This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: Science and Engineering Careers in the United States: An Analysis of Markets and Employment

Volume Author/Editor: Richard B. Freeman and Daniel L. Goroff, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-26189-1

Volume URL: http://www.nber.org/books/free09-1

Conference Date: October 19-20, 2005

Publication Date: June 2009

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

Chapter Author: Jinyoung Kim, Sangjoon John Lee, Gerald Marschke

Chapter URL: http://www.nber.org/chapters/c11626

Chapter pages in book: (p. 321 - 348)

10

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

Jinyoung Kim, Sangjoon John Lee, and
Gerald Marschke

10.1 Introduction

This chapter uses U.S. patent records to examine the nature and extent of knowledge spillovers from outside of the United States 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

Jinyoung Kim is a professor of economics at Korea University. Sangjoon John Lee is an assistant professor of economics at Alfred University. Gerald Marschke is an associate professor of economics at SUNY Albany and a research fellow of the NBER and IZA. Kim was supported by the Korea Research Foundation Grant funded by the Korean Government (KRF-2007-332-B00062).

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.

(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 United States 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.

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; 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 payoff to 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 previously mentioned literature is at the least suggestive of the im-

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 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.

portance 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 is the firm to which the patent is assigned and the assignee's nationality 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 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 and semiconductor firms are going to foreign countries to employ such researchers as opposed to such researchers immigrating to the United States to work for U.S. firms. We find, however, that inventors migrating to the United States 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 10.2 and 10.3 describe the sources for the data construction and the construction itself. Section 10.4 details some descriptive statistics of the data set. Section 10.5 describes our analysis on the influence of foreign R&D on U.S. innovation. Section 10.6 concludes.

10.2 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 Standard Industrial Classification [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.

3. Based on NBER-Case Western University data of U.S. patents from 1963 to 1999, 15.3 percent of industry patents were granted to the firms in the pharmaceutical industry and 14.8 percent 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.

Table 10.1 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 employees, 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

The data for this study come from five sources: (a) 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; (b) the Compact Disclosure/Securities and Exchange Commission (D/SEC) database from 1989 to 1997, which contains firm information taken primarily from 10-K reports filed with the Securities and Exchange Commission; (c) 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; (d) the Thomas Register, Mergent, and Corptech data, which report a firm's founding year, and finally (e) the National Bureau of Economic Research (NBER) Patent-Citations data collected by Hall, Jaffe, and Trajtenberg (2001), which contain all citations made by patents granted from 1975 to 1999. These data sources are described in detail following and the variables used in our study from each data source are in table 10.1.

10.2.1 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 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 United States 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).

10.2.2 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.

10.2.3 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.

10.2.4 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 (2001a, 2001b) 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.

10.3 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 nonuniqueness 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.

10.3.1 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 or 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: (a) the full addresses for the pair of names are the same; (b) 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 (c) the two names from the pair share the same coinventor. 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 the appendix 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: (a) the two have the same zip code; (b) they have the same full middle name; or (c) they reside in the same metropolitan statistical area (MSA). As an additional step beyond the aforementioned pairwise comparisons, we treat a pair of inventors as mismatched if the middle name initials of the pair are different.

Table 10.2 illustrates our name-matching procedure. Inventors 001 and 002 in table 10.2 have the same last and first names, and share the same coinventor. 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.

Table 10.2 Examples of name matching

Initial ID	Inventor name	Coinventor	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

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 are the first names of the next coinventors 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, Shiff, and Melamed (2006) 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, Shiff, and Melamed (2006). We identified 1.72 million unique inventors (34 percent) out of 5.1 million names in the entire patent data, while Trajtenberg, Shiff, and Melamed found 1.6 million distinctive inventors (37 percent) out of 4.3 million names. Note that our patent database is larger because it includes additional years, 2000 to 2002.

Matching Accuracy

We evaluated the accuracy of the algorithm using the curricula vitae (CV) 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 percent) of the 769 patents that belong to the 100 inventors in the CV sample. In other words, a reassignment of as few as eighty-four 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 thirty-eight 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 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.

5. The inventors in our CV sample have, on average, over seven 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 two.

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.

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 overassign patents to any of the eleven benchmark inventors with surnames of Asian ancestry, possibly because there is sufficient variation in their first names, which are also used in the matching.

