Wrapping It Up in a Person: The Mobility Patterns of New PhDs

Paula Stephan

pstephan@gsu.edu

March 2006

Prepared for NBER Innovation, Policy and the Economy Conference, April 2006

The author wishes to thank A. J. Sumell and Daniel Hall for data assistance. Financial support for this project was provided by the Andrew W. Mellon Foundation, the National Science Foundation, grant number 0244268, and the Science and Engineering Workforce Project, National Bureau of Economic Research. Bill Amis, Josh Lerner, Jennifer Ma and Scott Stern made helpful comments on an earlier draft. The author thanks David Kaiser for bringing the Oppenheimer quote to her attention. The use of NSF data does not imply NSF endorsement of the research methods or conclusions contained in this paper.

Abstract

The placement of new PhDs in industry provides one mechanism for transmitting tacit knowledge from universities to industry. This paper analyzes data concerning the placements of new PhDs who had definite plans to go to work in industry for the period 1997-2002. Data come from the Survey of Earned Doctorates overseen by the National Science Foundation.

We find knowledge sources to be heavily concentrated in certain regions and states. Moreover, the geographic distribution of knowledge sources, as measured by where PhDs going to work in industry are trained, is different than other measures of knowledge sources would suggest, such as university R&D-expenditure data. A major headline is the strong role played by Midwestern universities, which educate over 26% of all PhDs going to industry.

We find that only 37% of PhDs trained in S&E stay in their state of training. Stay patterns are particularly low among certain Midwestern states, many of whose students leave the state for employment on the Coasts. One can make the case that as the traditional industrial base of the United States shifts, a highly trained workforce will only be maintained if the Federal government increasingly steps in to provide financial support for graduate education, since state legislatures are unlikely to continue to fund these migration flows.

Firms most likely to hire new PhDs are found in computer and electrical products, followed by firms working in publishing and professional, scientific and technical services. The hiring data highlights the role that PhDs play in local economic development. Almost one out of five new PhDs going to work for industry heads to San Jose; 58% go to work in one of twenty cities. The placement data also suggest that small firms play a larger role in innovation than R&D expenditure data would suggest.

"The best way to send information is to wrap it up in a person"[1]

J. Robert Oppenheimer

Section I. Introduction

The mechanism by which knowledge flows from universities to firms is varied, involving formal means, such as publications, as well as less formal mechanisms, such as discussions between faculty and industrial scientists at professional meetings. Face-to-face transmission is most appropriate when tacit knowledge is involved, since, by definition, tacit knowledge cannot be codified. The placement of new PhDs in industry provides one mechanism for transmitting tacit knowledge. Much of a graduate student's training is of a tacit nature, acquired while working in her mentor's lab. These techniques, wrapped up in new PhDs, can be transmitted to industrial R&D labs when the PhD takes a position there upon graduating.[2]

Despite the role that PhD placements can play in the transmission of knowledge, we know very little about these knowledge flows. For example, we know little about the providence of new PhDs going to industry: What universities do they come from? Where do they go? Do they stay in the area where they were trained? By way of contrast, we know considerably more about the transmission of codified knowledge, due in large part to the citation trail left by both patents and articles which allow one to make inferences concerning patterns of transmission.

---

[1] J. Robert Oppenheimer, as quoted in Anon., "The eternal apprentice," Time magazine, vol. 52 (8 November 1948): 70-81, on p. 81.

[2] Dasgupta and David (1994, p. 511) state that the "export of scientists and engineers from the academy to industrial research is potentially the most important and salutary among the mechanisms available for effecting knowledge transfers."

The reason for this knowledge gap relates to the availability of data. Firm hires of new PhDs are not part of the public record. Nor, and more to the point, do the data collected by the National Science Foundation on new PhDs at the time of graduation capture the industrial destinations of new PhDs. The data has been collected but not coded. For the past four years, we have been coding this data, which, beginning in 1997, became available in verbatim records. We now have six years of data, ending with PhDs granted in 2002. The data are far from perfect, having several "holes." But they give a picture, partial as it may be, about which heretofore little has been known. They show a remarkable fluidity of knowledge flows; they also show that knowledge centers, as defined in terms of PhD production, exist in parts of the country that are no longer known for their industrial strength and that new PhDs working in industry are heavily clustered in certain cities.

Here we summarize findings from the six years of data that have just become available. In addition, we explore insights that human resource data can bring to the study of innovation, following up on a presentation that Stephan (2002) made at the National Research Council where she argued that human resource data could provide a lens for tracking innovation.

The plan of this paper is as follows: In Section II we describe the data. In Section III we explore issues related to geography. Where do the new PhDs come from? Where do they go? What do the patterns say in terms of the role of proximity in the transmission of knowledge spillovers? Section IV examines insights gained by using human resource data to illuminate patterns of innovation. We examine, for example, the industrial mix of hires, how hiring patterns changed between the two periods, and the diversity of fields hired within a given industry. Data issues are discussed in Section V. Conclusions are drawn in Section VI.

Section II. The Data

Since 1958 new PhDs at or near the time of graduation have been asked to complete the Survey of Earned Doctorates (SED), which is overseen by the National Science Foundation, Science Resources Statistics (SRS). The response rate has historically been quite high and is currently around 92%. Respondents are asked a number of questions concerning their training and field of work as well as plans subsequent to graduation.[3] Of particular interest for this study is the question that asks the recipient to "name the organization and geographic location where you will work or study." Although this question has been asked for many years, for those going to industry the names of firms, as well as the location of employment, have not been coded by NSF and have only been available in verbatim form since 1997.[4] As part of a larger project, we have coded the verbatim records by firm name and location for the six-year period 1997-2002. We have also coded whether the hiring firm is a top-200 R&D firm or a subsidiary of a top-200 R&D firm.[5] The data were coded for two different periods reflecting when the data became available. Period One covers 1997-1999 and Period Two covers 2000-2002. The number of new PhDs with definite plans to work in industry is remarkably similar between the two periods: 10,932 for period one and 10,833 for

[3] The most recent questionnaire is available at http://www.norc.uchicago.edu/issues/SED_Quex_05-06.pdf.

