NBER WORKING PAPERS SERIES

GEOGRAPHIC LOCALIZATION OF KNOWLEDGE SPILLOVERS
AS EVIDENCED BY PATENT CITATIONS

Adam B. Jaffe

Manuel Trajtenberg

Rebecca Henderson

Working Paper No. 3993

GEOGRAPHIC LOCALIZATION OF KNOWLEDGE SPILLOVERS
AS EVIDENCED BY PATENT CITATIONS

ABSTRACT

We compare the geographic location of patent citations to those of the cited patents, as evidence of the extent to which knowledge spillovers are geographically localized. We find that citations to U.S. patents are more likely to come from the U.S., and more likely to come from the same state and SMSA as the cited patents than one would expect based only on the preexisting concentration of related research activity. These effects are particularly significant at the local (SMSA) level, and are particularly apparent in early citations.

Adam B. Jaffe
Department of Economics
Harvard University
Cambridge, MA 02138
and NBER

Manuel Trajtenberg
Tel Aviv University
Ramat Aviv
Tel Aviv 69978
ISRAEL
and NBER

Rebecca Henderson
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02139
and NBER

GEOGRAPHIC LOCALIZATION OF KNOWLEDGE SPILLOVERS

AS EVIDENCED BY PATENT CITATIONS

by Adam B. Jaffe, Manuel Trajtenberg and Rebecca Henderson

The last decade has seen the development of a significant body of empirical research on R&D spillovers.[1] Generally speaking, this research has shown that the productivity of firms or industries is dependent not only on their R&D spending, but also on the R&D spending of other firms or other industries. In parallel, economic growth theorists have focussed new attention on the role of knowledge capital in aggregate economic growth, with a prominent modelling role for knowledge spillovers (e.g., Romer, 1986 and 1990; Grossman and Helpman, 1991).

We know very little, however, about where spillovers go. Is there any advantage to nearby firms, or even firms in the same country, or do spillovers waft into the ether, available for anyone around the globe to grab? The presumption that U.S. international competitiveness is affected by what goes on at Federal Labs and U.S. universities, and the belief that universities and other research centers can stimulate regional economic growth[2] are predicated on the existence of a geographic component to the spillover mechanism. The existing spillover literature is, however, virtually silent on this point.[3]

---

1. E.g., Jaffe (1986), and Nadiri and Bernstein (1988 and 1989). For a recent survey and evaluation of this literature, see Griliches (1991).

2. See, e.g., Minnesota Department of Trade and Economic Development (1988); Feller (1989); and Smilor, et al (1989).

3. Jaffe (1989) provides evidence that corporate patenting at the state level depends on university research spending, after controlling for corporate R&D. Mansfield (1991) surveyed industrial R&D about university research from which they benefitted. He

In the growth literature, it is typically assumed that knowledge spills over to other agents within the country, but not to other countries.[4] This implicit assumption clearly begs the fundamental question of whether and to what extent knowledge externalities are localized. As emphasized recently by Krugman (1991), acknowledging the importance of spillovers and increasing returns requires renewed attention by economists to issues of economic geography. Krugman revives and explores the explanations given by Marshall (1920) as to why industries are concentrated in cities. Marshall identified three factors favoring geographic concentration of industries: (1) the pooling of demands for specialized labor; (2) the development of specialized intermediate goods industries; and (3) knowledge spillovers among the firms in an industry. Krugman believes that economists should focus on the first two of these, partially because he perceives that "[k]nowledge flows, by contrast, are invisible; they leave no paper trail by which they may be measured and tracked, and there is nothing to prevent the theorist from assuming anything about them that she likes." (Krugman, p. 53)

Glaeser, et al (1991) characterize the "Marshall-Arrow-Romer" models as focussing on knowledge spillovers within the firms in a given industry. They examine the growth rate of industries in cities as a function of the concentration of industrial activity across cities, within-city industrial diversity, and within-city competition. They find that

---

found that they most often identified major research universities, but that there was some tendency to cite local universities even if they were not the best in their field.

4. The existence of this implicit assumption was noted by Glaeser, et al (1991): "After all, intellectual breakthroughs must cross hallways and streets more easily than oceans and continents." Grossman and Helpman (1991) consider international knowledge spillovers explicitly.

within-city diversity is positively associated with growth of industries in that city, while concentration of an industry within a city does not foster its growth. They interpret this contrast to mean that spillovers across industries are more important than spillovers within industries. As is discussed below, there is evidence from the R&D spillover literature to suggest that across-industry knowledge spillovers are, indeed, important. In this study, we do not consider the industrial identity of either generators or receivers of spillovers, though we do have some information on their technological similarity.

Our approach is to seek evidence of spillover-localization in patent citation patterns. Taking a citation from a later patent as evidence of a subsequent technological development that builds upon the result of the cited patent, it provides some evidence of the "paper trail" left by the "invisible" knowledge flow. Because patents contain detailed geographic information about their inventors, we can examine where these trails actually lead. We perform this examination for the citations of patents assigned to universities, and also for the citations of a sample of domestic corporate patents. If knowledge spillovers are localized within countries, then citations of patents generated within the U.S. should come disproportionately from within the U.S. To the extent that regional localization of spillovers is important, citations should come disproportionately from the same state or metropolitan area as the originating patent.

The most difficult problem confronted by the effort to test for spillover-localization is the difficulty of separating spillovers from correlations that may be due to a pre-existing pattern of geographic concentration of technologically related activities. That is, if a large fraction of citations to Stanford patents come from Silicon valley, we

would like to attribute this to localization of spillovers. A slightly different interpretation is that a lot of Stanford patents relate to semiconductors, and a disproportionate fraction of the people interested in semiconductors happen to be in Silicon valley, suggesting that we would observe localization of citations even if proximity offers no advantage in receiving spillovers. Of course, the ability to receive spillovers is probably one reason for this pre-existing concentration of activity. If it were the only possible reason, then, under the null hypothesis of no spillover localization we should still see no localization of citations. As discussed above, however, there are other sources of agglomeration effects that could explain the geographic concentration of technologically related activities without resort to localization of knowledge spillovers. We will show that the frequency with which citations are localized is significantly greater than a control frequency designed to capture the pre-existing geographic distribution of technologically related activities. Since this "control" frequency is, itself, likely to be partly the result of spillover-localization, we believe this to be a conservative test for the existence of localization.

The first section of the paper describes patents, and considers more carefully how citations might be used to infer spillovers. The following section explains the construction of the samples of patents used in this study. The third section presents an analysis of the frequency with which citations come from the same country, the same state and the same metropolitan area as the originating patent, and compares these to "control" frequencies. The fourth section examines whether the probability of geographic localization of any given citation can be explained by attributes of the originating or

citing patents, or of relationships between them. A concluding section follows.

