from one patent to the next patent. We are currently testing an improved version of the matching algorithm that allows for more variation in the way the middle name appears. Preliminary testing suggests that the new algorithm significantly reduces undermatching error at little cost in increased overmatching error.

should have assigned 158 patents, but instead assigned 308 patents. Two inventors whose last names are Johnson and Smith account for 138 of 150 overmatched patents, suggesting that overmatching error arises in our data for researchers with common names.

How might the under and overmatching errors affect our results? Error rates affect the accuracy of an estimate at a point in time, as when we estimate the fraction of patents in a given year that name an inventor with foreign research experience. On estimates of trends, however, matching errors may have qualitatively little effect if error rates are not changing over time. One concern that has been expressed to us is that inventor names on U.S. patents have become less varied. East Asian surnames, which are increasing in frequency among inventors on U.S. patents, tend to vary less than Western European names. Therefore, the overmatching error rate may be increasing. If names are getting more similar, we may be lumping under a single ID an inventor with foreign experience and an inventor without foreign experience. This will increase the frequency with which, for example, we find an inventor team with foreign experience and cause us to find a positive trend in foreign influence even if there is not. Our algorithm, however, did not over-assign patents to any of the 11 benchmark inventors with surnames of Asian ancestry, possibly because there is sufficient variation in their first names, which are also used in the matching.

3.2. Identifying the ownership structure and combining patent-inventor data with firm data

Because parent firms patent sometimes under their own names and at other times under the names of their subsidiaries, combining the Patents BIB data with firm-level data in the Compact D/SEC data is not straightforward. Mergers and acquisitions at both the parent firm and subsidiary levels, common in these two industries during the 1990s, and name changes

complicate linking the patent to firm-level data. (The USPTO does not maintain a unique identifier for each patenting assignee at the parent firm level nor does it track assignee name changes.) Thus, to use the firm-level information available in the Compact D/SEC data, the names of parent firms and their subsidiaries and the ownership of firms must be tracked over the entire period of the study.[8]

To start, we identify mergers and acquisitions, and name changes of firms in the two industries, pharmaceutical preparation (primary SIC code 2834) and semiconductor and related devices (3674), over the period between 1989 and 1997, using the Standard & Poor's data. We also identify the ownership structure of subsidiaries of firms using subsidiaries information available from the Compact D/SEC from 1989 to 1997.[9] We can then relate each assignee in the patent data to a firm in the Compact D/SEC data, which enables us to match each patent to a firm in the Compact D/SEC data. We then combine firms' founding years, obtained from Thomas Register, Mergent, and Corptech, with the other firm-level information.

As the final step, we add information on all citations from the NBER Patent-Citations data collected by Hall, Jaffe and Trajtenberg (2001) where each citing patent that was granted between 1975 and 1999 is matched to all patents cited by the patent.

4. Descriptive statistics

Figure 1 describes the distribution of U.S. patents granted by year of application. The figure shows the surge in patenting that began in the mid-1980s. The applicant flow for the

---

[8] NBER-CWRU researchers created a database of parent firms and their subsidiaries for all the names among USPTO patent assignees. However, they only linked subsidiaries based on the corporate ownership structure as it existed in 1989.

[9] The subsidiary list reported in the Compact D/SEC is not always complete. For example, some subsidiaries appear intermittently and some firms report subsidiaries every other year. Hence, we have treated a subsidiary as one for the firm throughout the period 1989-1997, if it is reported once as a subsidiary of the firm.

previous twenty-five years had been remarkably stable. The possible causes of this patent surge have been discussed in Hall and Ziedonis (2001), Kortum and Lerner (1999, 2003), and Kim and Marschke (2004). Because it covers the surge, the mid-1980s through the late 1990s is an interesting period to examine and is the period we study here.[10]

Figure 1 shows that the annual number of patents granted dips sharply after 1997. This dip reflects a lag between the application and granting dates. About 70 to 80 percent of all patent applications ultimately granted are granted within the first three years of the application and 97 percent of all patent applications are granted within the first four years of the application date (Hall, Griliches, and Hausman, 1986). For this reason, the last year covered by our analysis is 1997. Between January 1975 and February 2002, 45.5% were granted to U.S. assignees and 37.4% were granted to foreign assignees (see Table 3) with the rest unassigned. In Figure 2, we report the number of patents granted to firms in each of our two industries. Note that in both industries the number of patents granted annually rose over the period we study: the annual number of patents granted between 1989 and 1998 rose from about one thousand patents annually, but by a factor of two in the pharmaceutical industry and nearly seven in the semiconductor industry.

Table 4 shows that the number of inventors named as an inventor to at least one patent assigned to a firm in one of our two industries is 59,292 out of the 2,299,579 unique inventors in our data (25,609 inventors in the pharmaceutical and 33,683 in the semiconductor industry). Inventors working in the pharmaceutical and semiconductor industries are named as inventors on more patents on average than inventors in other industries (see Table 4). An inventor in a pharmaceutical firm is named as an inventor on average on 2.80 patents over our sample period, whereas an inventor in the semiconductor industry appears on average on 2.60 patents.

---

[10] The application rate has since appeared to level off, somewhat. The number of patents granted in 2004, 2005, and 2006 were approximately 165,000, 144,000, and 174,000, respectively.

We identified pharmaceutical and semiconductor firms in the Compact D/SEC data by their primary SIC. We identified 447 parent firms and 5,331 subsidiary firms in the pharmaceutical industry and 332 parent firms and 4,211 subsidiary firms in the semiconductor industry. Firm information starts in 1989 because we had access to the Compact D/SEC data only beginning in 1989. We dropped all patent applications filed after 1997 because as discussed above starting with application year 1998 the patent time series tailed off due to the review lag at the USPTO.

Some sample statistics from the firms in the two industries in our data—the number of selected firms and the number of employees, sales, and R&D expenditures—are reported in Table 5. For the year 1997, for example, the data show 221 firms in the pharmaceutical and 151 firms in the semiconductor industry, with 177 firms and 135 firms, respectively, reporting positive R&D expenditures. Pharmaceutical firms are larger in terms of number of employees, sales volume, and R&D expenditures.