[4] By way of contrast, for those going to academe, the institution of higher education has been coded for many years.

[5] The top-200 firm list was updated between Period One and Period Two.

period two. This represents 14.6% of degrees in S&E in period one and 15.2% in period two.

These numbers undercount placements of newly minted PhDs going to work in industry because a number of PhDs who take a job in industry do not have definite plans at the time they fill out the questionnaire. During Period One, 17,382 indicated that they planned to work in industry; thus the 10,932 with definite plans represents approximately 63% of those with plans to work in industry; during period two, 17,054 indicated that they planned to work in industry; thus the 10,833 represents a comparable 63%.[6]

The data on definite plans also undercounts placements of *recent* PhDs who work in industry but initially take a postdoctoral position upon graduating. This is particularly the case in the life sciences, where the percent of new PhDs taking a postdoctoral training position upon graduation exceeds 50%; yet approximately one-in-three of these postdocs eventually ends up working in industry.[7]

Some indication of the undercount is given by comparing the percentage of PhDs who reported working in industry four years after completing their PhD to the percentage with definite plans to work in industry at the time they received their PhD. Such a comparison shows that, although there is variation by field, about three times as many doctorates end up working in industry as do those who specify a firm at the time of

[6] This undercounting does not affect our conclusion unless at the time of the survey those with definite plans differ significantly from those without definite plans.
[7] The estimate for the percent of postdocs in biology who eventually take a position in industry comes from the 1995 Survey of Earned Doctorates, which contained retrospective questions concerning postdoctoral experience.

graduation.[8] Despite these limitations, much can be learned from analyzing the SED firm placement data.

Table 1 presents a summary of the data, showing (a) the number and percentage of all new PhDs in a field who had definite plans to work at a firm and (b) the number and percentage who identified a top-200 R&D firm or its subsidiary. Given that the underlying strength of the economy, especially in the high tech area, varied during the six-year period, the data are presented separately for the two periods. The slightly lower number of PhDs produced during Period Two compared to Period One in science and engineering undoubtedly reflects in part the strong market for non-PhD employment in science and engineering during the 1990s, especially in engineering, math and computer science. Only in the field of biology and medicine did the number of degree recipients increase, and then only marginally.

We see from Table 1 that the industrial placement rate of new PhDs is highest among engineers followed by computer scientists and chemists. This reflects underlying patterns among seasoned PhDs, where over 50% of both engineers and chemists work in industry. The field with the lowest percentage going to industry directly out of graduate school is biology. This is not surprising, given the extraordinarily high prevalence of academic postdoctoral positions in the life sciences and the relatively small percentage of seasoned biologists, compared to seasoned PhDs in other fields, working in industry.[9]

[8] The comparison made was between the percentage of 1995 PhDs who reported working in industry in 1999 (using the Survey of Earned Doctorates) and the percentage of Period One PhD placements in industry.

[9] In 1999 approximately 25% of all PhDs in the life sciences were working in industry compared to slightly over 50% in chemistry and in engineering, 30% in math and computer science, and 35% in physics and astronomy (Stephan et al, 2004).

Approximately 38% of the newly hired PhDs go to work at a top-200 R&D firm (or subsidiary) but there is considerable variation across fields. Relative to the underlying benchmark, engineers, chemists and computer scientists are most likely to work at large research-intensive firms. Biologists, those with degrees in agriculture, and psychologists and economists are least likely to work for large firms. The biology figure of 24% for the six years undoubtedly reflects the employment opportunities available in small start-up firms in biotechnology, many of which have a direct relationship with the campus where the individual trained.

Period Two comprises those who entered the labor market after the dot.com bust and during a period of recession. This depressed environment is no doubt responsible for the lower number of firm placements of new PhDs in engineering, computer science, and math. In two fields, however, the actual number placed (as well as the placement rate) rose considerably: In biology the number increased from 609 to 843 and in chemistry the number rose from 1216 to 1310. The underlying increase in biology (where the number of PhDs produced during the two periods remained almost constant) meant that the placement rate increased from 3.8% to 5.2%. While this is still a miniscule rate, it undoubtedly reflects the growing realization among doctoral students in the life sciences that industry, especially pharmaceuticals, represents a relatively favorable employment environment and reflects also the expansion of pharmaceutical firms during this period.[10] The underlying decline in PhD production in chemistry, coupled with an increase in the number of industrial placements, meant that the placement rate in chemistry increased substantially, going from 18.7% to 22.2%.

[10] The amount pharmaceutical industries spent on R&D grew considerably during the period 1999 to 2001. For example, Johnson & Johnson's R&D increased by 38.1%, Merck's by 18.8%, Lilly's by 25.3%, and Pharmacia's by 70.2% (National Science Board, 2004, p. 4-22.

The rate of those taking jobs at top-200 R&D firms is approximately the same in the two periods. But there are some noticeable differences, especially in the small fields of agriculture and astronomy. We also see that the number and percent of computer scientists going to work at large R&D firms decreased, undoubtedly a reflection of market conditions in the field after the dot.com bust.

Section III. Knowledge Sources and the Question of Proximity

Knowledge sources, by region of country where trained, are presented in Table 2.[11] Many of the PhDs going to work in firms are educated in geographic centers associated with innovation. For example, one in four is educated in New England and the Middle Atlantic states; about one in six is educated in the Pacific states. But the headline here is the extraordinarily strong role Midwest institutions (East North Central and West North Central) play, educating 26.5% of those going to industry.

Public knowledge sources are often measured in terms of university R&D expenditure data. Column 3 of Table 2 shows the distribution of these expenditures by region. A comparison of column 3 with column 2 indicates that public knowledge sources as measured by human resource flows to industry are concentrated in somewhat different geographic regions from those that university R&D expenditure data would suggest, and the differences are statistically significant. For example, the South Atlantic region produces about 15% of those going to industry but accounts for 19% of university R&D expenditures; the East North Central produces 19.6% of new PhDs going to industry but accounts for only 14.4% of university R&D. We conclude that the spatial distribution of knowledge sources embodied in newly minted talent is somewhat different

[11] Regions are defined in the Appendix.

from the distribution of knowledge sources stemming from university research, as measured by university R&D expenditure data. Part of this difference may be an artifact of our inability to count new PhDs who go to industry after taking a postdoctoral position, but this is unlikely to account for the striking differences in the mid-west.