I. Patents and Patent Citations

A patent is a property right in the commercial use of a device.[5] For a patent to be granted, the invention must be non-trivial, meaning that it would not appear obvious to a skilled practitioner of the relevant technology, and it must be useful, meaning that it has potential commercial value. If a patent is granted,[6] a public document is created containing extensive information about the inventor, her employer, and the technological antecedents of the invention, all of which can be accessed in computerized form. Among this information are "references" or "citations." What citations a patent must include is determined by the patent examiner. The citations serve the legal function of delimiting the scope of the property right that the patent constitutes. In theory, the granting of the patent is a legal statement that the idea embodied in the patent represents a novel and useful contribution over and above the previous state of knowledge, as represented by the citations. Thus, in principle, a citation of Patent X by Patent Y means that X represents a piece of previously existing knowledge upon which Y builds.

The examiner has several ways of identifying potential citations. The applicant has a legal duty to disclose any knowledge of the prior art that she may have. In addition, the examiner is supposed to be an expert in the technological area and be able

---

5. Ideas are not patentable; nor are algorithms or computer programs, though a chip with a particular program coded into it might be. The definition of a device was recently broadened to include genetically engineered organisms.

6. There is no public record of unsuccessful patent applications.

to identify relevant prior art that the applicant misses or conceals. The framework for the search of the prior art is the patent classification system. Every patent is assigned to a 9-digit patent class (of which there are about 100,000) as well as an unlimited number of additional or "cross-referenced" classes. An examiner will typically begin the search of prior art using her knowledge of the relevant classes. For the purpose of identifying distinct technical areas, we utilize aggregations of subclasses to a 3-digit level; at this level there are currently about 400 technical classes.[7]

The main advantage of patent data can be stated simply: They are easily available and they provide a tremendous amount of information about the invention, the inventor and her employer.[8] Every major research organization holds some patents, and the associated data are publicly available in computerized form. There is no other form of data that gives such broad coverage of the output of the research enterprise. Further, the data available for each patent are quite extensive: In addition to the citation and classification information discussed above, one knows the application date, the name and exact address of each inventor, and the name of the organization to which the patent right is assigned, if any. The combination of the citation information with detailed institutional and geographic information about each applicant provides a unique mechanism for tracing the diffusion of technology across time, space, and types of

---

7. Examples of 3-digit patent classes are "Batteries, Thermoelectric and Photoelectric;" "Distillation: Apparatus;" "Robots;" 17 distinct classes of "Organic Compounds;" and the ever-popular "Whips and Whip Apparatus."

8. For a general discussion of the value and problems of patent data, see Griliches (1990) and Trajtenberg (1990, Chapter 5).

institutions.

There are, however, limits to the value of patent data for our purposes. Most fundamentally, much of the output of research cannot be patented, and this is particularly true for basic research, which may generate the greatest spillovers. Beyond the question of what is patentable, there is a question of what is patented. An inventor in possession of what she judges to be a patentable idea decides whether or not to apply for a patent. Though the decision to apply, and favorable action by the Patent Office, may create a presumption of potential value to the invention, a decision not to apply does not mean that the invention is valueless. Patenting is a strategic decision. In addition, firms and universities face quite different incentives in this regard. Until 1980, universities could not grant exclusive licenses to commercialize patents derived from federally funded research. This restriction greatly limited the effective monopoly power that a patent is intended to confer, and hence greatly reduced incentives to apply for patents derived from federally funded research, which is about 90% of university research. Firms, on the other hand, may elect not to patent and rely on secrecy to protect their property rights.

None of these limitations seem particularly troubling for the narrow purpose at hand. We do not purport in this paper to measure the knowledge output of firms or universities, or the fraction that "spills out." We simply take a set of patents (described further below) as evidence of a set of potentially economically useful inventions, and then examine where subsequent related inventions were developed. While this set is surely a non-random sample of the universe of new knowledge creations, it still seems

informative to examine the geographic patterns that emerge.

Separate from the question of biases created by looking at spillovers from patented inventions, there is the question of whether it is appropriate to use patent citations as indicating knowledge spillovers in the way that we propose. The role of the examiner in identifying citations means that the citing inventor may not actually have been aware of the work of the cited inventor. Further, even if the citing inventor was aware of the cited work, she may not, in fact, have benefitted from her knowledge of it.[9] In using citations to trace the pattern of knowledge spillovers, we risk imputing spillovers that did not really exist. For our purposes, however, this implies a conservative bias against finding substantive results that only underscores the importance of those results if found. That is, if many citations do not actually correspond to true spillovers, then citations would be an extremely noisy indicator of spillovers, suggesting that one might not find a geographic pattern to citations even if there really is a geographic pattern to spillovers. But if we _do_ find that there is a geographic pattern to citations, the fact that citations mis-measure spillovers only means that our results understate the importance of geography.[10]

---

9. As in any paper, the Bibliography at the end of this paper contains references that we feel have to be included for completeness, but from which we may have received little direct intellectual benefit.

10. This reasoning would break down if non-spillover citations were _more_ geographically localized than spillovers. It is difficult for us to see how this might be the case. Certainly, for citations to previous patents that represent work of which the citing inventor is unaware, one would expect no geographic connection (other than the localization due to concentration of the underlying technologically activities, which we control for directly).

II. The Data

We begin with two sets of university patents: 316 comprising the universe of successful applications from the year 1975, and 482 comprising the universe of successful applications from the year 1980.[11] We are particularly interested in universities because of a prior belief in their importance in generating spillovers. In order to compare the citation patterns of university patents with those of corporate patents, we also drew two "matching" samples of corporate patents to correspond to each of these university sets. One sample (the "Top Corporate" sample) was drawn from patents granted to the 200 U.S. firms with the greatest R&D spending in 1986, according to Compustat. The "Other Corporate" sample was drawn from the universe of all other patents assigned to U.S. corporations. In order to make the matching samples as similar as possible except for their institutional origin, the corporate samples were drawn as follows:

1. For each university patent, we identified all patents in both the Top Corporate and Other Corporate groups that had the same patent class and application year as the university patent.

2. From each of these two sets of patents matched by class and application year, we then drew the patent that minimized the absolute value of the difference in patent numbers between the university patent and the matching sample patent.

3. Step (2) was performed without replacement, that is, if a patent class had n university patents, we drew n distinct matching sample patents.

---

11. These patents may have been granted anytime between their application date and the end of 1989. In practice, most patents are granted (or denied) within about 3 years of application. We have no information on unsuccessful applications.

The result of this exercise is that for both the 1975 and 1980 university cohorts, we have

samples of Top Corporate and Other Corporate patents with the same application year

and the same patent class distribution. By matching on patent classes, we control for

variations in citation practices across technological areas. Because patent numbers are

assigned sequentially, choosing matching sample patents with close patent numbers

results in matches that were granted very close in time to the originating patent. This

is desirable so that the matching samples will have had the same amount of time to be

cited. About 90% of the matching sample patents were granted within 1 month of the

matching university patent, and over 99% were granted within 1 year.

These 6 distinct sets of patents (1975 and 1980 cohorts for each of university, Top

Corporate, Other Corporate) represent the potential generators of spillovers; we call

them "originating patents." The next step was to identify all of the patents citing any of

these originating patents, of which there were about 10,000 by the end of 1989. As a

prelude to the geographic analysis, Tables One and Two and Figure One present some

descriptive data about the citations and their relationship to the originating patents.