5. International Knowledge Flows

A small literature uses patents to examine international knowledge flows. Jaffe and Trajtenberg (1998) find, for example, that patents are more likely to cite other patents from the same country and that citations to other country's patents occur after a lag, suggesting that international borders impede or slow knowledge diffusion. Singh (2007) uses patent citations to examine knowledge flows between foreign-based subsidiaries of multinational corporations (MNCs) and their host countries. He finds that host country patents cite the patents of local foreign MNC subsidiaries at high rates. He also finds that MNC patents cite host country patents at an even higher rate, especially when the host country is technologically advanced. He

17

interprets this to mean that significant amounts of knowledge flow from MNCs to the host country, and even greater quantities flow from the host country to the MNC. He also finds that inventor flows between MNCs and host countries correlate with the direction of the flow of research personnel.[11]

We focus on knowledge flows to the U.S. from abroad and are interested in how they have changed in recent years, especially over the period of the patent surge that began in the mid-1980s. We use the geographic mobility of researchers and their location at the time of invention to track the transmission of foreign knowledge from other countries to the U.S. In addition, we test if the international migration of researchers facilitates knowledge transfers across borders.

Table 6 shows the annual number of unique inventors named on U.S. domestic patents for the years 1985 through 1997. It also shows the percentages of inventors who at the time of the patent application (1) resided in a foreign country, (2) resided in the U.S. and had been previously listed as a foreign-residing inventor on a successful patent application, and (3) resided in the U.S. but had never been previously listed as a foreign residing inventor on a successful patent application. Because our data included patents granted in 1975 and later, we imposed a cut-off for the patents used to define whether an inventor has foreign-experience at the time of the patent's application. We consider as "foreign-experienced" only those inventors who are currently foreign residents or had been foreign residents sometime in the ten-year period prior to the date of the patent's application because ten years still leaves us a long period over which to conduct our analysis and because knowledge acquired in a foreign country far in the past may not be very valuable.

---

[11] Singh also uses the USPTO patent data to produce an inventor panel. His name-matching strategy differs in several ways from our (see Singh). Singh does not report on the accuracy of his match.

Table 6 shows a dramatic increase in the number of unique inventors on U.S. domestic patents between 1985 and 1997, from 42,368 to 119,556, which translates to an average annual growth rate of 9 percent. This increase in this period is expected given the timing of the patent surge. Among those inventors with foreign experience, the percentage of inventors with current foreign addresses increased steadily during the period from 8.15 percent to 9.11 percent while the percentage of U.S.-residing inventors with foreign experience increased from 0.99 percent in 1985 to 1.30 percent in 1992, then dropping to 1.01 percent in 1997. Overall, the percentage of inventors with foreign experience increased (from 9.14 percent in 1985 to 10.13 percent in 1997).

Table 6 shows that the growth in the number of inventors in the pharmaceutical (13 percent annually) and semiconductor (31 percent annually) industries has been significantly faster than for all industries combined. In the pharmaceutical industry the share of inventors with foreign experience grew rapidly although the increase is mostly in the share of inventors with current foreign addresses and there is a decrease in the fraction of U.S.-residing inventors with past foreign experience. This finding is not surprising given the increasing rate at which U.S. pharmaceutical firms have been citing new laboratories abroad (Chacar and Lieberman, 2003) and findings that collaborations among academic scientists have become more dispersed, possibly due to improvements in telecommunications (Adams, Black, Clemmons, and Stephan, 2004). The semiconductor industry shows a similar pattern, but the changes are less pronounced than in pharmaceutical industry.[12]

Figure 3A shows the average annual patent productivity of inventors in U.S. domestic patents by foreign-experience type for all patents. Figures 3B and 3C repeat the analysis of Figure 3A but for the pharmaceutical and semiconductor industries alone. The calculations in

---

[12] Phene and Almeida (2003) note a dramatic rise in overseas patenting by the five leading U.S. semiconductor companies between 1986 and 1995.

19

these figures are based on inventors named to at least one patent in each year and to at least one additional patent during the previous 10 year period. The latter restriction is imposed because inventors with foreign experience have at least one patent earlier by design. We first note in these figures that patent productivity for all three types has been increasing. However, the growth rate of patent productivity is the highest for the U.S.-residing inventors with past foreign experience and those inventors in later years have significantly higher patent-inventor ratio than other types of inventors. On the other hand, the growth in patent productivity among current foreign residents has been the slowest. There are a number of possible explanations for this. First, inventors with higher productivity are more likely to migrate to the U.S., especially in recent years, because of better compensation for skilled labor in the U.S. labor market or because of U.S. immigration policies. Second, as shown in Table 6, the share of current foreign residents has been rising while that of U.S. residents with foreign experience has been falling, especially in our two industries. These changes may be associated with a more selective migration of researchers with higher productivity. Third, foreign experience somehow improves the productivity of researchers, proportionally more in recent years, or inventors with foreign experience happen to be working in technological areas with higher patent propensities.

Figures 3B and 3C show qualitatively similar changes in productivity among the different types of inventors. Note that because we are looking within an industry with rather homogenous technology, these gaps are less likely due to heterogeneity in technology class. Figure 3C more clearly demonstrates the changes in patent productivity across the types of inventors in the semiconductor industry.

Where Figures 3A-C track the productivity of inventors by their patent output, Figures 4A-C track how the quality of inventors' output changes by inventor type. There is evidence that

20

citations received reflect the economic value of the patent (Trajtenberg, 1991). Figure 4A shows the citations received in the 5-year period following application per patent by inventor type for all industries over time. This figure covers only years of application through 1992 because the NBER citation data contain citations made by patents granted in years up to 1999 and we take into account the 5-year period of citation and a 2-year gap between application and granting dates. Between 1985 and 1992, the citations per patent rose for all three classes of inventors. Throughout the 1985-1992 period, the average citations per patent produced was the highest and grew fastest for U.S. residents with foreign experience and was the lowest and grew slowest for foreign residing inventors. In 1992, the number of citations attracted by the average patent of a U.S. residing inventor with foreign-patenting experience, U.S.-residing inventor without foreign patenting experience, and foreign residing inventor, was about 6.5, 5, and 3.5 respectively. Thus, taken together, Figures 3A and 4A show that U.S.-residing inventors with foreign experience produce more patents on average and patents of higher quality than the other two classes of inventors between 1985 and 1992. Figures 4B and 4C conduct the analysis separately for the pharmaceutical and semiconductor industries. The semiconductor industry shows the same ordering of inventor types, though the levels are higher for each type. The pharmaceutical industry however shows no clear and consistent distinction between the two classes of U.S.-residing inventors. Figure 4C does show that foreign-residing inventors produce the lowest quality patents, by the citation measure and of U.S.-residing inventors, those with foreign patenting experience produce more valuable patents. Thus it appears that in both industries in the late 1980s into the 1990s U.S.-residing inventors with foreign experience have higher patent rates. In the semiconductor industry, their patents also attract more citations.