The top-20 universities training PhDs hired by firms are given in Table 3. We see that the knowledge sources are quite concentrated; the top 20 educate 40% of those going to industry; the top ten educate 25%. Again we see the important role that the Midwest plays. Seven of the top twenty institutions are in the Midwest; five of the top-ten institutions are in the Midwest. The dominant role played by California is also evident. Four of the top-20 universities are in California.

Considerable research has focused on the role that geographic proximity plays in transmitting knowledge. Early work by Jaffe (1989), for example, used university research and development expenditures as a proxy for the availability of local knowledge spillovers as did work by Audretsch and Feldman (1996a, 1996b). More recent work by Feldman and Audretsch (19999), Anselin, Varga and Acs (1997, 2000) and Black (2001) has followed suit, shifting the analysis from the state to the CMSA. In each study a significant relationship is found between the dependent variable, which is a measure of innovation, and the proxy measure for local knowledge.

These and countless other studies go a long way toward establishing that geographic proximity promotes the transmission of knowledge. They do not, however, address the extent to which knowledge spillovers are local. One of the few papers to examine this question was written by Audretsch and Stephan (1996) and examines academic scientists affiliated with biotech companies. Because the authors knew the location of both the scientist and the firm, they were able to establish the geographic origins of spillovers embodied in this knowledge-transfer process. Their research shows that although proximity matters in establishing formal ties between university-based scientists and companies, its influence is not overwhelming. Approximately 70% of the links between biotech companies and university-based scientists in their study were non-local.

Knowledge sources and knowledge destinations, as proxied by PhD flows, are given in Table 4 by region. The table can be used to examine the question of the degree to which spillovers, as proxied by the employment location of newly trained PhDs, are local. Entries that lie on the diagonal represent "local" links, showing those who take employment within their region of training. Here we find that 48% of the entries lie on the diagonal. There is considerable variation by region, however. The Pacific Region retains slightly over 70%; and the Mid-Atlantic is second, retaining 51%; New England is a close third with a 46% retention rate. By way of contrast, the East South Central region retains only 32%. The East North Central—with its heavy production of new PhDs--retains 38%; and its sister region, the West North Central, retains 34%.

Appendix A.1 drills down to the state level, showing training, employment, and retention patterns (where confidentially permits) by state. Compared to the mean state retention rate of 37.1%, the Midwest states are low: Iowa retains 13.6% of those it trains; Indiana retains only 11.8% of the 771 PhDs it trains that go to work in industry and Wisconsin retains only 17.7% of the 492 it trains. By way of contrast, the retention rate is extremely high in California, with almost seven out of ten PhDs staying to take a job in California.

Overall, the state stay rates are low compared to those for bachelor and master degree recipients in science and engineering. Among those taking jobs in industry, for example, the stay rate is 64.4% in science and 62.3% in engineering.[12] The PhD state stay rate is also low compared to recent law school graduates for whom 57% with known employment status remain in the state of training (National Association for Law Placement, 1998).

New PhDs who leave their state of training tend to go a reasonable distance. This is clearly seen from Table 4, which shows migration flows between regions. As noted above, the Pacific region attracts a considerable number of new PhDs from the mid-west and mid-and-south-Atlantic states. In earlier work, and for the period 1997-1999, we found that, among those who left their PMSA of training, the average distance between location of training and location of employment was 835 miles.

Elsewhere (Sumell et al 2006) we have examined factors affecting the propensity for PhDs hired by industry during Period One to leave the state of training and transfer their knowledge to another state. We find that mobility relates to field and quality of the PhD program. For example, compared to the benchmark of biology, individuals trained in agriculture, engineering, chemistry, computer science and earth science are more likely to leave their state of training. Among those trained in engineering, biology, chemistry, math and medicine, those trained in top programs are more likely to leave. We also find that those who were supported on a fellowship or a dissertation grant are more likely to

[12] Interstate Migration Patterns of Recent Recipients of Bachelor's and Master's Degrees in Science and Engineering. http://www.nsf.gov/statistics/nsf05318/sect3.htm

leave their state of training.[13] Those who worked part-time during their last year in graduate school or are returning to a previous employer are also more likely to stay. Those on temporary visas are more likely to leave their state of training, as are Asians, regardless of visa status, and underrepresented minorities in science and engineering. On the other hand, individuals who went to both college and high school in their PhD state of training are considerably (17%) more likely to remain in state than those who did not receive both degrees from the same state.[14]

Our finding that only 37.1% stay in the state of training raises the question of whether the role of proximity to the university is overemphasized in the transmission of public knowledge from universities to industry. The top source of public knowledge, according to the Carnegie Mellon survey of firms (Cohen, Nelson, Wash 2002), is publications and reports. Neither requires proximity to the scientist/engineer. The second source, informal information exchange, public meetings, or conferences and consulting, is facilitated by proximity but proximity is not essential. The next tier includes recently-hired graduate students. Our research clearly shows that in this respect proximity does not play a major role.[15]

---

[13] Top fields are based on the 1993 National Research Council (NRC) rankings for all fields except medicine and agriculture. The rankings for the majority of fields are based on the "scholarly quality" scores in the NRC rankings for each relevant program at the institution. For field definitions that were broader than the program definitions in the NRC rankings (such as biology), we calculated the means for each rated program applicable to our broader field for each institution. For the fields of medicine and agriculture, we used the 1998 NSF CASPAR data to rank institutions, due to the absence of data for these fields in the NRC rankings. Institutions in these fields were ranked by total federal R&D expenditures at each institution. In the case of biology and medicine, which have a very large number of PhD programs, 75 institutions were included among the top programs. For smaller fields, such as astronomy, the top category includes the top 25 programs. In most other fields, the top category includes the top 50 programs.