Table One shows that about 80-90 percent of the 1975 patents and 70-80 percent of the

1980 patents had received at least one citation by the end of 1989, with the higher

proportion in each case applying to the university patents. Mean citations received

(including zeros) were 4-6 for 1975 and 3-4 for 1980, again with the higher numbers

corresponding to the university patents.[12] The average lag between the originating

---

12. Our companion paper (Trajtenberg, Henderson and Jaffe, 1992) explores in detail
the inferences that can be drawn about the nature of university and corporate research
from differences in citation intensity and related measures.

TABLE ONE
DESCRIPTIVE STATISTICS

Originating Dataset	Percent Receiving Citations	Total No. of Citations	Mean Citations Received	Average Citation Lag[1,2]	Percent Self Citations[2]	Percent Same Patent Class[2]
1975						
University	88.6	1933	6.12	6.53	5.6	54.3
Top Corporate	84.2	1476	4.70	7.17	18.6	55.7
Other Corporate	82.3	1341	4.22	7.82	9.1	57.5
1980						
University	79.9	2093	4.34	4.36	8.9	56.3
Top Corporate	79.9	1701	3.54	4.41	24.6	58.3
Other Corporate	74.1	1424	2.95	4.46	12.6	57.2

Notes:

1. Application year of citing patent minus application year of originating patent

2. For those patents receiving any citations

Table 2

Originating and Citing Patents by Technological Field

Origin Field	Origin Distribution (Percent of Column)	Field Distribution of Citations (Percent of Row)					Distribution of Citations by Origin Field (Percent of Column)
		1 Drugs and Medical	2 Chemicals exc. Drugs	3 Electronics Optics and Nuclear	4 Mechanical Arts	5 All Other	
1975							
1	28.2	83.7	9.3	1.8	4.0	1.2	32.4
2	22.2	9.5	73.3	9.5	7.2	0.5	16.9
3	26.3	2.0	5.1	88.3	4.2	0.4	29.3
4	16.1	2.6	15.0	8.2	71.4	2.9	16.1
5	7.3	4.1	7.1	4.0	9.0	75.9	5.2
Total	100.0	30.0	19.7	29.6	15.7	5.0	100.0
1980							
1	36.9	78.7	11.8	4.8	3.5	1.2	34.6
2	22.8	9.9	70.8	3.2	14.6	1.5	22.5
3	22.0	2.1	4.8	84.2	8.5	0.4	27.3
4	13.9	3.6	10.6	7.8	75.3	2.7	11.7
5	4.4	6.5	3.6	2.2	9.6	78.1	3.9
Total	100.0	30.8	22.7	26.3	16.0	4.3	100.0

application year and the application year of the citing patent is 6.5 to 8 years for the 1975 cohort, and a little over 4 years for the 1980 cohort.

The inference that a citation indicates a possible knowledge spillover is much less clear in the case where the citing patent is owned by the same organization as the originating patent. For this reason, we distinguish what we call "self-citations." A self-citation is defined as a citing patent assigned by its inventors to the same pa.ty as the originating patent, which is, by construction, either a university or a domestic corporation. Not surprisingly, the self-citation rate differs for the different sources of originating patents, with universities having the lowest and Top Corporations the highest rates.[13] Finally, Table One shows that 55 to 60 percent of citations have a primary patent class that is the same as the primary patent class of the originating patent, indicating that the originating and citing patents are technologically close to one another.

The technological relationships between the citing and originating patents are summarized in a different way in Table Two. This Table uses a very broad 5-way technological classification, based on the underlying patent classes: (1) Drugs and Medical Technology; (2) Chemicals and Chemical Processes Excluding Drugs; (3) Electronics, Optics. and Nuclear Technologies; (4) Mechanical Arts; and (5) All Other. Even at this broad classification level, one cannot assume that a citing patent is in the same category as the originating patent. The Table shows a cross-classification of frequencies across these fields for the originating and citing patents, with the university
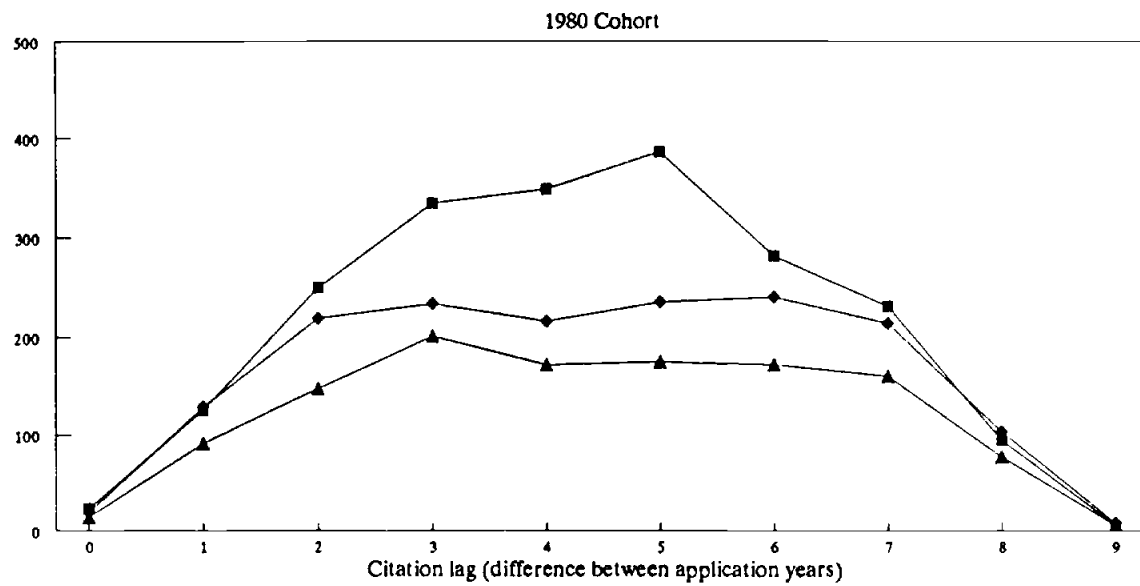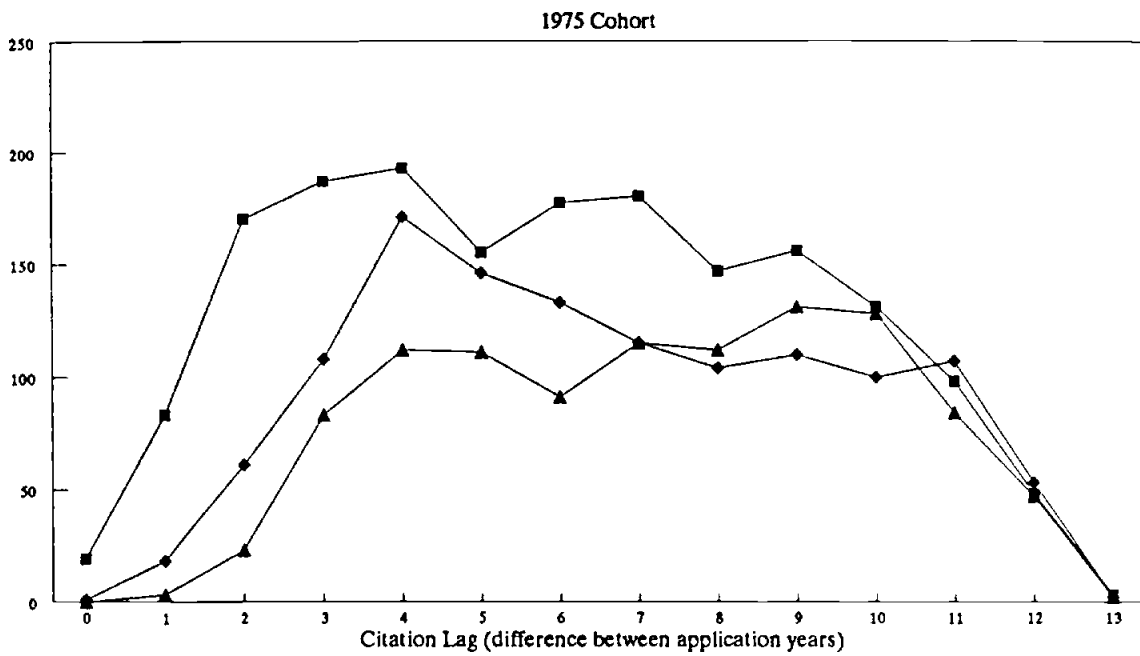
---