10.3.2 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.

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, if a firm is reported as a subsidiary of another firm once between the years 1989 to 1997, we consider it a subsidiary of that firm for the entire period.

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.

10.4 Descriptive Statistics

Figure 10.1 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 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 10.1 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 percent were granted to U.S. assignees and 37.4 percent were granted to foreign assignees (see table 10.3) with the rest unassigned. In figure 10.2, 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 1,000 patents annually, but by a factor of two in the pharmaceutical industry and nearly seven in the semiconductor industry.

Table 10.4 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 10.4). 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.

We identified pharmaceutical and semiconductor firms in the Compact

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.

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 previously, 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

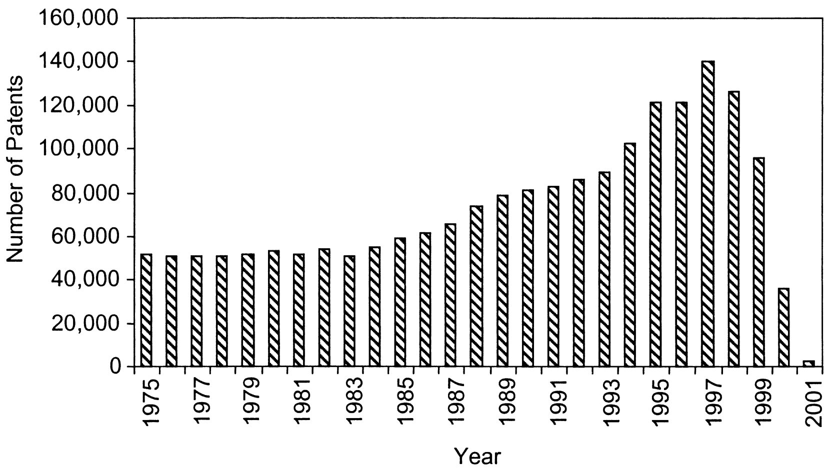

Fig. 10.1 Number of patents granted by year of application (1975–2001)

Table 10.3 Number of patents by assignee type (January 1975–February 2002)

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

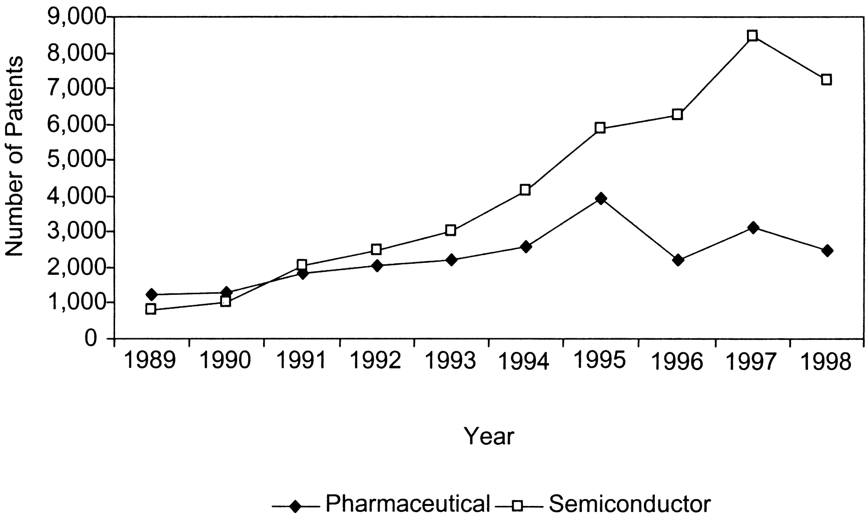

Fig. 10.2 Number of patents granted by year of application in two industries

Table 10.4 Patent statistics for all inventors (January 1975–February 2002)

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

expenditures—are reported in table 10.5. 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.