[14] The logit analysis also includes controls measuring the innovative character of the state, such as patent counts, academic R&D expenditures, industrial R&D expenditure, and a measure of job opportunities for PhDs in the state. In addition, we control for per capita income, population, and the educational level of the state.

[15] This discussion also raises the question of the degree to which spillovers from public institutions result from nonappropriability. We have argued that tacit knowledge comprises an important component of the

The finding that nearly two out of three PhDs who go to work in industry leave their state of training and that more than one out of two leave their region of training highlights the degree to which the market for PhDs working in industry is national. It also underscores the degree to which the quality and scale of doctoral S&E training programs requires, at least in part, a tolerance on the part of Midwestern states of the fact that a good portion of their most prized "talent" emigrate to the Coasts. Many of these doctoral programs were initially developed and designed to meet state needs. As the labor market expands, and as the traditional industrial base of the United States shifts, one can make the case that a highly trained S&E workforce will only be maintained if the Federal government increasingly steps in to provide financial support for graduate education, since state legislatures are unlikely to continue to fund these migration flows over the long run.[16]

Section IV. Using Human Resource Data to Illuminate Patterns of Innovation

Firms hire new PhDs not only for the new knowledge that they possess but also for their ability more generally to contribute to the innovative activities of the firm. Tracking the placement of PhDs can also inform our understanding of patterns of innovation. This can be useful given that changes are occurring in patterns of innovation which traditional measures, such as patent counts and research and development expenditure data, are increasingly unable to illuminate. To quote Mowery (1999, p. 46),

---

knowledge that new PhDs transmit to firms. Yet tacit knowledge, as Zucker, Darby and Brewer (1998) point out, facilitates excludability. Thus knowledge transmission, to paraphrase the aforementioned authors, can result from the maximizing behavior of scientists who have the ability to appropriate the returns to their knowledge rather than from nonappropriability.

[16] This is not to say that the Federal government does not already provide considerable support for the training of PhD students. But much of this, with the exception of training grants from NIH, comes indirectly through the support for research assistantships on faculty member's grants.

"Without substantial change in the content and coverage of data collection, our portrait of innovative activity in the U.S. economy is likely to become less and less accurate."

Here we explore how data concerning the placement of new PhDs with firms can illuminate our understanding of patterns of innovation. Of particular interest is how such data inform our understanding of the location of innovation, the source of innovative inputs, and the degree to which human resource data relate to other measures of innovation. Before doing so, we place the discussion in context by summarizing major changes occurring in patterns of innovation.

Changing Patterns of Innovation. Four trends characterize the change that has occurred in patterns of innovation in recent years: (1) the increased reliance on external R&D, such as that performed by universities, consortia and government laboratories (Mowery, 1999, p. 44): (2) increased collaboration in the development of new products and processes with domestic and foreign competitors and customers (Mowery, 1999, p. 44); (3) a decentralization of in-house R&D activities (Merrill and Cooper, 1999); and (4) the movement of innovative activities to functions in the firm typically not thought of as being drivers of innovation. The latter is fueled in part by the development of technologies that impact the operation and marketing of the firm's production. Although all four changes contribute to the growing inadequacy of traditional measures to describe innovative activity, it is the latter two that we explore here because they can best be illuminated by examining HR data.

Increasingly firms have chosen to locate research activities at the plant level, instead of at a central R&D lab. This decentralization creates fuzziness in the current R&D data since the location of where the actual innovation is developed corresponds less and less to corporate headquarters, yet the data are collected at the corporate level. Knowing the location of PhDs working in industry can help solve the "location" problem since the placement data reflects actual location, not the location of the company's headquarters.

Another organizational change with regard to patterns of innovation is the movement of innovative activities to functions within the firm not typically regarded as drivers of innovation. One example is the assignment of scientific personnel to evaluate and seek R&D opportunities through mergers and acquisitions. Another is the involvement of technically teamed personnel in marketing and distribution. The important innovations that firms make in these areas are generally missed in standard measures of R&D. Measuring flows of new PhDs to industry regardless of their organizational assignment provides the opportunity of learning something about these sources of innovation that are not typically counted in R&D expenditure data.

Location. Table 5 shows the regional distribution of new PhDs going to work in industry. The region where the largest number of new PhDs plan to work is the Pacific (25.9%). The strong presence of IT firms in the Pacific region, as well as the heavy proportion of engineers in the database, contribute no doubt to this finding. The Mid-Atlantic region is the second largest employer of new PhDs. The East North Central is a distant third. Column 3 gives the distribution of industrial R&D expenditures by region. A comparison of the spatial distribution of new hires with the spatial distribution of R&D industrial expenditures is consistent with the argument above, showing that the distributions are spatially different. For example, we see that expenditure data undercounts innovative activity in the South Atlantic and the West South Central, and overcounts innovative activity in the Pacific region and the East North Central. While some of these differences are undoubtedly due to our inability to fully measure PhD flows to industry, the differences are suggestive that R&D expenditure data alone fail to capture regional differences.

The work location of new PhDs going to industry can also inform our understanding of the location of innovative activity at the city level—something that is not possible to obtain from industrial R&D expenditure data. Table 6 shows the top-20 PMSA destinations of new PhDs hired by firms. The data are striking on several counts. First, almost 60 percent of the placements went to one of the top 20 PMSAs. Second, there is substantial disparity in counts between the top-ranked PMSA and all others, with San Jose employing almost twice as many scientists and engineers as Boston, the second most popular destination. Third, and related, California has a high prevalence in the counts. Five of the top 20 destinations are in California. Combined, these five PMSAs capture approximately 25% of those going to a top 20 PMSA and slightly more than 16% of those going to any MSA.

The employment data are less geographically concentrated than other measures of innovation. For example, while 35% of utility patents are issued in five cities (New York, San Francisco, Los Angeles, Chicago and Boston), only 30% of industrial hires are employed in the top five cities.[17] SBIR Phase II awards are even more heavily concentrated than the patent data, with approximately one in two being awarded to firms located in San Francisco, Boston, Los Angeles, the District of Columbia and New York.

---

[17] Note that here we include Oakland and San Francisco with San Jose since the patent count data are for MSAs and not PMSAs. Distributions are taken from Black (2001).