13. The apparent increase in self-citation rates between 1975 and 1980 is probably spurious; self-citations tend to come earlier than other citations. See Trajtenberg, Henderson, and Jaffe (1992) for more on this issue.

and corporate patents combined for this purpose. For example, the Table shows that 83.7 percent of those citations received by all of our 1975 Drug and Medical Patents were themselves classified as Drug and Medical; about 9.3 percent were classified in Chemicals. Overall, the 28.2 percent of our 1975 patents that were classified as Drugs and Medical generated 32.4 percent of citations to our 1975 patents; of all the citations (regardless of origin field) to our 1975 patents, 30 percent were Drug and Medical. The diagonal elements of the matrices in each panel, which correspond to the fraction of citations that are within broad technical fields, range from 70.8 to 88.3.

Figure One provides additional detail on the distribution of lags between originating and citing patents, again defined as the difference in application years. The Figure shows that citations are few in the early years,[14] and reach a plateau after about 3 years. It is not possible to tell from these data when (if ever) that plateau tails off; the apparent tail-off in both panels of the Figure is due at least in part to the 1989 observational cutoff. For 1975, the higher citation rate for university patents is particularly pronounced in the early years; this pattern is not apparent in the 1980 cohort.

The easiest way to examine the locus of the citing patents is using an assignment code that is provided by the patent office on the public datasets. The code identifies those patents that are unassigned, meaning that the property right resides with the

---

14. Recall that patents are typically granted 1 to 3 years after application; thus a citation lag of 0 or 1 implies that the citing patent may well have been applied for before the originating patent was actually granted. Pending applications are not public, so in this case the citation would almost surely have been identified by the examiner.

1975 Cohort

Citation Lag (difference between application years)

1980 Cohort

Citation lag (difference between application years)

- University citations
- Top Corporate Citations
- Other Corporate Citations

inventor(s), and classifies the remainder according to whether the assignee was U.S. or foreign, and whether it was an individual, a government or a corporation. Table Three compares the assignee distributions of the citation datasets to the universe of all patents. It shows that citations of university patents are themselves more likely to be assigned to universities than the typical patent, and are also more likely to be cited by a patent assigned to the U.S. government.

The identified citations of corporate patents are also slightly more likely to be assigned to a university than a typical patent, probably because the patent class distribution of the corporate originating datasets was chosen to reflect the distribution of university research activity. Hence, these citations are concentrated in areas where universities are important. All of these patents, chosen because they cited a patent assigned to a U.S. university or U.S. firm, are more likely to be themselves assigned to a U.S. university or firm than a randomly drawn patent. Note that self-citations are not excluded from Table Three.

The meaning of a geographic assignment based on assignee is somewhat unclear, however, in a world of multinational corporations. An invention developed at an IBM lab in Switzerland could be categorized as U.S. corporate, while one from a Toyota lab in Kentucky could be categorized as Foreign Corporate.[15] For this reason, and also our interest at looking at smaller geographic units, we turn to the geographic information that relates to the inventors themselves.

---

15. "Could" rather than "would" in each case because the categorization would depend on whether the inventor legally assigned the patent to the parent multinational corporation or to a host-country subsidiary.

TABLE THREE
ASSIGNEE DISTRIBUTIONS

Dataset	Percent Assigned to:					
	Unassigned	Universities	U.S. Corporations	Foreign Corporations	U.S. Government	Other*
Citations to:						
1975 University	12.9	9.9	44.7	26.1	4.5	1.8
1975 Top Corporate	8.5	1.6	62.5	23.9	2.2	1.3
1975 Other Corporate	14.8	1.3	57.9	23.0	1.6	1.3
1980 University	10.5	14.4	47.2	22.4	3.3	2.2
1980 Top Corporate	6.3	2.2	68.4	20.8	1.8	.5
1980 Other Corporate	9.0	3.0	62.9	21.8	1.9	1.4
All 1982 Application Year Patents	16.1	.8	44.0	35.7	1.9	1.7
All 1985 Application Year Patents	16.7	.9	40.1	39.3	1.2	1.6

* includes U.S. and foreign individuals and foreign governments.

III.   The Extent of Geographic Localization

The patent data contain the country of residence of each inventor, and the city and state of residence for U.S. inventors. Use of this information is complicated by the fact that patents can have multiple inventors who can live in different places. The following procedure was followed:

1. For U.S. inventors, city/state combinations were placed in counties using a commercially available city directory; each U.S. inventor was then assigned to an SMSA[16] based on state and county. For this purpose an additional "phantom" SMSA was created in each state, encompassing all counties in the state outside of defined SMSA's. Approximately 98% of inventors were successfully assigned to SMSAs.

2. Assignments of each patent to a country, a state and an SMSA were then made based on pluralities of inventors. So, for example, a patent with one inventor living in Bethesda MD, one in Alexandria VA and one in rural Virginia would be assigned VA for its state and Washington DC for its SMSA. Ties were assigned arbitrarily, except that ties between true SMSAs and phantom SMSAs were resolved for the true one and ties between U.S. and foreign were resolved in favor of foreign.[17]

Having assigned all of the patents to countries, states and SMSAs, we can then ask the question: how often is the citing patent in the same locale (country, state, SMSA) as the originating patent? But to ask that question meaningfully we have to consider how often we would expect them to match under some "null" hypothesis. That

----

16. These assignments were made based on the 1981 SMSA definitions. In areas where Consolidated Metropolitan Statistical Areas were defined in 1981, these were used; elsewhere Metropolitan Statistical Areas were used. Hence we use the generic term "SMSA."

17. At the country level, 98% of patents were assigned unanimously. At the state level, 90% were assigned unanimously; an additional 4% had more than half of inventors in a single state. At the SMSA level, 86% were assigned unanimously, and an additional 6% had a clear majority.

is, we need to compare the probability of a patent matching the originating patent by geographic area, conditional on its citing the originating patent, to the unconditional probability. This unconditional probability gives us a baseline or reference value against which to compare the actual proportions that match. We now consider how to estimate this unconditional or "null hypothesis" probability. As indicated in the introduction, a key issue is the extent to which we allow the null-hypothesis probability to reflect the pre-existing concentration of technological activity.