10.5 International Knowledge Flows

A small literature uses patents to examine international knowledge flows. Jaffe and Trajtenberg (1999) find, for example, that patents are more likely to cite other patents from the same country and that citations to other countries' 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

Table 10.5 **Summary statistics from the pharmaceutical and semiconductor industry samples (Units in sales and R&D: $1,000)**

		No. of firms reporting			Mean			Standard deviation		
Year	No. of firms	Employment	Sales	R&D	Employment	Sales	R&D	Employment	Sales	R&D
					Pharmaceutical industry					
1989	88	85	78	69	5,903	895,924	85,151	13,058	1,940,353	180,615
1990	88	85	81	64	5,722	794,134	78,612	13,726	2,036,458	188,652
1991	146	137	124	98	4,741	884,836	95,712	12,690	2,187,517	217,398
1992	151	145	123	109	4,694	987,210	101,187	12,374	2,404,976	237,929
1993	161	155	132	126	4,297	1,609,557	105,501	11,764	7,440,985	250,355
1994	179	170	149	136	4,668	1,670,350	104,193	13,395	7,693,383	255,514
1995	184	171	150	142	4,460	1,924,112	124,575	13,263	7,897,844	323,828
1996	193	170	158	152	4,114	1,996,128	126,510	12,652	8,085,530	355,602
1997	221	196	193	177	4,078	2,161,856	194,237	13,980	9,156,677	797,319
1998	209	180	186	170	4,391	2,341,263	210,768	14,307	10,268,340	830,591
					Semiconductor industry					
1989	71	70	71	55	3,108	275,944	25,530	10,043	885,337	64,583
1990	67	65	67	52	3,273	309,869	30,203	10,102	959,953	82,405
1991	87	86	84	70	3,492	410,065	28,720	13,365	1,521,497	81,478
1992	93	92	91	79	3,244	423,890	28,546	13,226	1,698,277	94,122
1993	107	107	103	95	2,919	477,073	28,395	13,307	2,044,170	105,515
1994	114	112	108	100	2,057	590,914	33,344	6,922	2,588,848	119,556
1995	131	124	127	115	2,050	720,290	43,822	7,156	3,103,003	151,508
1996	136	123	131	122	3,201	746,105	62,304	14,466	3,243,984	215,290
1997	151	141	147	135	3,277	1,081,964	87,658	14,750	4,956,544	336,703
1998	154	125	153	139	3,328	1,095,652	102,042	14,013	5,255,021	392,838

advanced. He 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 United States 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 United States. In addition, we test if the international migration of researchers facilitates knowledge transfers across borders.

Table 10.6 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 (a) resided in a foreign country, (b) resided in the United States and had been previously listed as a foreign-residing inventor on a successful patent application, and (c) resided in the United States 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 cutoff 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.

Table 10.6 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 dropped 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 10.6 shows that the growth in the number of inventors in the pharmaceutical (13 percent annually) and semiconductor (31 percent annually)

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

Table 10.6 **Inventors on U.S. domestic patents with foreign experience**

				Percentage of inventors by foreign-experience type (%)								
	Number of inventors			Current foreign residents			Current U.S. residents w/ foreign experience			Current U.S. residents w/o foreign experience		
Year	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

Notes: 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 ten years while residing at a foreign address.

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 et al. 2004). The semiconductor industry shows a similar pattern, but the changes are less pronounced than in pharmaceutical industry.[12]

Figure 10.3, panel A shows the average annual patent productivity of inventors in U.S. domestic patents by foreign-experience type for all patents. Panels B and C of figure 10.3 repeat the analysis of panel A, but for the pharmaceutical and semiconductor industries alone. The calculations in 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 ten-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 United States, 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 10.6, 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.

Figure 10.3, panels B and C, 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. Panel C of figure