21

Is a patent from a domestic firm more likely to cite foreign-assignee patents when its inventors have foreign experience? We are interested in learning if knowledge spillovers from foreign countries are facilitated by direct exposure to inventors with foreign experience. Table 8 presents the results of our estimation of the determinants of accessing foreign knowledge in the pharmaceutical and semiconductor industry. The unit of observation of the regression is a patent applied for in year 1997. The dependent variable is the fraction of citations to the patent that are assigned to foreign assignees (CITE_FRGN). The key regressor in these regressions is a binary variable which takes 1 if at least one inventor on the patent is currently residing or formerly resided in one of the foreign countries where foreign assignees of cited patents are located (FRGN_EXP). Note that this regressor reflects not just whether an inventor has foreign experience but which country the inventor has experience from. We speculate that knowledge spillover is country-specific.

The regressions in Table 8 also include as right-hand side variables firm-level characteristics in year 1997. A measure of the size of the research operation, proxied by the number of unique inventors named to patents awarded to the firm in 1997 (INVENTOR), is included to examine whether large-scale R&D enterprises are more likely to rely on foreign knowledge. We use the number of employees (EMPLOYEE) as an alternative measure of organizational size at the firm level. Included are the R&D-inventor ratio (R&D/INV) and the number of business lines in the firm (NSIC), measured by the number of secondary SIC's identified with the firm. We include the R&D-inventor ratio (R&D/INV) as a regressor because a highly capitalized firm may rely on more advanced technology and thus may be more open to foreign technology. We include NSIC as a regressor to estimate the impact of economies of scope in the firm's use of foreign knowledge. Our regressions also include the median

experience of all inventors in the firm (MEXP) and years elapsed since the founding year of the firm (FIRMAGE). The means and standard deviations of the independent and dependent variables, along with their definitions, are described in Table 7.[13]

Column 1 in Table 8 for each industry panel shows the estimated relationship between the fraction of a patent's citations to foreign-assigned patents and the existence of foreign-experienced inventors using OLS. Column 2 for each industry panel reports the estimates of the determinants of the citation to foreign patents.

One concern for our regression is that inventors are more likely to cite their own past patents than other inventors' patents, which may drive the estimated relationship between our dependent variable and the key regressor, FRGN_EXP. In column 3 for each industry panel we thus exclude patents which have the same inventors as those in their cited patents.

The results in Table 8 show that a patent by inventors with foreign experience in both industries is more likely to cite patents assigned to foreign firms from the same country where the inventors are residing or resided in the past: FRGN_EXP has a significantly positive effect in all models. This effect is still significant with the data without self-citing patents. The coefficient estimates suggest that having an inventor on the patent with patenting experience in a particular (non-U.S.) country increases the fraction of citations to that country's patents by between .22 and .29.

The results show a negative effect of the size of the R&D enterprise on the fraction of citations to foreign patents in the semiconductor industry. There is no significant effect of the size of the R&D enterprise in the pharmaceutical industry. On the other hand, the coefficient estimate on the firm size variable (EMPLOYEE) is insignificant in all models. The coefficient

---

[13] Note that the means of the variables reported in Table 7 are not the averages across firms because our regressions are at the patent level, not at the firm level. For instance, the mean value of INVENTOR and EMPLOYEE is greater than the firm mean in 1997 because larger firms tend to have more patents.

estimate on log R&D/INV is generally positive but insignificant in all regressions. The coefficient estimate on log NSIC is never significant by conventional criteria of significance. The coefficient estimate on log MEXP is negative and significant for both industries. This may partly reflect that it is more costly for older inventors to learn new technologies from abroad, or it may be due to a vintage or a composition effect (e.g., areas of technology that experienced innovators innovate in are somehow more "domestic"). The coefficient estimate on log FIRMAGE is significant for both industries but has different signs for the two industries. The effect is negative in pharmaceutical industry while it is positive in semiconductor industry. That is, we find that in the semiconductor industry older firms, and in the pharmaceutical industry, younger firms are more likely to make use of the output of non-U.S. R&D.

6. Conclusion

We describe the construction of a panel data set that links inventors to the U.S. pharmaceutical and semiconductor firms for whom they work. These data contain measures of inventors' R&D productivity—patents and patent citations—as well as information on the firms to which their patents are assigned. In this chapter we use these data to examine the role of research personnel as a pathway for the diffusion of ideas from foreign to U.S. innovators. We envision that local knowledge abroad that is tacit can be accessed or imported in two ways. The first is for U.S. firms to move closer to the sources of local foreign knowledge, possibly by setting up subsidiaries abroad (see Phene and Almeida, 2003) and by hiring local scientists, or by sending firms' U.S. scientists abroad to the subsidiaries. In our analysis this is captured by the number of inventors who are foreign-residing at the time of invention. The second way is for firms to hire scientists who had previously worked in laboratories abroad; that is, for scientists

24

with a foreign background to move to U.S. firms on U.S. soil. This is captured by the number of inventors who are residing in the U.S. but have foreign experience.

Table 6 suggests that U.S. domestic firms are relying more on the first way than the second way of accessing foreign sources of knowledge. In any year, the fraction of inventors who are residing in the U.S. and have foreign patenting experience is always less than .02. The fraction of inventors on U.S. patents assigned to U.S. firms who are abroad at the time of invention rises from about .08 in 1987 to about .09 in 1997. In the pharmaceutical and semiconductor industries, however, foreign-residing inventors are used more extensively and the growth has been more dramatic. Between 1989 and 1997, the fraction of inventors who were foreign-residing at the time of invention rose from .15 to .30 and from .09 to .15 for the pharmaceutical and semiconductor industries, respectively. This may be consistent with the argument by Phene and Almeida (2003) that the ability of local subsidiaries of multinational enterprises to tap the stock of knowledge in their host countries increases as part of a maturation process.

While U.S. innovating firms' employment of migrant workers with foreign research experience has fallen in relative terms, compared to foreign-residing and U.S. domestic researchers without foreign experience, these migrant workers are highly productive. Moreover, their productivity has increased over the period that we study, possibly because of either accelerating knowledge spillovers or more selective migration of high-productivity inventors.

In Table 8, we present evidence that either employing researchers abroad or foreign-experienced researchers in the U.S. contributes to the import of foreign knowledge. The citation of the patent of a foreign assignee by a U.S. firm's patent represents either an assimilation of foreign knowledge by the parent firm in the U.S. or the assimilation of the foreign knowledge by

foreign subsidiaries of the U.S. firm. Evidence from the international business literature suggests that the multinational corporation's home base learns from its foreign-based subsidiaries (Singh, 2007; Kogut and Zander, 1993; Dunning, 1992). Thus we interpret the evidence presented in Table 8 as evidence that when U.S. innovating firms employ either foreign-experienced researchers in the U.S. or those at a foreign-based subsidiary, transmission of foreign knowledge from their foreign origins to the U.S. results. Tables 6 and 8 together, we believe, suggest that this transmission has been increasing in the U.S. pharmaceutical and semiconductor industries from the late 1980s through the late 1990s.