To be concrete, assume that both the originating and citing patents are drawn randomly from some set P, and that the elements of P are distributed across N distinct geographic areas such that the fraction in area i is f_i, i = 1...N. For the moment we can think of these different areas as countries, states or cities; in each case let f_N be the fraction that are foreign. Suppose first that we choose a sample of originating patents that are, by design, all from a given area i. If there is no geographic relationship between originating and citing patents, then the probability of a match for any given citation is simply the probability that a randomly drawn patent comes from area i, that is, f_i. For the country-level match, that is what we did: all the originating patents are of U.S. origin. Hence the probability of a country-level match under the null hypothesis is the fraction of patents in some (properly chosen) universe P that are domestic or $(1-f_N)$.

The null probability for the state and SMSA matches is slightly more complicated. Suppose now that we chose originating patents at random from the geographic distribution characterized by the f_i's. The expected probability of a match would be the

probability of picking a patent from a given area (f_i), times the probability that the citation is from that area (f_i), summed over all areas. Thus, if originating patents were drawn geographically at random, the null probability would be the sum across areas of the squared area proportions, or the Herfindahl index of concentration across geographic areas. Of course, we did not choose the originating patents at random, because we excluded foreign patents. This implies that the probability of a match if there were no geographic relationship is the sum over i of $[f_i/(1 - f_N)]f_i$ or $(1 - f_N)$ times the Herfindahl index of concentration across states or SMSAs within the U.S. We will refer below to this statistic as an "adjusted domestic Herfindahl."

The question then becomes: what is the appropriate universe P from which originating and citing patents are drawn? This depends on the null hypothesis one wishes to test. One version is that the appropriate P is the universe of all patents. This corresponds to the null hypothesis that there is no relationship between the geographic location of citing and originating patents.

One could argue, however, that this is not an appropriate test of localization of knowledge spillovers. We know from Table One that over half of all citations are in the same patent class as the originating patent. We suspect that the probability of a given geographic location conditional on patent class is not the same as the unconditional. In other words, concentration of inventive activity across geographic areas is probably higher within technical areas than it is over all. Silicon valley has a higher proportion of the world's semiconductor researchers than it does of the world's researchers. Of course, part of the reason for this is probably the existence of knowledge spillovers. To

the extent that is true, the effect of this pre-existing concentration of activity on the probability of a match might appropriately be viewed as part of the phenomenon of interest, and the appropriate null probability would remain at that predicted by drawing at random from the universe of all patents. On the other hand, as discussed above, there may be reasons for this pre-existing concentration other than knowledge spillovers. To this extent, the null probability should reflect the localization predicted by the likelihood that the originating and citing patents are, on average, more technologically similar than two randomly drawn patents.

If citations were always from the same class as originating patents, then we could view the patent class as the appropriate universe P. We could calculate the measures discussed above within each patent class, and then average across patent classes to yield an expected value for the null probability. The results of such calculations are discussed below. But this does not seem quite right either, given that almost half the time citations are not assigned to the same primary class as the originating patent.[18] For this reason, we focus primarily on an alternative method for calculating the null probability. This method allows for localization caused solely by geographic concentration of technologically related activities, but does not rely on the assumption that citing and originating patents are necessarily in the same patent class. Instead, we took every citation that we identified, and drew a control patent for it in the way described in section II above, this time drawing from the universe of all U.S. patents. That is, for

---

18. We also show below that the probability of a geographic match is not, in fact, higher for citations in the same patent class.

each citation, we found another patent in the same class, with the same application year, granted as close in time as possible to the citing patent. This patent has nothing in common with the originating patent, except that it is temporally and technologically very close to the citation. We then examined the frequency with which these control patents came from the U.S., and from the same state and SMSA as the originating patent that was cited by the patent for which the control was drawn, and compared these frequencies to those for the citations. If it were true that citations are close to originating patents only because of the technological areas they represent, then the frequencies with which citations and controls match the originating patents by geographic area should be the same.

Before getting to such comparisons, examination of Table Four is useful merely to get a sense for the extent of geographic concentration. It shows the fraction of patents coming from abroad and from a selection of major U.S. SMSAs for several of the datasets. Not surprisingly, a measurable fraction of university patents comes from Madison, WI; this is not true for corporate patents. Somewhat more substantively, a measurable (though smaller) fraction of the citations of university patents comes from Madison, and this fraction is larger than that for the controls. Indeed, the controls for the university citations look generally "more like" corporate patents than do the citations, suggesting that localization may be present. Other qualitative evidence of localization is apparent in the Table, including the high percentage of NY SMSA citations that come from the NY SMSA.

Quantitative comparisons of the matching proportions are presented in Table

TABLE FOUR
SMSA DISTRIBUTIONS FOR SOME DATASETS

Location	1975 University Originating	1975 Top Corporate Originating	Citations to 1975 University	Citations to 1975 Top Corporate	Controls for Citations to 1975 University	All Citations to Patents Originating in NY SMSA
Foreign	-	-	31.8	31.4	35.8	31.2
Boston	15.0	3.1	7.5	4.6	5.1	4.0
Los Angeles/ Anaheim	7.0	4.8	9.0	5.7	6.1	3.9
S. Francisco/ Oakland	5.1	1.4	3.8	3.7	6.1	3.5
Madison, WI	4.2	-	1.6	-	.5	.6
Philadelphia/ Wilmington	4.2	9.3	5.4	8.2	4.5	9.1
Rural Iowa	3.8	-	1.6	.6	.2	-
San Jose	3.5	2.8	4.0	3.4	-	1.9
New York/ NJ\Conn	3.2	13.5	9.7	11.7	13.7	28.5
Salt Lake City	3.2	-	2.1	-	.5	.4
Detroit/ Ann Arbor	2.6	2.4	2.6	1.7	1.7	1.2
Minneapolis/ St. Paul	1.3	5.2	2.8	2.9	1.9	2.1
Chicago	1.9	4.2	3.9	5.7	5.6	4.2
Albany	.6	3.1	1.9	2.1	1.3	.8

Note: All figures are percentages. SMSA percentages for citations and controls are relative to domestic total.

Five. For each geographic area and each originating dataset, it presents the proportion of citations that geographically matched the originating patent. These proportions are shown both with and without self-citations. The matching proportions for the control samples are then shown, as well as a t-statistic testing the equality of the control proportions and the citation proportions (excluding self-citations).[19] The null-hypothesis probabilities based on Herfindahls are not shown in the Table, but are discussed in the text.

We focus first on the 1975 results on the left of the Table. At the country level, it turns out that the different tests proposed above make little difference. The proportion of all patents that are domestic in the period corresponding to citations of from 1975 is 63.6%; calculating this percentage by patent class and then taking a weighted average across classes using the originating dataset class distribution for weights yields 64.0%. These proportions are quite similar to the predictions based on the fraction of controls that matched, shown in the Table (62.8, 63.1 and 66.3 percent for university, Top Corporate and Other Corporate, respectively).