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

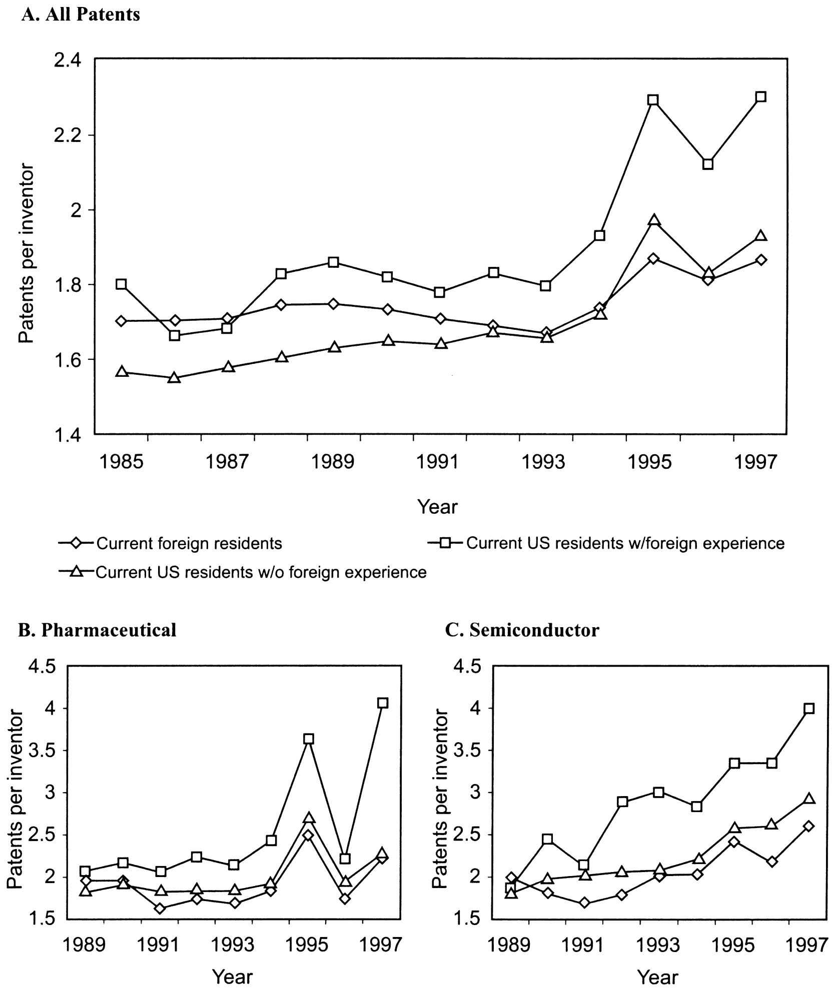

Fig. 10.3 Patent-inventor ratio by foreign-experience type

10.3 more clearly demonstrates the changes in patent productivity across the types of inventors in the semiconductor industry.

Where figure 10.3, panels A through C track the productivity of inventors by their patent output, figure 10.4, panels A through C track how the quality of inventors' output changes by inventor type. There is evidence that citations received reflect the economic value of the patent (Trajtenberg 1990). Figure 10.4, panel A shows the citations received in the five-year period following application per patent by inventor type for all industries over time. This figure covers only years of application through 1992 be-