Our findings are consistent with reports that during this period the U.S. pharmaceutical industry has been increasing the pace at which it is establishing laboratories on foreign soil. Our findings are also consistent with arguments that foreign subsidiaries of U.S. semiconductor firms are becoming more adept at extracting foreign based knowledge.

We anticipate this data set will be useful in addressing other important questions. These data will allow us to investigate the consequences of the mobility of R&D personnel on firm R&D. What is the impact, for example, of the arrival of a researcher with a particular set of R&D experiences on the character and quantity R&D done by a firm? We will be able to address this question because we know each researcher's patenting history, both in terms of quantity but we also know the kinds of technologies underlying the innovations. This data set will allow us to directly observe the importance of inter-firm mobility for technological diffusion. From the perspective of the researcher, this data set will allow us to examine the determinants of inter-firm mobility. The panel nature of these data will allow us to investigate the productivity profiles of researchers working in industry over their careers. Because we observe all the inventors responsible for a patent we will be able to use this data set to investigate how firms

organize the R&D enterprise, the extent of collaboration among researchers who are geographically dispersed, and the extent of interaction among researchers with different backgrounds.

Appendix A. The SOUNDEX Coding System

The SOUNDEX is a coded index for last names based on the way a last name sounds rather than the way it is spelled. Last names that sound the same, but are spelled differently, such as SMITH and SMYTH, have the same SOUNDEX code. We use the SOUNDEX coding method to expand the list of similar last names to overcome the potential for misspellings and inconsistent foreign name translations into English; misspellings are common in the USPTO data, as are names of non-Western European origin.

A SOUNDEX code for a last name takes an upper case initial followed by 6-digit numeric codes. For example, the SOUNDEX code for Keynes is K520000. The rules for generating a SOUNDEX code are[14]:

1. Take the first letter of the last name and capitalize it.

2. Go through each of the following letters giving them numerical values from 1 to 6 if they are found in the Scoring Letter table (1 for B, F, P, V; 2 for C, G, J, K, Q, S, X, Z; 3 for D, T; 4 for L; 5 for M, N; 6 for R; 0 for Vowels, punctuation, H, W, Y).

3. Ignore any letter if it is not a scoring character. This means that all vowels as well as the letters h, y and w are ignored.

4. If the value of a scoring character is the same as the previous letter, ignore it. Thus, if two 't's come together in the middle of a name they are treated as a single 't' or a single 'd'. If they are separated by another non-scoring character then the same score can follow in the final code. The name PETTIT is coded as P330000. The second 'T' is ignored but the third one is not, since a non-scoring 'I' intervenes.

5. Add the number onto the end of the SOUNDEX code if it is not to be ignored.

6. Keep working through the name until you have created a code of 6 characters maximum.

7. If you come to the end of the name before you reach 6 characters, pad out the end of the code with zeros.

8. You may choose to ignore a possessive prefix such as 'Von' or 'Des'.

See "Using the Census SOUNDEX," General Information Leaflet 55 (Washington, DC: National Archives and Records Administration, 1995) for the detailed method.

---

[14] The strings of '-, ., +, /, (,), %, ?, #, &, ", _' in all name fields have been translated to blank space in advance and then last names are SOUNDEX coded.

REFERENCES

Adams, J.D., Black, G.C., Clemmons, J.R., and Stephan, P.E. (2004). "Scientific Teams and Institution Collaborations: Evidence from U.S. Universities, 1981-1999," NBER Working Paper no. 10640.

Agrawal, Ajay K., Cockburn, Iain M., and McHale John 2003 "Gone but Not Forgotten: Labor Flows, Knowledge Spillovers, and Enduring Social Capital," NBER Working Paper no. 9950

Almeida, Paul and Kogut, Bruce. "Localization of Knowledge and the Mobility of Engineers in Regional Networks," *Management Science* 45(7), July 1999.

Audretsch, D. B., and Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *The American Economic Review*, *86*, 630–640.

Branstetter, L.G. (2001). "Are Knowledge Spillovers International or Intranational in Scope?" *Journal of International Economics*, 53(1): 53-79.

Chacar, A. S. and Lieberman, M.B. (2003). "Organizing for technological innovation in the U.S. Pharmaceutical Industry." *Geography and Strategy: Advances in Strategic Management*, 20, 299–322.

Cohen, Wesley M., Nelson, Richard R., Walsh, John P. 2002. "Links and Impacts: The Influence of Public Research on Industrial R&D." *Management Science*, 48(1), pp. 1–23

Duguet, Emmanuel and MacGarvie, Megan. 2005. "How well do patent citations measure flows of technology? Evidence from French innovation surveys," *Economics of Innovation and New Technology*, 14(5), 375-393

Dunning, J.H. (1992). *Multinational Enterprises and the Global Economy*. Addison-Wesley: Wokingham, England; Reading, Mass.

Feldman, M. P. (1994). *The geography of innovation*. Boston: Kluwer Academic Publishers.

Freeman, Richard B. (June 2005) "Does Globalization of the Scientific/Engineering Workforce Threaten U.S. Economic Leadership?" NBER Working Paper No. 11457.

Griliches, Z. (1992). The search for R&D spillovers. *Scandinavian Journal of Economics* (Supplement), 29–47.

Hall, B., Griliches, Z. and Hausman, J. "Patents and R&D: Is There a Lag?" *International Economic Review*, Vol. 27 (1986), pp. 265-283.

Hall, B., Jaffe, A. and Trajtenberg, M. (2000). "Market Value and Patent Citations: A First Look." *NBER Working Paper* No. 7741.

Hall, B. H., Jaffe, A. B. & Trajtenberg, M. (2001). The NBER patent citation data file: Lessons, insights and methodological tools. NBER Working Paper 8498.

Hall, B. and Ziedonis, R. (2001) The Determinants of Patenting in the U.S. Semiconductor Industry, 1980–94. *Rand Journal of Economics*, 32(1), 101–128.

Hu, A. G.Z and Jaffe, A. B. (2003) "Patent citations and international knowledge flow: the cases of Korea and Taiwan." *International Journal of Industrial Organization*, 21, pp. 849-880.

Jaffe, Adam B. 1989. "Real Effects of Academic Research." *The American Economic Review*, 79(5), pp. 957-970.

Jaffe, Adam B., Fogarty, Michael S., and Banks, Bruce A. (1998) "Evidence from Patents and Patent Citations on the Impact of NASA and Other Federal Labs on Commercial Innovation," *Journal of Industrial Economics*, 46 (2) , 183–205

Jaffe, A. B and M. Tratjenberg (1999), "International Knowledge Flows: Evidence from Patent Citations," *Economics of Innovation and New Technology*, 8, pp. 105-136.