Table Five shows that, including self-citations, citations are domestic about 6 or

---

19. Let p_c be the probability that a citation comes from the same geographic unit as the originating patent; let p_o be the corresponding probability for a randomly drawn patent in the same patent class (control). We test H_o: $p_c = p_o$ versus H_a: $p_c > p_o$ using the test statistic:

$$t = \frac{\hat{p}_c - \hat{p}_o}{\sqrt{[\hat{p}_c(1 - \hat{p}_c) + \hat{p}_o(1 - \hat{p}_o)]/n}}$$

where $\hat{p}_c$ and $\hat{p}_o$ are the sample proportion estimates of p_c and p_o. This statistic tests for the difference between two independently drawn binomial proportions; it is distributed as t.

TABLE FIVE
GEOGRAPHIC MATCHING FRACTIONS

Originating Cohorts

| | 1975 | | | | 1980 | |
	University	Top Corporate	Other Corporate	University	Top Corporate	Other Corporate
Number of Citations	1759	1235	1050	2046	1614	1210

Matching by Country

Overall Citation Matching Percentage	68.3	68.7	71.7	71.4	74.6	73.0
Citations Excluding Self-cites	66.5	62.9	69.5	69.3	68.9	70.4
Controls	62.8	63.1	66.3	58.5	60.0	59.6
t-statistic	2.28	-0.1	1.61	7.24	5.31	5.59

Matching by State

Overall Citation Matching Percentage	10.4	18.9	15.4	16.3	27.3	18.4
Citations Excluding Self-cites	6.0	6.8	10.7	10.5	13.6	11.3
Controls	2.9	6.8	6.4	4.1	7.0	5.2
t-statistic	4.55	.09	3.50	7.90	6.28	5.51

Matching by SMSA

Overall Citation Matching Percentage	8.6	16.9	13.3	12.6	21.9	14.3
Citations Excluding Self-cites	4.3	4.5	8.7	6.9	8.8	7.0
Controls	1.0	1.3	1.2	1.1	3.6	2.3
t-statistic	6.43	4.80	8.24	9.57	6.28	5.52

Note: Number of citations is less than in Table One because of missing geographic data for some patents. The t-statistic tests equality of the citation proportion excluding self-cites and the control proportion. See text for details.

7 percent more often than the controls. Excluding self-citations eliminates this difference for the Top Corporate citations and cuts it roughly in half for the others. The remaining difference between the citations excluding self-cites and the controls is only marginally significant statistically.

Looking at the 1975 results for states, we find that citations of university patents come from the same state about 10 percent of the time; this rises to 15% for Other Corporate and 19% for Top Corporate. Excluding self-citations, however, makes a big difference. The university and Top Corporate proportions are cut to 6-7 percent, and the Other Corporate to just over 10. For comparison, the adjusted domestic Herfindahl across states is 4.3 percent for the universe of patents, and is 6.5% for the weighted average of within-patent-class values. The latter of these two figures is again quite close to the actual match frequency using the control patents. For the university and Other Corporate cohorts, the matching frequencies excluding self-citations are significantly greater than the matching control proportions.

At the SMSA level, 9 to 17 percent of total citations are localized. This again drops significantly when self-citations are excluded, but 4.3 percent of university citations, 4.5 percent of Top Corporate citations and 8.7 percent of Other Corporate citations are localized excluding self-cites. This compares to control matching proportions of about 1 percent, and these differences are highly significant. The overall adjusted domestic Herfindahl at the SMSA level is about 1.6%; the within-patent-class Herfindahl is about 3.4%. Note that the latter is higher than the control frequency reported in the Table, and is not significantly different from the citation matching frequencies except for Other

Corporate.

The results for citations of 1980 patents are even stronger and more significant. For every dataset, for every geographic level, the citations are quantitatively and statistically significantly more localized than the controls. It is well known that the proportion of all U.S. patents taken by foreigners has been increasing; this is reflected in a decline of 3 to 6 percent in the control percentages matching by country. The citation matching percentages actually rise, however, particularly for Top Corporate citations. It is impossible to tell from this comparison whether this represents a real change, or whether it is the result of the 1980 citations having shorter average citation lags. Since this gets to the issue of explaining which citations are localized, we postpone discussion until the next section.

Before moving on, the results on the extent of localization can be summarized as follows. For citations observed by 1989 of 1980 patents, there is a clear pattern of localization at the country, state and SMSA levels. Citations are 5 to 10 times as likely to come from the same SMSA as control patents; 2 to 6 times as likely excluding self-citations. They are 3 to 4 times as likely to come from the same state as the originating patent; roughly twice as likely excluding self-cites. Whereas about 60 percent of control patents are domestic, 70 to 75 percent of citations and 69 to 70 percent of citations excluding self-cites are domestic. Once self-cites are excluded, universities and firms have about the same domestic citation fraction; at the state and SMSA level there is weak evidence that university citations are less localized. For citations of 1975 patents, the same pattern, but weaker, emerges for citations of university and Other Corporate

patents. For Top Corporate, there is no evidence of localization at the state or country levels, though the SMSA fraction is significantly localized. Thus we find significant evidence that citations are even more localized than one would expect based on the pre-existing concentration of technological activity, particularly in the early years after the originating patent.

IV. Factors Affecting the Probability of Localization

The contrast between the 1975 and 1980 results suggests that localization of early citations is more likely than localization of later ones. This accords with intuition, since whatever advantages are created by geographic proximity for learning about the work of others should fade as the work is used and disseminated. Another hypothesis that is implicit in the previous discussion is that citations that represent research that is technologically similar to the originating research are more likely to be localized, because the individuals pursuing these related research lines may be localized. In addition, attributes of the originating invention or the institution that produced it may affect the probability that its spillovers are localized.

To explore these issues, we pooled the citations (excluding self-cites) to university and corporate patents for each cohort, and ran a probit estimation with geographic match/no match between the originating and citing patents as the dependent variable. As independent variables we included the log of the citation lag (set to zero for lags of zero), dummy variables for Top Corporate and Other Corporate originating patents, interactions of the lag and these dummies, and a dummy variable equal to unity if the

citation has the same primary class as the originating patent. To prevent the measurement of the effect of time from being contaminated by the fact that all patents are becoming more likely to be foreign over time, we included as a control a dummy variable that is unity if the control patent corresponding to this citation matches geographically with the originating patent.

We also included two variables relating to the originating patent suggested by our work on basicness and appropriability of inventions (Trajtenberg, Henderson, and Jaffe, 1992). The first, "generality" is one minus the Herfindahl index across patent classes of the citations received. It attempts to capture the extent to which the technological "children" of an originating patent are diverse in terms of their own technological location. Thus an originating patent with generality approaching 1 has citations that are very widely dispersed across patent classes; generality of zero corresponds to all citations in a single class. We argue elsewhere that generality is one aspect of the "basicness" of an invention. One might hypothesize that basic research results are less likely to be localized, because there spread is more likely to be through communication mechanisms (e.g. journals) that are not localized. The other variable characterizing the originating invention is the fraction of the originating patent's citations that were self-cites. We take a high proportion of self-cites as evidence of relatively successful efforts by the original inventor to appropriate the invention. We expect that the non-self-citations to such a patent are more likely to be confined to suppliers, customers, or other firms that the inventing firm has a relationship with, and may therefore tend to be localized.