Relationship between R&D Expenditures and Hiring Patterns. Another way to examine how the hiring data informs our understanding of innovation is to compare rankings between R&D expenditure data and rankings using the hiring data. Such a comparison shows that innovation is less concentrated than the R&D data would suggest. For example, while the top 20 R&D firms (National Science Board 2004) account for 36.2% of industrial R&D in the United States, the top 20 hiring firms account for only 22.4% of all industrial hires of new PhDs.[18] Moreover, although overlap between the top 20 R&D firms and the top 20 hiring firms exists, there are considerable differences. Only ten of the top 20 R&D firms appear on the top 20 hiring list. Clearly the PhD-hiring variable is related to the R&D expenditure variable but also captures a somewhat different dimension of innovation.

PhD placements are given in Table 7 for the top 32 hiring firms, grouped by NAIC classification.[19] In accordance with SRS guidelines, all cells contain three or more firms with no firm hiring 50 percent or more of the new PhDs. Together, these 32 firms hire approximately one-fourth of all new PhDs going to industry during the period studied.

Firms making the largest number of hires among the 32 were located in computer and electrical products, followed by firms working in publishing and professional, scientific and technical services. Five firms in pharmaceuticals and medicine are among the top 32, employing 746 new PhDs. This is particularly notable given the under-

---

[18] We measure the top 20 using R&D expenditure data for 1999, 2000 and 2001 (National Science Board, 2004, Table 4.4).

[19] We choose the number 32 in order to maximize our ability to display the data and comply with SRS's policy concerning display of data. Each cell on Table 7 contains three or more firms and no firm in any cell hires 50% or more of the new PhDs. Analysis is restricted to individuals going to work in the United States.

representation of new PhDs in biology in the data and the fact that firms in pharmaceuticals hire many freshly trained PhDs after they complete their postdoctoral training--not directly out of graduate school.

The top 32 firms recruited approximately the same number of new PhDs in Period One (2858) and Period Two (2873). However, there were large differences across NAIC classifications, reflecting changes in the underlying economy. Growth was greatest in chemicals and pharmaceuticals (37.5%). This mirrors our earlier finding that hiring increased among biologists and chemists between the two periods. Employment of new PhDs fell 42% between the two periods in aerospace products and parts. Employment remained relatively constant among the other NAIC groupings.

The SED data also provides insight into the mix of expertise that firms hire. Pharmaceuticals provide an illustrative case. During the six-year period, top-200 R&D pharmaceutical companies hired 1047 new PhDs. The dominant field of training was chemistry (402), but 100 or more were hired from four other fields: 193 from biology; 147 from engineering, 140 from medicine and 132 from math. The hires in math undoubtedly reflect the importance of modeling in drug discovery.

Foreign. Approximately 5% (1096) of the new PhDs with definite plans to go to industry indicate that they are taking a position with a firm located outside the United States. The number (and percent) going abroad is slightly lower in Period Two than in Period One. The most common foreign destination is Korea, where 22.5% of those with plans to work in industry abroad indicate that they will go; the next most likely destination is Germany (8.8%), followed by Japan at 8.5%. Canada attracts about 6% and Taiwan close to 5%. In light of recent discussions concerning increased innovative activity in developing Asian countries, it is interesting to note that approximately 6% are headed to the countries of China (1.8%), India (2.1%) or Thailand (2.0%).

Section V. Data Issues

As noted earlier, the data used for this paper undercount new PhDs going to work for industry in two respects. First, they undercount in the sense that not all PhDs have definite work plans at the time they graduate. Second, they undercount in the sense that in certain fields, especially the life sciences, it is common practice for individuals to first take a position as a post doc before eventually taking a job in industry. While we can learn something about both groups by examining patterns in the Survey of Doctorate Recipients (SDR), this is far from a perfect substitute, since the SDR only samples about 8% of PhDs for follow-up study. We would learn far more if resources were available for follow-up with those who do not have definite plans. We could also learn considerably more if a survey were done of postdocs, especially postdocs at the time they leave the postdoctoral position. Science Resources Statistics at NSF is currently in the process of reviewing and studying the possibility of fielding a postdoc survey. SRS's goal is to provide an integrated approach to surveying postdocs in order to fill in current gaps.

SRS has made some changes in data collection and its policy towards data use which have the potential to increase our knowledge about industrial placements and, by inference, the innovation process in the United States. First, SRS is in the process of adding a "salary offer" question to the SED for those with definite plans.[20] When

[20] SRS plans a limited field test of possible salary-offer-question wording and formats for the July 2006-June 2007 SED. The test will ask some respondents to identify their salary offer in ranges and others to provide a specific salary figure. Using the results of that test, SRS plans to add a "salary offer" question to the SED for the academic year beginning July 2007 through June 2008.

implemented, it will be the first time that information has been collected at the national level on starting salaries for PhDs in science and engineering. Second, SRS has established guidelines for how SRS data can be matched to other data, such as patent databases or publication counts.[21] The ability to link the PhD records with, for example, patent counts will provide another window for examining patterns of innovation. Third, SRS is exploring the possibility of coding information concerning the industrial placements of respondents to the SDR.

Section VI. Conclusion

Here we have examined hiring patterns of recently trained PhDs in science and engineering who have definite plans to work in industry after graduation. The period of analysis is 1997-2002. Data are taken from the Survey of Earned Doctorates, a census of recent PhDs which has a response rate of approximately 92%. While respondents have long been asked to identify the name and location of where they will work, prior to 1997 the data was not coded for those with plans to go to industry and since 1997 it has only been collected in verbatim form. We have now coded the verbatim records by firm name and location for the six-year period 1997-2002 and identified placements made at top-200 R&D firms. During the period analyzed, almost 22,000 new PhDs indicated that they had definite plans to work for a firm after graduation and identified the firm and the location of the firm. This represents approximately 15% of all newly minted PhDs during this time period and approximately 23% of all PhDs who had definite plans at the time of graduation.