Jaffe, Adam, Trajtenberg, Manuel and Henderson, Rebecca, (1993) "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics* 108 (3), pp 577-598, 1993.

Keller, W. (2002). "Geographical Localization of International Technology Diffusion," *American Economic Review*, 92(1), pp. 120-142.

Kim, Jinyoung and Marschke, Gerald, 2004. "Accounting for the recent surge in U.S. Patenting: Changes in R&D expenditures, patent yields, and the high tech sector," *Economics of Innovation and New Technology*, Vol. 13, 543-58.

Kim, Jinyoung and Marschke, Gerald, 2005. "Labor Mobility of Scientists, Technological Diffusion, and the Firm's Patenting Decision," *The Rand Journal of Economics*, 36 (2), pp. 298-317

Kogut, B. and Zander, U. (1993) "Knowledge of the Firm and the Evolutionary Theory of the Multinational Corporation," *Journal of International Business Studies*, 24(4), pp. 625-645.

Kortum, S., and Lerner, J. (1999) What is Behind the Recent Surge in Patenting? *Research Policy*, 28(1), 1-22.

Kortum, S. and Lerner, J. (2003) "Unraveling the Patent Paradox." Presented at the American Economic Association Meetings.

Moen, Jarle. (2005) "Is Mobility of Technical Personnel a Source of R&D Spillovers?" *Journal of Labor Economics*, 23(1), pp. 81-114.

Mowery, David C. and Ziedonis, Arvids A. (October 2001). "The Geographic Reach of Market and Non-Market Channels of Technology Transfer: Comparing Citations and Licenses of University Patents" NBER Working Paper No. W8568.

National Archives and Records Administration. (1995). General Information Leaflet 55.

Polanyi, Michael. *Personal Knowledge: Towards a Post-Critical Philosophy* (University of Chicago Press, Chicago, 1958)

Polanyi, Michael. *The Tacit Dimension* (Doubleday, Garden City, NY, 1966)

Regets, Mark C. (June 2007). "Research Issues in the International Migration of Highly Skilled Workers: A Perspective with Data from the United States," Division of Science Resources Statistics, National Science Foundation Working Paper, SRS 07-203.

Singh, Jasjit. (2007) "Asymmetry of Knowledge Spillovers between MNCs and Host Country Firms," *Journal of International Business Studies,* 38, pp. 764-786.

Thompson, Peter and Fox-Kean, Melanie. (2005). "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment." *American Economic Review*, 95(1), pp. 450-460.

Trajtenberg, M. (Spring 1990). "A Penny for Your Quotes: Patent Citations and the Value of Innovations" *Rand Journal of Economics, 21(1),* 172-187.

Trajtenberg, Manuel; Shiff, Gil; and Melamed, Ran. (2006). "The Names Game: Harnessing Inventors' Patent Data for Economic Research" NBER Working Paper No. 12479, August 2006.

U.S. Patent and Trademark Office User's Guide. (October 2001). "Patents BIB: Selected Bibliographic Information from U.S. Patent Grant Publications and U.S. Patent Application Publications 1969 to Present," Draft.

Von Hippel, E. (1988). *The sources of innovation.* New York: Oxford University Press.

Zucker, Lynne G.; Darby, Michael R.; Armstrong, Jeff S. "Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance in Biotechnology" NBER Working Paper No. 8499, October 2001

Zucker, L. G., Darby,M. R., & Brewer,M. B. (1998). Intellectual capital and the birth of U.S. biotechnology enterprises. *American Economic Review*, *88*, 290–306.

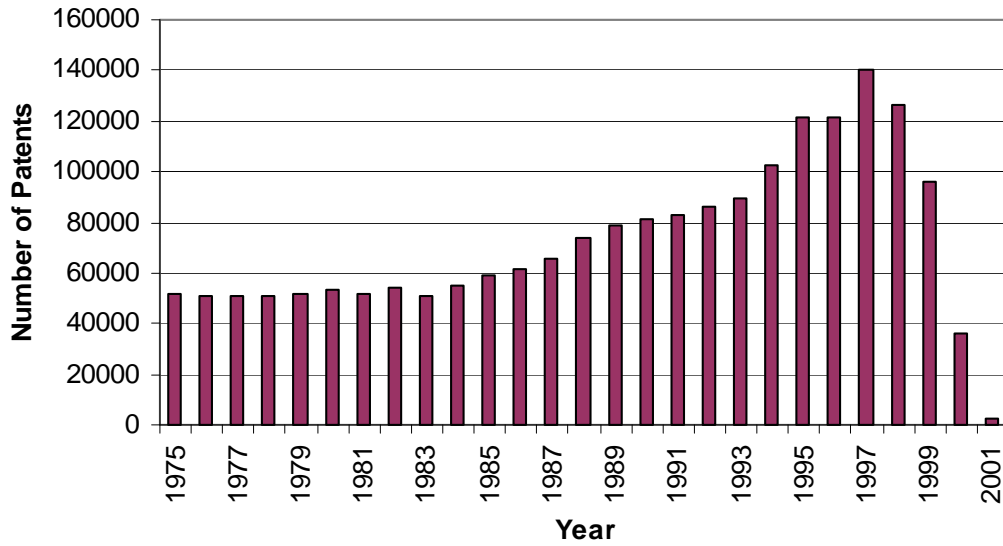**Figure 1. Number of Patents Granted by Year of Application (1975-2001)**