Finally, the extent of localization depends fundamentally on the mechanisms by

which information flows, and these mechanisms may be different in different technical fields. For this reason, we also included dummy variables for the broad technological fields discussed above in the context of Table Two.

The results are presented in Table Six. Because of the presence of the interaction terms between the lag and the corporate dummies, the coefficient on the lag itself corresponds to the fading of localization of citations of university patents. There is evidence in the 1975 results of such fading. This effect is statistically significant at the state and SMSA levels; its quantitative significance is discussed further below. For the citations of corporate patents, the interaction terms measure the difference between their fading rates and those of university citations. These terms are generally not statistically significant. In only one case (Other Corporate, 1975) could we reject the hypothesis of equality of fading rates at traditional confidence levels. There is, however, weak evidence that the corporate citations do not fade as rapidly as those of university patents, at least at the state and SMSA levels. The coefficients on the corporate dummies themselves capture differences in the predicted probability of localization for citations with lags of 0 or 1 year. These are all insignificant, and there is no clear pattern.

The matching patent class and generality measures do not work well. The effects are generally insignificant, and show no consistent pattern. The effect of the self-citation fraction is, however, strong and puzzling. At the state and local level, there is a very significant effect in the predicted direction: citations of patents with a high self-citation fraction are more likely to be localized. This is not just saying that self-citations are localized, since they are excluded; it is the other citations that are more localized. At

TABLE SIX
GEOGRAPHIC PROBIT RESULTS

	Country Match		State Match		SMSA Match	
	1975	1980	1975	1980	1975	1980
Dummy for Control Sample Match	.139 (.045)	.085 (.041)	.396 (.124)	.300 (.102)	*	.283 (.172)
Log of Citation Lag	-.078 (.049)	.094 (.056)	-.264 (.073)	.198 (.079)	-.123 (.057)	.037 (.086)
Dummy for Top Corporate	-.114 (.168)	-.010 (.127)	-.383 (.249)	.013 (.177)	-.234 (.288)	-.208 (.200)
Dummy for Other Corporate	.069 (.209)	.053 (.134)	-.214 (.277)	-.007 (.189)	.325 (.291)	-.042 (.207)
Log-lag *Top Corp. Dummy	.046 (.091)	-.016 (.086)	.226 (.138)	.007 (.115)	.102 (.156)	.156 (.131)
Log-lag *Other Corp. Dummy	.008 (.108)	-.026 (.091)	.307 (.147)	.036 (.124)	.037 (.155)	.039 (.138)
Dummy for Matching Patent Class	-.085 (.050)	.069 (.045)	-.013 (.073)	.034 (.058)	-.057 (.080)	-.016 (.068)
Generality of Origin Patent	.092 (.091)	.177 (.088)	.026 (.136)	-.140 (.111)	.013 (.150)	-.298 (.130)
Origin Fraction Self-citations	-.813 (.180)	.162 (.124)	.815 (.246)	.883 (.134)	1.174 (.237)	.828 (.154)
# of Observations	3581	4217	3573	4215	3566	3972
# of Matches	2363	2925	256	490	197	298
Log Likelihood	-2269	-2559	-894.2	-1459	-735.8	-1022

* The number of observations for which the control patent matched at the SMSA level was so small that this parameter could not be estimated.

Standard errors in parentheses. All equations also included 5 technological field dummies.

the country level, at least in 1975, this effect is reversed and is significant. Taking all results together it suggests that for patents with a lot of self-citations, the non-self-citations are more likely to be foreign, but those that are domestic are more likely to be in the same state and SMSA as the originating patent.

The 1980 results are disappointing. The coefficient on the time lag term switches sign, though it is generally insignificant. One possibility is that these citations span too short a time period to capture the lag effect well. To test this possibility, we re-ran the estimation in Table Six on the 1975 citations, excluding all that were granted after 1984. These results show what we would have believed about citations of 1975 patents if we had looked for them only as long as we have looked for the citations of the 1980 patents. The results are presented in Table Seven. As expected, they look "more like" the 1980 results than the original 1975 results did. In particular, the coefficient on the lag term is now insignificant, and is positive at the SMSA level. Thus it may be that the "perverse" results for the 1980 sample would go away if we had later citations to include.

A probit coefficient does not have an economically meaningful magnitude, because of the need to standardize the variance of the underlying error distribution. We can, however, calculate what the coefficients imply about changes in the predicted probabilities. This is done in Table Eight, using the 1975 lag coefficient.[20] Table Eight was constructed by calculating the predicted localization probability using the results of

---

20. As discussed above, this is the point estimate of the lag coefficient for citations of university patents. The point estimates are different for the corporate originating patents, but since these differences are generally insignificant we have not performed separate calculations for each dataset.

TABLE SEVEN
Geographic Probit Results

"Trimmmed" 1975 Citation Sample

	Country Match	State Match	SMSA Match
Dummy for Control Sample Match	.249 (.061)	.456 (.158)	*
Log of Citation Lag	-.077 (.072)	-.088 (.100)	.093 (.117)
Dummy for Top Corporate	-.361 (.212)	-.467 (.327)	-.326 (.397)
Dummy for Other Corporate	-.104 (.294)	-.403 (.410)	.197 (.425)
Log-lag *Top Corp. Dummy	.275 (.144)	.252 (.218)	.118 (.261)
Log-lag *Other Corp. Dummy	.137 (.190)	.346 (.260)	.026 (.271)
Dummy for Matching Patent Class	-.136 (.067)	-.149 (.095)	-.187 (.103)
Generality of Origin Patent	-.018 (.124)	-.130 (.175)	-.180 (.193)
Origin Fraction Self-citations	-.739 (.234)	.655 (.310)	.985 (.318)
# of Observations	2005	2003	1986
# of Matches	1354	166	122
Log Likelihood	-1244	-560.9	-444.7

* The number of observations for which the control patent matched at the SMSA level was so small that this parameter could not be estimated.

Standard errors in parenthesis. All equations also included 5 technological field dummies.

TABLE EIGHT

PREDICTED LOCALIZATION PERCENTAGES OVER TIME

Based on 1975 Probit Results
for Citations of University Patents

Predicted Percentage for:

Lag	Same Country	Same State	Same SMSA
0 or 1 Year	67.1	9.7	4.8
5 Years	65.5	6.5	4.0
10 Years	64.6	5.3	3.7
25 Years	63.5	4.0	3.3

Table Six, evaluating the citation lag at different values, and evaluating the other independent variables at the mean of the data. It shows that the estimates correspond to a reduction in the localization fraction after, for, for example, 10 years, from 67.1% to 64.6% at the country level, 9.7% to 5.3% at the state level, and 4.8% to 3.7% at the SMSA level.