[21] The policy is described at the following web site:
http://www.nsf.gov/statistics/database.cfm

Data on firm placements provide insights that other data do not provide. One such insight relates to where these newly minted and hired PhDs trained. This is of interest since newly trained PhDs provide one means by which knowledge, especially tacit knowledge, is transferred from the public sector to the private sector. We find these knowledge sources to be heavily concentrated in certain regions and states. Moreover, the geographic distribution of knowledge sources, as measured by where PhDs going to work in industry trained, is different than other measures of knowledge sources would suggest, such as university R&D-expenditure data. We conclude that the spatial distribution of knowledge sources embodied in newly minted talent is different from the distribution of knowledge sources stemming from university research, as measured by university R&D expenditures.

A major headline here is the strong role played by Midwestern universities, which educate over 26% of all PhDs going to industry. Indeed, seven of the top twenty institutions educating PhDs to work in industry are located in the Midwest. We also find that PhDS working in industry are not particularly likely to remain in the state where they received their PhD training. Compared to master-degree recipients going to work in industry, PhDs are almost 45% less likely to remain in the state of training. To wit, the state stay rate for PhDs working in industry is 37%; that for masters is 65%. The finding suggests that it is important to rethink the role that proximity to the university plays in the transmission of knowledge.

Stay patterns are particularly low among certain Midwestern states and universities located in these states. Some of these states have seen a considerable decline in their industrial prowess in recent years. As Nathan Rosenberg has pointed out, it is not accidental that athletes at Purdue University bear the nickname of “boilermakers,” reflecting Purdue’s early commitment to engineering education supporting industry in the state of Indiana. While the name persists, Purdue’s PhDs now overwhelmingly leave the state to take employment elsewhere—many as far away as the west coast. One can make the case that as the traditional industrial base of the United States shifts, a highly trained S&E workforce will only be maintained if the Federal government increasingly steps in to provide financial support for graduate education, since state legislatures are unlikely to continue to fund these migration flows over the long run. It is risky as a nation to continue to rely on the “kindness” of Midwestern states to educate the high-quality S&E workforce that heads out-of-state upon graduation.

Hiring data also inform our understanding of patterns of innovation. This is particularly useful given that R&D data are often collected at the corporate level and thus do not reflect the decentralization that is occurring in research and development, as companies move away from large central labs. Hiring patterns also provide information on scientists and engineers working in industry, regardless of their organizational assignment. This provides the opportunity for learning something about resources employed in innovative activity that are not typically counted in R&D expenditure data.

Firms most likely to hire new PhDs are found in computer and electrical products, followed by firms working in publishing and professional, scientific and technical services. Five firms in pharmaceuticals and medicine are among the top hiring firms. Apropos to the above argument, while we find some overlap between top hiring firms and top R&D firms, there are also considerable differences. Only ten of the top 20 R&D

firms appear on the top-20 hiring list. Clearly the PhD hiring variable is related to R&D expenditures but also captures a somewhat different dimension of innovation.

New PhDs working for industry are most likely to head to San Jose. Indeed, almost one out of five new PhDs going to work for industry heads to San Jose. It is no wonder that the San Jose newspaper has a fulltime science reporter! Other top-destination cities include Boston, New York, Washington, D.C., Portland-Seattle, and Chicago. While industrial employment of newly trained scientists and engineers is heavily concentrated in a handful of cities, it is not nearly as concentrated as are counts of patents or SBIR Phase II awards.

The location data highlights the role that PhDs play in local economic development, not only through their contribution to innovation, but also through the economic impact that their relatively high wages exert on the local economy. Sumell (2005), for example, estimates that a newly trained PhD in computer science working in industry earns $86,700 a year; a newly trained PhD in electrical engineering earns $78,500. More than 300 new PhDs a year go to work in industry in San Jose alone. Many of these are electrical engineers and computer scientists. Hired to work on products that will have a global market, they spend much of their income locally. Through the multiplier effect, their spending contributes to regional economic growth.

Finally, our data suggest that small firms play a larger role in innovation than R&D data would suggest. For example, while the top 200 R&D firms expend more than 70% of all R&D in the U.S., they hire only 39% of all new PhDs. The difference reflects in part the degree to which small firms are "knowledge-intensive" and the degree to which R&D statistics are dominated by development costs associated with large firms, as opposed to research costs. It is difficult to know the extent to which this small-firm effect reflects Federal policies such as the SBIR program that are aimed specifically at small innovative firms. But the knowledge that small firms contribute substantially to innovation[22] and are hiring newly-minted PhDs suggests that the Federal government might consider further leveraging the benefits coming from small knowledge-intensive firms by investing additional resources in programs aimed at small innovative firms. Such a policy not only has the potential of contributing to innovation and subsequent economic growth. It could also augment the number of research positions available for scientists and engineers and send a positive signal to those contemplating careers in science and engineering.

[22] See, for example, the work of Acs and Audretsch (1990) which discusses the increased importance small firms play in generating innovation, especially in certain industries.

References:

Acs, Zoltan and David Audretsch. Innovation and Small Firms, MIT Press, Cambridge, 1990.

Anselin, Luc, Attila Varga and Zoltan J. Acs. "Local Geographic Spillovers between University Research and High Technology Innovations." *Journal of Urban Economics*, v. 42, pp. 422-448, Nov. 1997.

__________________, "Geographic Spillovers and University Research, A Spatial Econometric Perspective." *Growth and Change*, v. 31, pp. 501-515, Fall 2000.

Audretsch, David and Marianne Feldman. "Innovation Clusters and the Industry Life Cycle," *Review of Industrial Organization*, v. 11, pp. 253-273, 1996a.

_______________, "R&D Spillovers and the Geography of Innovation and Production." *American Economic Review*, v. 63, pp. 630-640, 1996b.

Audretsch, David and Paula Stephan, "Company-Scientist Locational Links: The Case of Biotechnology." *American Economic Review*, v. 86, pp. 641-652, 1996.

Black, Grant. The Geography of Small Firm Innovation. Doctoral dissertation, Georgia State University, 2001.

Cohen, Wesley. R. Nelson and J. Walsh. "Links and Impacts: The Influence of Public Research on Industrial R&D. *Management Science*, v. 48, pp. 1-23, 2002.