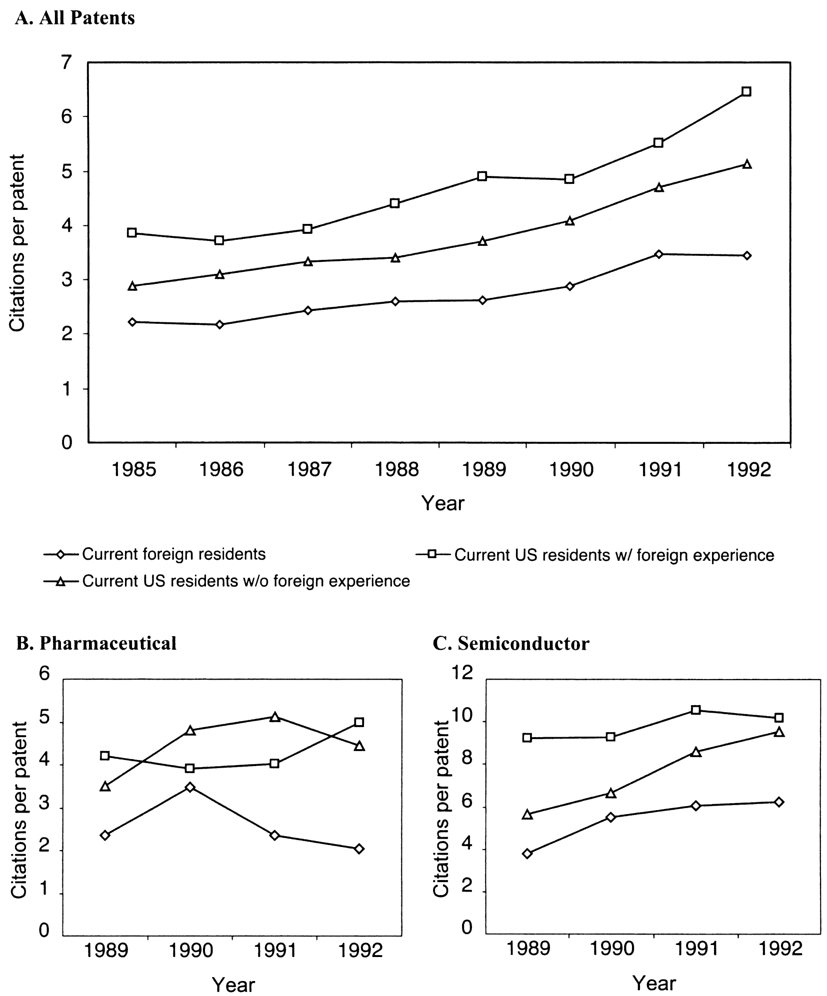

Fig. 10.4 Citations per patent by foreign-experience type

cause the NBER citation data contain citations made by patents granted in years up to 1999 and we take into account the five-year period of citation and a two-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 to 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 10.3, panel A, and 10.4, panel A, 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. Figure 10.4, panels B and C, 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 10.4, panel C, does show that foreign-residing inventors produce the lowest quality patents, as measured by citations, 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 and into the 1990s, U.S.-residing inventors with foreign experience have higher patent rates. In the semiconductor industry, their patents also attract more citations.

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 10.7 presents the results of our estimation of the determinants of accessing foreign

Table 10.7 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	1,430	1,247	1,215	4,316	4,186	4,112
R^2	0.0325	0.1794	0.1772	0.0237	0.1157	0.1165

Notes: 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.

knowledge in the pharmaceutical and semiconductor industry. The unit of observation in 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 means and standard deviations of the independent and dependent variables, along with their definitions, are described in table 10A.1.[13] The key regressor in these regressions is a binary variable that 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 10.7 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).

Column (1) in table 10.7 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 ordinary least squares. 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 that have the same inventors as those in their cited patents.

The results in table 10.7 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

13. Note that the means of the variables reported in table 10.7 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.

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 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.

10.6 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 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 United States but have foreign experience.

Table 10.6 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 United States 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 10.7, we present evidence that either employing researchers abroad or foreign-experienced researchers in the United States 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 10.7 as evidence that when U.S. innovating firms employ either foreign-experienced researchers in the United States or those at a foreign-based subsidiary, transmission of foreign knowledge from their foreign origins to the United States results. Table 10.6 and 10.7 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, and we also know the kinds of technologies underlying the innovations. This data set will allow us to directly observe the importance of interfirm mobility for technological diffusion. From the perspective of the researcher, this data set will allow us to examine the determinants of interfirm 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

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 nonscoring 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 nonscoring "i" intervenes.

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

Table 10A.1 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 SICs 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)

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 National Archives and Records Administration (1995) for the detailed method.

References

Adams, J. D., G. C. Black, J. R. Clemmons, and P. E. Stephan. 2004. Scientific teams and institution collaborations: Evidence from U.S. universities, 1981–1999. NBER Working Paper no. 10640. Cambridge, MA: National Bureau of Economic Research, July.

Agrawal, A. K., I. M. Cockburn, and J. McHale. 2003. Gone but not forgotten: Labor flows, knowledge spillovers, and enduring social capital. NBER Working Paper no. 9950. Cambridge, MA: National Bureau of Economic Research, September.

Almeida, P., and B. Kogut. 1999. Localization of knowledge and the mobility of engineers in regional networks. *Management Science* 45 (7): 905–17.

Audretsch, D. B., and M. P. Feldman. 1996. R&D spillovers and the geography of innovation and production. *The American Economic Review* 86 (3): 630–40.

Branstetter, L. G. 2001. Are knowledge spillovers international or intranational in scope? *Journal of International Economics* 53 (1): 53–79.

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

Cohen, W. M., R. R. Nelson, and J. P. Walsh. 2002. Links and impacts: The influence of public research on industrial R&D. *Management Science* 48 (1): 1–23.

Duguet, E., and M. MacGarvie. 2005. How well do patent citations measure flows of technology? Evidence from French innovation surveys. *Economics of Innovation and New Technology* 14 (5): 375–93.

Dunning, J. H. 1992. *Multinational enterprises and the global economy.* Wokingham, England: Addison-Wesley.

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

Freeman, R. B. 2005. Does globalization of the scientific/engineering workforce threaten U.S. economic leadership? NBER Working Paper no. 11457. Cambridge, MA: National Bureau of Economic Research, June.