V.   Discussion and Conclusion

Despite the invisibility of knowledge spillovers, they do leave a paper trail in the form of citations. We find evidence that these trails, at least, are geographically localized. The results, particularly for the 1980 cohort, suggest that these effects are quite large and quite significant statistically. Because of our interest in true externalities, we have focussed on citations excluding self-cites. For some purposes, however, this is probably overly conservative. From the point of view of the Regional Development Administrator, it may not matter whether the subsequent development that flows from an invention is performed by the inventing firm, as long as it is performed in her state or city. Our results are also conservative because we attribute none of the localization present in the control samples to spillovers, despite the likelihood that spillovers are, indeed, one of the major reasons for the pre-existing concentration of research activity.

We also find evidence that geographic localization fades over time. The 1980 citations, which have shorter average citation lags, are systematically more localized than the 1975 citations. By using a probit analysis, we produced estimates of the rate of fading. These estimates seem to suggest a rate of fading that is both smaller than one

would expect, and smaller than would be necessary to explain the difference between the 1975 and 1980 overall matching fractions. One possibility is that the difficulty of measuring the rate of fading is due to the "contamination" of citations by the patent examiner. As noted above (Footnote 14), it is particularly likely that citations with very short lags were added by the examiner. If we believe that such citations are less likely to represent spillovers and less likely to be localized, then this would tend to bias towards zero our measure of the effect of time on localization.

We find less evidence of the effect of technological area on the localization process. Citations in the same class are no more likely to be localized. These non-results are also consistent with the relative insensitivity of our estimates of the "null" probabilities to whether or not we look within classes. Overall, there is not really any evidence in these data that the probability of coming from a given geographic location conditional on patent class is different from the unconditional probability. This may be due to the arbitrary use of the "primary" patent class, to the exclusion of the "cross-referenced" classes. There is no legal difference in significance between the primary and cross-referenced classes, and in many cases the examiners do not place any significance on which class is designated primary. In future work, we hope to explore whether using the full range of information contained in the cross-referenced classes provides a better technological characterization of the patents.

In this context, it is worth noting that part of what is going on is probably that knowledge spillovers are not confined to closely related regions of technology space. As shown in Table 2, citations come to some extent from different technological areas, even

at a very broad level of technological categorization. This is consistent with previous research (Jaffe, 1986) that found that a significant fraction of the total "flow" of spillovers affecting firms' own research productivity comes from firms outside of the receiving firm's immediate technological neighborhood.

We find surprisingly little evidence of differences in localization between the citations of university and corporate patents. The largest difference is that corporate patents are more often self-cited, and self-cites are more often localized. The probit results do not allow rejection of the hypothesis that the initial localization rates for non-self-citations are indistinguishable for the different groups. They do provide some weak evidence that this initial localization is more likely to fade for the university patents, at least at the state and local levels.

In order to provide a true foundation for public policy and economic theorizing, we would ultimately like to be able to say more about the mechanisms of knowledge transfer, and about something resembling social rates of return at different levels of geographic aggregation. The limitations of patent and citation data make it difficult to go much further with such questions within this research approach. *Ex post*, the vast majority of patents are seen to generate negligible private (and probably social) returns. In future work, we plan to identify a small number of patents that are extremely highly cited. It is likely that such patents are both technologically and economically important (Trajtenberg, 1990). Case studies of such patents and their citations could prove highly informative about both the mechanisms of knowledge transfer, and the extent to which citations do indeed correspond to externalities in an economic sense.

BIBLIOGRAPHY

Bania, N.: "Technological Spillovers and Innovation in Research and Development Laboratories," *REI Working Paper Series*, Center for Regional Economic Issues, Case-Western Reserve University, 1989

Bania, N., R. Eberts and M. Fogarty: "The Role of Technical Capital in Regional Growth," presented at the Western Economic Association Meetings, July 1987.

Bernstein, J.I. and M.I. Nadiri: "Interindustry R&D Spillovers, Rates of Return, and Production in High-Tech Industries," American Economic Review Papers and Proceedings, 1988

Bernstein, J.I. and M.I. Nadiri: "Research and Development and Intra-industry Spillovers: An Empirical Application of Dynamic Duality," Review of Economic Studies, 1989

Carpenter, M. et. al.: "Citation Rates to Technologically Important Patents," World Patent Information, Vol. 3. No. 4., 1980.

Carpenter, M. and F. Narin.: "Validation Study: Patent Citations as Indicators of Science and Foreign Dependence," World Patent Information, Vol.5, No. 3, pp. 180-185, 1983.

Dorfman, N: "Route 128: The Development of a Regional High-Technology Economy" in D. Lampe (ed.), The Massachusetts Miracle: High Technology and Economic Revitalization, MIT Press, 1988

Feller, Irwin, "R&D Theories and State Advanced Technology Programs," Paper prepared for the American Association for the Advancement of Science Annual Meeting, January 1989

Glaeser, Edward, H.D. Kallal, J.A. Scheinkman and A. Shleifer: "Growth in Cities," National Bureau of Economic Research Working Paper No. 3787, July 1991

Griliches, Z.: "Patent Statistics as Economic Indicators: a Survey," Journal of Economic Literature, 28, page 1661, 1990

Griliches, Z.: "The Search for R&D Spillovers," National Bureau of Economic Research Working Paper No. 3768, July 1991

Grossman, G. and E. Helpman, Innovation and Growth in the Global Economy, Cambridge: M.I.T. Press, 1991

Jacobs, Jane: The Economy of Cities, New York, Vintage Books, 1969

Jaffe, A.: "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits and Market Value," American Economic Review, 1986

Jaffe, A.: "Real Effects of Academic Research," American Economic Review, 1989.

Krugman, P.: Geography and Trade, Cambridge: M.I.T. Press, 1991

Mansfield, E.: "Sources and Characteristics of Academic Research Underlying Industrial Innovations," mimeo, University of Pennsylvania, 1991

Marshall, A.: Principles of Economics, London: MacMillan, 1920

Minnesota Department of Trade and Economic Development, Office of Science and Technology, State Technology Programs in the United States, 1988

Romer, P.: "Increasing Returns and Long-Run Growth," Journal of Political Economy, October 1986

Romer, P.: "Endogenous Technological Change," Journal of Political Economy, 1990

Shimshoni, D.: "Regional Development and Science-Based Industry," in J. Kain and J. Meyer (eds.), Essays in Regional Economics, Harvard University Press, 1971

Smilor, R., G. Kozmetsky and D. Gibson: Creating the Technopolis: Linking Technology, Commercialization and Economic Development, Ballinger Publishing Co., 1988

Teplitz, P.: Spin-off Enterprises form a Large Government Sponsored Laboratory, unpublished M.S. thesis, M.I.T., 1965

Trajtenberg, M.: "A Penny for Your Quotes: Patent Citations and the Value of Innovations," Rand Journal of Economics, 1990a, Vol. 21, No. 1, pp 172-187

Trajtenberg, M.: Economic Analysis of Product Innovation: The Case of CT Scanners, Harvard University Press, 1990

Trajtenberg, M., R. Henderson and A. Jaffe: "Quantifying Basicness and Appropriability of Innovations with the Aid of Patent Data -- A Comparison of University and corporate Research," National Bureau of Economic Research Working Paper No. xxxx, 1992

Wainer, H.: The Spin-off of Technology from Government Sponsored Research Laboratories: Lincoln Laboratory, unpublished M.S. thesis, M.I.T., 1965

</ref_section>