Dasgupta, Partha and Paul David."Towards a New Economics of Science." *Research Policy* v. 23, no. 5, pp. 487-521, 1994.

Feldman, Maryann and David B. Audretsch. "Innovation in Cities: Science Based Diversity, Specialization and Localized Competition." *European Economic Review*, v. 43, pp. 409-429, Feb. 1999.

Feldman, Maryann. *The Geography of Innovation*. Dordrecht, The Netherlands: Kluwer Academic Publishers. 1994.

Jaffe, Adam. "Real Effects of Academic Research," *American Economic Review*, v. 70, pp. 957-970, 1989.

Merrill, Stephen A. and Ronald S. Cooper. "Trends in Industrial Research and Development: Evidence from National Data Sources" in National Research Council, Securing America's Industrial Strength. National Academy Press, Washington, DC. 1999.

Mowery, David. "America's Industrial Resurgence: How Strong, How Durable?" *Issues in Science and Technology*, 15(3): 41-48, 1999.

National Association for Law Placement. "Class of 1997 Employment Report and Salary Survey." National Association for Law Placement, Washington D.C., 1998.

National Science Board, *Science and Engineering Indicators 2002*. Two volumes, Arlington, VA: National Science Foundation, 2002.

___________ *Science and Engineering Indicators 2004*. Two volumes. Arlington, VA: National Science Foundation, 2004.

National Science Foundation, "Interstate Migration Patterns of Recent Recipients of Bachelor's and Master's Degrees in Science and Engineering." http://www.nsf.gov/statistics/nsf05318/sect3.htm

Stephan, Paula, Albert Sumell, Grant Black, and James Adams. "Doctoral Education and Economic Development: The Flow of New PhDs to Industry." *Economic Development Quarterly*, v. 18, pp. 151-167, 2004.

Stephan, Paula, "Using Human Resource Data to Illuminate Innovation and Research Utilization," Using Human Resource Data to Track Innovation, Stephen Merril and Michael McGeary, eds., National Academy Press, Washington D.C., 2002.

Sumell, Albert J., "The Location Decisions of the Highly Educated: The Role of Amenities and the Implications for State Investment Policies." Unpublished dissertation, Georgia State University, 2005.

Sumell, Albert J., Paula Stephan, and James Adams, "Capturing Knowledge: The Location Decision of New PhDs Working in Industry." Working Paper, Department of Economics, Georgia State University, 2006.

Zucker, Lynn, Michael Darby and M. Brewer. "Intellectual Capital and the Birth of the U.S. Biotechnology Enterprise." *American Economic Review*, v. 88, pp. 290-396, 1998.

Table 1
Firm Placements of New Science and Engineering PhDs: 1997-2002

	Period One				Period Two			
Field of PhD	Percent PhDs awarded who identified a firm	Number PhDs awarded who identifie d a firm	Percent Going to a top 200 R&D Firm	Number Going to a top 200 R&D firm	Percent PhDs awarded who identified a firm	Number awarded who identified a firm	Percent Going to a Top 200 R&D firm	Number Going to a top 200 R&D Firm
All science and engineering	14.5	10,932	37.8	4134	15.2	10,83 3	40.0	4333
All engineering	30.7	5,364	44.7	2400	31.9	5,089	47.8	2435
Agriculture	9.0	308	14.9	46	8.2	256	31.2	80
Astronomy	7.8	44	36.4	16	6.8	35	48.6	17
Biology	3.8	609	23.2	141	5.2	843	24.6	207
Chemistry	18.7	1216	45.0	547	22.2	1310	45.0	589
Computer science	28.4	762	50.3	383	27.9	697	45.3	316
Earth science	12.3	252	29.7	100	13.1	192	31.2	60
Math	12.5	477	32.3	154	12.3	417	35.5	148
Medicine	5.0	435	20.0	87	5.4	486	26.1	127
Other (Economics and psychology)	8.3	811	10.7	87	9.0	1037	8.8	91
Physics	16.1	654	33.2	217	18.1	638	41.2	263

Table 2
Region of Training 1997-2002 of Those Working in the U.S.

Region Trained	Percent Trained	University R&D (percent)*
New England	8.3	8.2
Mid Atlantic	16.9	14.7
East North Central	19.7	14.4
West North Central	6.9	6.7
South Atlantic	15.4	19.0
East South Central	2.6	4.4
West South Central	8.2	9.3
Mountain	5.0	6.3
Pacific	16.9	17.0

*Expenditure data are for 1997-1999 and come from National Science Board (2002)

Table 3
Top-20 Producing Universities of PhDs Headed to Industry*
1997-2000

University	Number Trained
Stanford	732
Illinois—Urbana/Champaign	670
California, Berkeley	579
Texas, Austin	576
Purdue, main campus	528
MIT	527
Minnesota, Twin Cities	521
Michigan, Ann Arbor	489
Georgia Institute of Technology	451
Wisconsin, Madison	430
Pennsylvania State	388
North Carolina State	381
UCLA	365
Cornell	335
Ohio State, main campus	302
Northwestern	299
Carnegie Mellon	288
Texas A&M	278
Maryland, College Park	277
Southern California	264

Table 4
Regional Flows of New PhDs Going to Industry: 1997-2002*
(Represents percent staying in region of training)

Region of Employment/Region of Training	New England	Mid Atlantic	East North Central	West North Central	South Atlantic	East South Central	West South Central	Mountain	Pacific	Total Employed
New England	842 (45.8)	306	228	61	215	34	60	49	127	1922
Mid Atlantic	341	1871 (51.1)	602	182	467	42	142	87	315	4049
East North Central	66	213	1622 (38.2)	191	257	76	88	55	103	2671
West North Central	31	57	168	504 (33.8)	65	23	37	23	41	949
South Atlantic	111	277	268	101	1402 (42.3)	69	101	53	110	2492
East South Central	s	19	36	16	49	180 (31.7)	27	9	s	345
West South Central	59	135	232	99	208	47	939 (52.1)	88	103	1910
Mountain	26	63	110	41	84	18	53	457 (42.1)	62	914
Pacific	229	538	781	220	429	53	247	212	2610 (71.5)	5319
Out of U.S.	132	181	201	75	141	26	108	52	180	1096
Total Trained	1837**	3660	4248	1490	3317	568	1802	1085	3651**	21667

s=suppressed. Not reported if counts are 6 or less or a specific firm contributes 50% or more to a cell.
*Counts exclude those trained in Puerto Rico or going to Puerto Rico as well as those with an unknown employment location (total of 72 cases). **Counts do not include suppressed numbers.