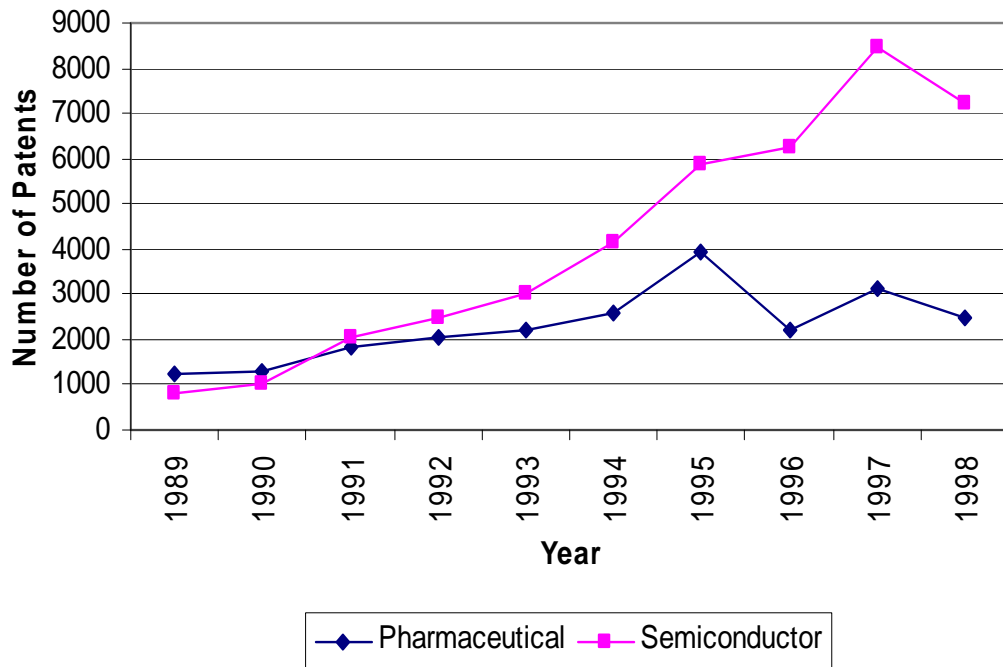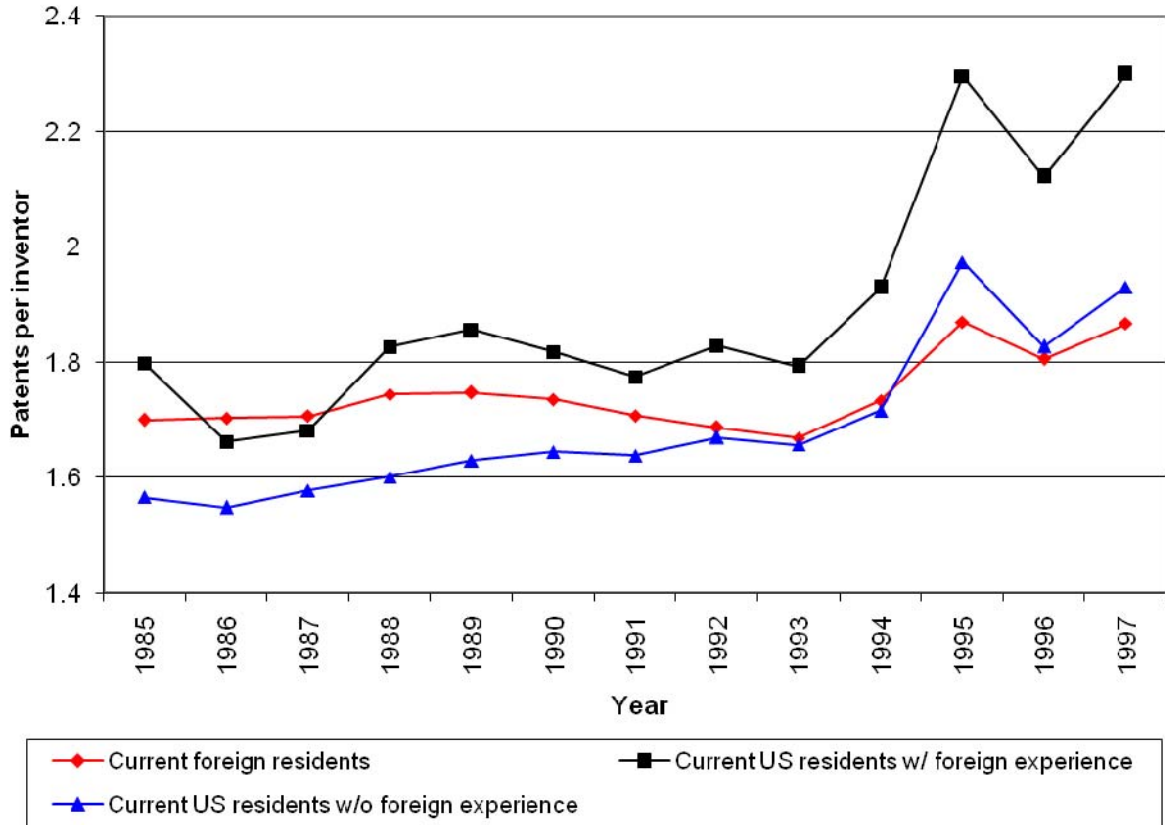**Figure 2. Number of Patents Granted by Year of Application in Two Industries**

Figure 3. Patent-Inventor Ratio by Foreign-Experience type

A. All Patents
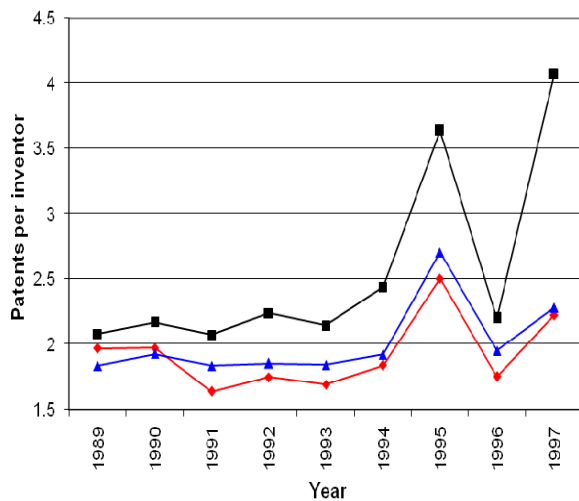
B. Pharmaceutical
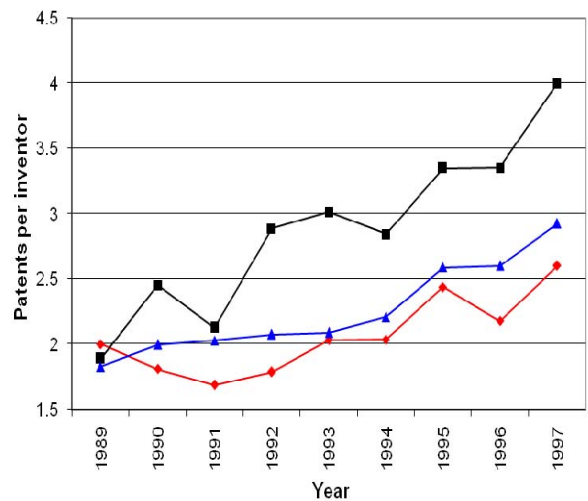
C. Semiconductor