Griliches, Z. 1992. The search for R&D spillovers. *Scandinavian Journal of Economics* 94 (Supplement): 29–47.

Hall, B., Z. Griliches, and J. Hausman. 1986. Patents and R&D: Is there a lag? *International Economic Review* 27 (2): 265–83.

Hall, B., A. Jaffe, and M. Trajtenberg. 2001a. Market value and patent citations: A first look. NBER Working Paper no. 7741. Cambridge, MA: National Bureau of Economic Research, June.

———. 2001b. The NBER patent citation data file: Lessons, insights and methodological tools. NBER Working Paper no. 8498. Cambridge, MA: National Bureau of Economic Research, October.

Hall, B., and R. Ziedonis. 2001. The determinants of patenting in the U.S. semiconductor industry, 1980–94. *RAND Journal of Economics* 32 (1): 101–28.

Hu, A. G. Z., and A. B. Jaffe. 2003. Patent citations and international knowledge flow: The cases of Korea and Taiwan. *International Journal of Industrial Organization* 21 (6): 849–80.

Jaffe, A. B. 1989. Real effects of academic research. *The American Economic Review* 79 (5): 957–70.

Jaffe, A. B., M. S. Fogarty, and B. A. Banks. 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* 8:105–36.

Jaffe, A., M. Trajtenberg, and R. Henderson. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics* 108 (3): 577–98.

Keller, W. 2002. Geographical localization of international technology diffusion. *American Economic Review* 92 (1): 120–42.

Kim, J., and G. Marschke. 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* 13 (6): 543–58.

———. 2005. Labor mobility of scientists, technological diffusion, and the firm's patenting decision. *The RAND Journal of Economics* 36 (2): 298–317.

Kogut, B., and U. Zander. 1993. Knowledge of the firm and the evolutionary theory of the multinational corporation. *Journal of International Business Studies* 24 (4): 625–45.

Kortum, S., and J. Lerner. 1999. What is behind the recent surge in patenting? *Research Policy* 28 (1): 1–22.

———. 2003. Unraveling the patent paradox. Paper presented at the American Economic Association Annual Meeting. Washington, D.C.

Moen, J. 2005. Is mobility of technical personnel a source of R&D spillovers? *Journal of Labor Economics* 23 (1): 81–114.

Mowery, D.C., and A. A. Ziedonis. 2001. The geographic reach of market and nonmarket channels of technology transfer: Comparing citations and licenses of university patents. NBER Working Paper no. 8568. Cambridge, MA: National Bureau of Economic Research, October.

National Archives and Records Administration. 1995. Using the Census Soundex. General information leaflet 55. College Park, MD: National Archives and Records Administration.

Phene, A., and P. Almeida. 2003. How do firms evolve? The patterns of technological evolution of semiconductor subsidiaries. *International Business Review* 12 (3): 349–67.

Polanyi, M. 1958. *Personal knowledge: Towards a post-critical philosophy.* Chicago: University of Chicago Press.

Polanyi, M. 1966. *The tacit dimension.* Garden City, NY: Doubleday.

Regets, M. C. 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, J. 2007. Asymmetry of knowledge spillovers between MNCs and host country firms. *Journal of International Business Studies* 38 (5): 764–86.

Thompson, P., and M. Fox-Kean. 2005. Patent citations and the geography of knowledge spillovers: A reassessment. *American Economic Review* 95 (1): 450–60.

Trajtenberg, M. 1990. A penny for your quotes: Patent citations and the value of innovations. *RAND Journal of Economics* 21 (1): 172–87.

Trajtenberg, M., G. Shiff, and R. Melamed. 2006. The names game: Harnessing inventors' patent data for economic research. NBER Working Paper no. 12479. Cambridge, MA: National Bureau of Economic Research, August.

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

Zucker, L. G., M. R. Darby, and J. S. Armstrong. 2001. Commercializing knowledge: University science, knowledge capture, and firm performance in biotechnology. NBER Working Paper no. 8499. Cambridge, MA: National Bureau of Economic Research, October.

Zucker, L. G., M. R. Darby, and M. B. Brewer. 1998. Intellectual capital and the birth of U.S. biotechnology enterprises. *American Economic Review* 88 (1): 290–306.