Table 5
Region of Employment 1997-2002

Region Employed	Percent Employed	Industrial R&D (percent)*
New England	9.3	9.5
Mid Atlantic	19.7	18.0
East North Central	13.0	17.2
West North Central	4.6	4.1
South Atlantic	12.1	9.5
East South Central	1.8	1.8
West South Central	9.3	5.8
Mountain	4.4	6.0
Pacific	25.9	28.1

*Expenditure data are for 1997-1999 and come from National Science Board (2002)

Table 6*

Top 20 Metropolitan Statistical Area Locations of Industrial Hires: 1997-2002

PMSA	Number	Percent
San Jose	1878	9.1
Boston	1015	4.9
New York	937	4.5
Washington DC MD VA	758	3.7
Portland-Seattle	694	3.4
Chicago	669	3.2
Los Angeles-Long Beach	622	3.0
Houston	586	2.8
Newark	547	2.6
San Francisco	534	2.6
Dallas	505	2.4
Minneapolis	439	2.1
Detroit	429	2.1
Oakland, CA	424	2.1
Philadelphia PA-NJ	377	1.8
San Diego	345	1.7
Austin	341	1.7
Raleigh-Durham	320	1.5
Atlanta	309	1.5
Middlesex-Somerset-Hunterdo	299	1.4
Total Top 20	12028	58.2
Other PMSAs	7272	35.2
U.S. NON PMSA	1360	6.7
Total in U.S.	20660	1.00

* Each cell represents hiring by three or more firms and no firm in any cell hires 50% or more of the new PhDs reported in that cell.

Table 7
Top 32 Firms Hiring New PhDs by NAIC Classification: 1997-2002*

NAIC Classification	Industry	Number
3254	Pharmaceuticals	746
325	Chemical other than pharmaceuticals	418
331, 333	Primary metal; machinery	304
334	Computer and electrical products	1541
3364	Aerospace Products and parts	316
336	Transportation other than aerospace	349
511, 541	Publishing industries and Professional, Scientific and Technical Services,	1244
32,513, 99	Other manufacturing; Broadcasting and telecommunications; conglomerate	813
Total		5731

*Each cell reports data on three or more firms and no firm in any cell hires 50% or more of the new PhDs reported in that cell.

Appendix
Inter-State and Inter-Regional Migration Patterns of New Industrial PhDs
1997-2002

State/***Region***	Number of New PhDs Trained In State/***Region***	Number of New PhDs Working In State/***Region***	Number of New PhDs Produced that Stay In State/***Region***	Percent of New PhDs Produced that Stay In State/***Region***
New England	***1846***	***1922***	***842***	***45.7***
Connecticut	268	429	79	29.5
Maine	18	19	*s*	*s*
Massachusetts	1358	1283	550	40.5
New Hampshire	61	79	17	27.9
Rhode Island	121	46	16	13.2
Vermont	20	*66*	8	40.0
Mid Atlantic	***3668***	***4050***	***1871***	***50.9***
New Jersey	618	1455	299	48.4
New York	1735	1730	635	36.6
Pennsylvania	1315	865	327	24.9
East North Central	***4270***	***2672***	***1622***	***38.0***
Illinois	1306	881	367	28.1
Indiana	711	311	84	11.8
Michigan	871	696	316	35.6
Ohio	890	558	268	25.4
Wisconsin	492	226	87	17.7
West North Central	***1497***	***953***	***504***	***33.7***
Iowa	317	90	43	13.6
Kansas	202	94	50	24.8
Minnesota	552	484	190	34.4
Missouri	304	218	85	28.0
Nebraska	70	43	20	28.6
North Dakota	37	9	*s*	*s*
South Dakota	15	11	*s*	*s*
South Atlantic	***3328***	***2492***	***1402***	***42.1***
Delaware	131	*147*	*s*	*s*
Florida	506	301	156	30.8
Georgia	618	348	185	29.9
Maryland	486	437	128	26.3
North Carolina	701	433	211	30.1
South Carolina1	170	122	36	21.2
Virginia	529	464	153	28.9
West Virginia	48	56	8	16.7

Washington D.C.	139	184	*20*	14.4
East South Central	***570***	***345***	***180***	***31.7***
Alabama	194	102	48	24.7
Kentucky	91	58	13	14.3
Mississippi	90	24	12	13.3
Tennessee	195	161	72	36.9
West South Central	***1806***	***1910***	***939***	***49.2***
Arkansas	41	27	12	29.3
Louisiana	172	135	40	23.3
Oklahoma	161	79	39	24.2
Texas	1432	1669	738	51.5
Mountain	***1081****	***914***	***457***	***42.3***
Arizona	373	339	146	39.1
Colorado	375	313	153	40.1
Idaho	25	50	7	28.0
Montana	26	12	*s*	*s*
New Mexico	79	80	26	32.9
Utah	185	91	60	32.4
Nevada	*s*	22	*s*	*s*
Wyoming	25	7	*s*	*S*
Pacific	***3657****	***5319***	***2610***	***71.4***
Alaska	*s*	9	*s*	*s*
California	3176	4465	2200	69.3
Oregon	154	*s*	s	*s*
Washington	304	353	107	35.2
Hawaii	23	*s*	*s*	*s*
Puerto Rico	***28***	***30***	***21***	***75.0***

**Does not include suppressed counts.*
s=suppressed. At the request of Science Resources Statistics, National Science foundation, counts not reported if 6 or less or if a specific firm contributes half or more of the count in a cell.
Note that counts differ from those of Table 4 which excludes those trained or going to Puerto Rico as well as those with an unknown location.