Figure 4. Citations per Patent by Foreign-Experience type

A. All Patents
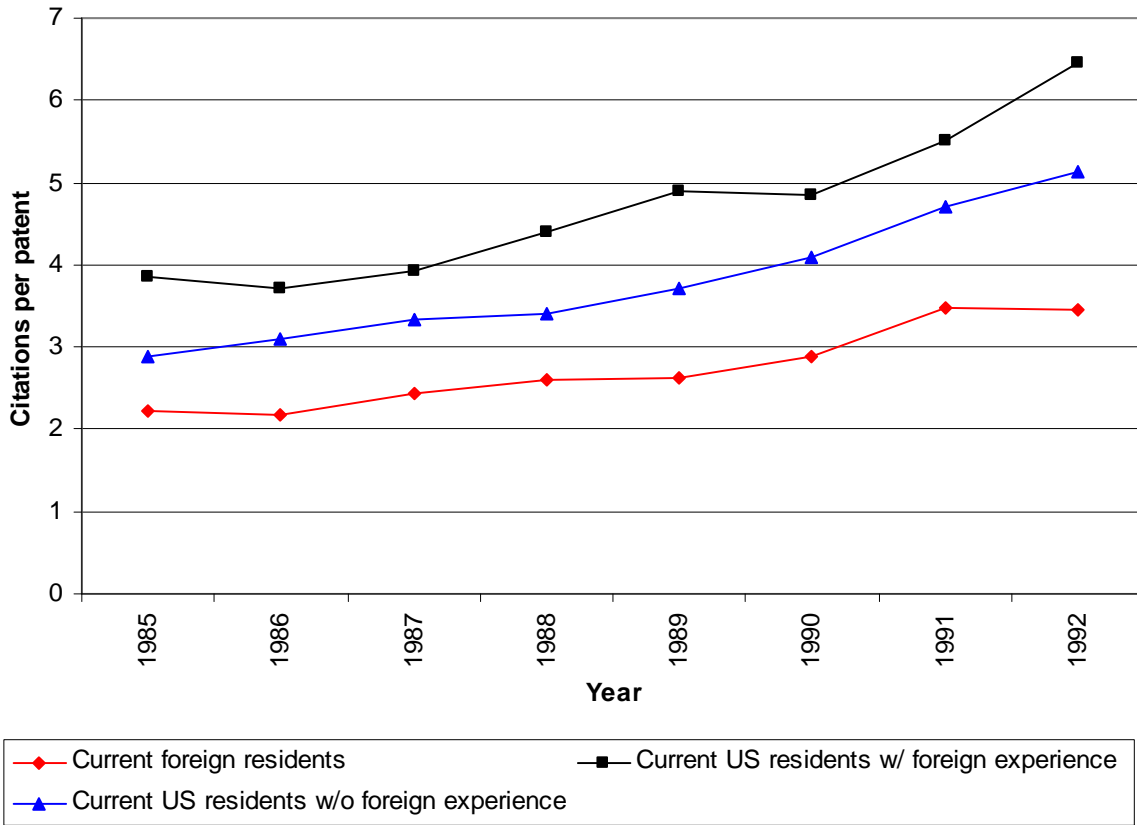
B. Pharmaceutical
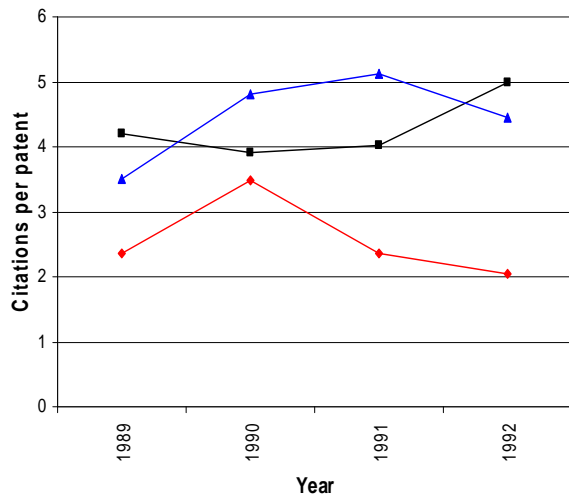
C. Semiconductor
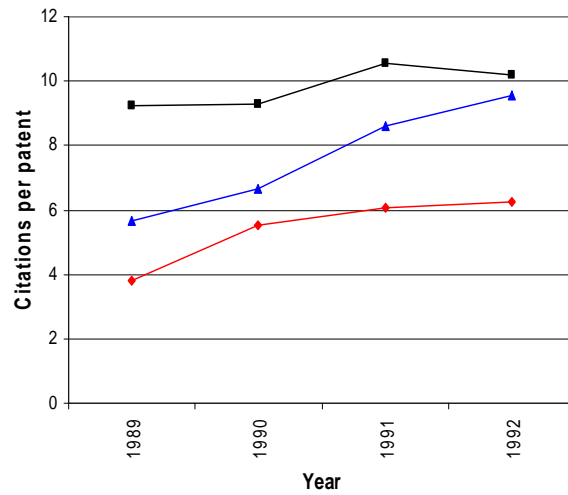
Table 1. Variables from Each Data Source

Data Source	Variables
Patents BIB	Patent ID number, application year, inventors' names, address, city, state, country, assignee ID and assignee name
Compact D/SEC	Firm name, primary and other corresponding SIC codes, R&D expenditures, sales, number of employee, capital, and subsidiaries of the firms
S&P	Firm name and ownership changes due from merger and acquisition, and obsolete securities due to bankruptcy or dissolution
Thomas Register	Founding year of firm
Citation	Citing patent number and cited patent number

Table 2. Examples of Name Matching

Initial ID	Inventor name	Co-inventor	Middle name	ZIP	Final ID
001	Adam Smith	John Keynes		20012	001
002	Adam Smith	John Keynes	Emmanuel	14228	001
003	Adam Smith		E	14228	001
004	Adam Smith		Emmanuel	14214	001
005	Adam Smith	John Keynes	J	14228	005
006	Adam Smyth	John Keynes		14228	001

Table 3. Number of Patents by Assignee Type (January 1975 – February 2002)

Assignee	Description	# Observations	Percentage	
US	Assigned to U.S. organization and state/local governments	1,090,194	43.7	45.5
	Assigned to a U.S. resident (individual)	15,849	0.6	
	Assigned to a U.S. Federal Government organization	30,431	1.2	
Foreign	Assigned to a non-U.S., non-government organization	914,826	36.7	37.4
	Assigned to a non-U.S. resident (individual)	7,873	0.3	
	Assigned to a non-U.S. government organization (all levels)	8,613	0.4	
Others	Unassigned	412,621	16.6	17.1
	Missing observations	13,203	0.5	
Total		2,493,610		

Table 4. Patent Statistics for All Inventors
(January 1975 – February 2002)

	Total	Pharmaceutical	Semiconductor
Inventors	2,299,579	25,609	33,683
No. of Patent per Inventor	2.22	2.80	2.60

Table 5. Summary Statistics from the Pharmaceutical and Semiconductor Industry Samples

(Units in sales and R&D: $1000)

	Year	No. of firms	No. of firms reporting			Mean			Standard Deviation		
			employment	sales	R&D	employment	sales	R&D	employment	sales	R&D
	1989	88	85	78	69	5903	895924	85151	13058	1940353	180615
Pharmaceutical	1990	88	85	81	64	5722	794134	78612	13726	2036458	188652
Industry	1991	146	137	124	98	4741	884836	95712	12690	2187517	217398
	1992	151	145	123	109	4694	987210	101187	12374	2404976	237929
	1993	161	155	132	126	4297	1609557	105501	11764	7440985	250355
	1994	179	170	149	136	4668	1670350	104193	13395	7693383	255514
	1995	184	171	150	142	4460	1924112	124575	13263	7897844	323828
	1996	193	170	158	152	4114	1996128	126510	12652	8085530	355602
	1997	221	196	193	177	4078	2161856	194237	13980	9156677	797319
	1998	209	180	186	170	4391	2341263	210768	14307	10268340	830591

	Year	No. of firms	No. of firms reporting			Mean			Standard Deviation		
			employment	sales	R&D	employment	sales	R&D	employment	sales	R&D
	1989	71	70	71	55	3108	275944	25530	10043	885337	64583
Semiconductor	1990	67	65	67	52	3273	309869	30203	10102	959953	82405
Industry	1991	87	86	84	70	3492	410065	28720	13365	1521497	81478
	1992	93	92	91	79	3244	423890	28546	13226	1698277	94122
	1993	107	107	103	95	2919	477073	28395	13307	2044170	105515
	1994	114	112	108	100	2057	590914	33344	6922	2588848	119556
	1995	131	124	127	115	2050	720290	43822	7156	3103003	151508
	1996	136	123	131	122	3201	746105	62304	14466	3243984	215290
	1997	151	141	147	135	3277	1081964	87658	14750	4956544	336703
	1998	154	125	153	139	3328	1095652	102042	14013	5255021	392838

Table 6. Inventors on U.S. Domestic Patents with Foreign Experience

| Year | Number of Inventors | | | Percentage of Inventors by Foreign-Experience Type (%) | | | | | | | | |
| | | | | Current Foreign Residents | | | Current US Residents w/ Foreign Experience | | | Current US Residents w/o Foreign Experience | | |
	All	Pharma	Semi	All	Pharma	Semi	All	Pharma	Semi	All	Pharma	Semi
1985	42,368			8.15%			0.99%			90.86%		
1986	44,828			8.30			1.07			90.63		
1987	48,810			8.21			1.13			90.66		
1988	54,947			8.49			1.13			90.37		
1989	59,164	2,143	1,139	8.60	14.47	9.04	1.17	2.01	1.14	90.23	83.53	89.82
1990	63,812	2,259	1,362	8.02	17.35	7.78	1.22	1.51	1.25	90.76	81.14	90.97
1991	67,657	3,332	2,791	7.76	19.09	6.02	1.26	1.23	1.22	90.98	79.68	92.76
1992	73,640	3,876	3,370	7.86	20.38	7.15	1.30	1.21	1.13	90.85	78.41	91.72
1993	80,428	4,505	4,190	8.06	25.88	7.06	1.21	1.31	1.03	90.73	72.81	91.91
1994	90,910	5,320	5,739	8.44	26.86	14.76	1.20	0.98	0.94	90.36	72.16	84.30
1995	104,775	6,629	7,450	8.78	28.87	15.18	1.13	0.87	0.86	90.08	70.25	83.96
1996	104,829	4,894	7,916	9.19	31.55	13.26	1.07	0.90	0.78	89.75	67.55	85.95
1997	119,556	6,093	9,993	9.11	29.71	15.31	1.01	0.75	0.80	89.87	69.54	83.89

Note: Columns 2-4 show the number of unique inventors in all U.S. domestic patents, in pharmaceutical patents, and in semiconductor patents, respectively. In columns 8-10, we report the percent of inventors with current addresses in the U.S. who have at least one patent in the past 10 years while residing at a foreign address.

Table 7. Variable Definitions and Sample Statistics

	Definition	Mean (Standard Deviation)	
		Pharmaceutical	Semiconductor
CITE_FRGN	Fraction of citations to patents that are assigned to foreign assignees	0.5505 (0.3319)	0.4760 (0.2850)
FRGN_EXP	= 1 if at least one inventor is residing or has resided in the past in one of the foreign countries where foreign assignees of cited patents are located	0.0734 (0.2609)	0.0290 (0.1677)
INVENTOR	Number of all inventors in the patenting firm	326.0 (195.7)	923.5 (728.6)
EMPLOYEE	Number of employees in the patenting firm	35,979 (21,833)	41,538 (52,501)
R&D/INV	Real R&D expenditures in 1996 constant dollars divided by the number of inventors in the patenting firm (thousands of dollars per inventor)	31.67 (24.51)	12.04 (27.34)
NSIC	Number of secondary SIC's assigned to the patenting firm	3.791 (1.991)	3.154 (1.944)
MEXP	Median experience of all inventors in the patenting firm where experience is measured as the number of years elapsed after the application year of an inventor's first patent	5.292 (1.582)	3.832 (1.067)
FIRMAGE	Years elapsed since the founding year of the patenting firm	77.40 (51.51)	36.17 (23.40)

Table 8 Determinants of Citations to Foreign-Assigned Patents

Dependent variable = CITE_FRGN

	Pharmaceutical			Semiconductor		
	(1)	(2)	(3)	(1)	(2)	(3)
FRGN_EXP	0.2295	0.2225	0.2299	0.2615	0.2514	0.2899
	8.54	7.48	5.77	4.91	4.43	4.70
Log INVENTOR		0.0366	0.0415		-0.0566	-0.0546
		0.97	1.08		-2.75	-2.66
Log EMPLOYEE		0.0200	0.0172		0.0167	0.0139
		0.92	0.76		0.87	0.74
Log R&D/INV		0.0019	0.0015		0.0028	0.0029
		0.56	0.43		0.95	0.98
Log NSIC		0.0061	0.0017		0.0490	0.0487
		0.19	0.05		1.46	1.44
Log MEXP		-0.3289	-0.3260		-0.3212	-0.3168
		-5.67	-5.57		-2.73	-2.61
Log FIRMAGE		-0.0505	-0.0516		0.1151	0.1160
		-2.54	-2.54		2.95	2.91
Observations	1430	1247	1215	4316	4186	4112
R^2	0.0325	0.1794	0.1772	0.0237	0.1157	0.1165

Note: Rows show the estimated coefficient and the t statistic for each regressor. The result for a constant term is suppressed. Column 3 shows the results from a regression that omits patents for which an inventor is listed as an inventor on a cited patent. The t statistic is based on the Huber-White sandwich estimator of variance.