WORKING PAPER

It's a Small(er) World:
The Role of Geography in Biotechnology Innovation

By

Daniel K.N. Johnson

Colorado College Working Paper 2006-03

August, 2006

Department of Economics and Business

Colorado College

Colorado Springs, Colorado 80903-3298

www.coloradocollege.edu/dept/EC

It's a Small(er) World: The Role of Geography in Biotechnology Innovation

Abstract

Using patent citation data for the U.S., we test whether knowledge spillovers in biotechnology are sensitive to distance, and whether that sensitivity has changed over time. Controlling for self-citation by inventor, assignee and examiner, cohort-based regression analysis shows that physical distance is becoming less important for spillovers with time.

JEL codes: L6 --- Industry Studies: Manufacturing
N9 --- Regional and Urban History
O3 --- Technological Change
R1 --- Urban, Rural and Regional Economics

Daniel K.N. Johnson is an Assistant Professor of Economics at Colorado College, 14 East Cache La Poudre Street, Colorado Springs, CO 80903, tel: (719) 389-6654, fax: (719) 389-6927, email:djohnson@coloradocollege.edu.

Special thanks to Milena Mareva and David Popp for suggestions about data and programming. Financial support was provided by a National Science Foundation Award for the Integration of Research and Education, a Wellesley College Faculty Grant, a Colorado College Faculty Grant and a Mrachek Research Fellowship.

It's a Small(er) World:
The Role of Geography in Biotechnology Innovation

I. Introduction

It has long been noted that firms within an industry often cluster geographically, for a variety of reasons. Localization economies, or Marshall-Arrow-Romer externalities which reduce the cost of inputs to firms in the local industry, have been studied in a variety of contexts (see for example Henderson (1986) or Smith and Florida (1994)). For some industries, it is the nature of the knowledge itself, its speed of progress and the degree to which it is tacit, that encourages firms to locate near other firms in the sector. Both of these aspects of knowledge, diffusion speed and tacitness, have been identified in the literature (see Caballero and Jaffe (1993) for diffusion speed and Von Hippel (1994) for tacitness) but to our knowledge no analysis has tested their change over time, nor measured their strength in biotechnology.

This paper examines knowledge flows within biotechnology, confirming the traditional evidence that inter-firm knowledge transfers decrease with distance, while adding the important caveat that the impact of physical distance has been diminishing over time. Thus, historically there has been a knowledge spillover-based reason for biotechnology firms to cluster geographically, although that reason is less powerful now than ever before.

Using all biotechnology patents granted in the U.S. between 1975 and 1994, we show that there is a marked tendency to cite patents from nearby areas. Thus, this paper follows the "paper trail" literature (Feldman, 1999) on inter-firm knowledge transfers or "spillovers" in documenting the creation of knowledge via patent citations. We test whether this pattern could naturally arise from a tendency to cite other patents listing the same inventor, the same firm assignee, the same patent examiner, or the same technology class. We conclude that the geographic clustering of citations holds over and above the effects of these other factors, suggesting that there is a local nature to knowledge spillovers (at least insofar as patent citations reflect knowledge flows).

In section II of the paper, we briefly review the relevant literature on biotechnology clustering and the geographic nature of knowledge spillovers. Section III describes our data set, designed for compatibility with the literature, and Section IV presents regression analysis that controls for non-geographic effects in presenting the declining role of distance. Section V concludes with implications for policy and further research.

II. Literature review

The literature suggests that knowledge spillovers cluster geographically, with higher spillovers (shown by more patent citations) within a short distance. The underlying supposition is that inventors are more aware of (or find more use for) inventions located close to them, and therefore build more heavily on local inventions, the result being a geographic clustering of citations.

Empirical evidence stresses the important role of geography in the spillover of knowledge from one member of an innovation network to another (see for example a review by Gelsing, 1992). It also emphasizes the importance of frequent personal contact and research collaboration. In particular, Zucker et al. (1998) show that firms using biotechnology are strongly influenced by the location of superstars in academic research institutions.

Lundvall (1992) points out that the importance of geography should differ predictably by technology type. While geography has little impact on stationary technologies (facing constant needs and opportunities), that importance grows quickly for technologies undergoing incremental innovation and radical innovation. During technological revolutions, there is a dramatic effect on the geographic pattern of subsequent innovations. Since biotechnology has enjoyed aggressive growth, we expect a large geographic impact on knowledge flows.

Geographic proximity has already been used to explain the location of R&D-intensive activities (for Boston's high-technology district, see Dorfman (1988); for biotechnology in particular see Zucker et al., (1998); for France see Carrincazeaux et al., (2001)), due to evidence of localized spillovers within an industry. However, the location of *firms* is not always a good predictor of the location of *innovation*. Feldman (1994) tests whether U.S. state patent levels can be predicted simply by the presence of particular industries in that state. Test statistics are decisive, perhaps because industries have different propensities to patent, so that in fact there is only a 0.42 correlation between innovation measures and value-added in each state by industry. This result is confirmed for the 1975-1994 period in the U.S. (Johnson and Brown, 2004) in an exploration of why the northeastern states lost a dramatic share of the national patenting total. It occurred not only due to the location of industries, but also due to industries that were unable to maintain the patenting rates enjoyed by the same sector in other regions.

Localization of patent citations has been firmly established by the leading paper on the topic (Jaffe et al., 1993), with a random sample of patents clearly more likely to cite local patents than others at every geographically aggregated level. The effects are small but statistically significant, and are more intense where knowledge becomes obsolescent rapidly, like electronics, optics and nuclear technology (Jaffe and Trajtenberg, 1996). The result has been confirmed for semiconductors (Almeida and Kogut, 1997).

Since biotechnology knowledge becomes obsolescent very rapidly (see Johnson and Santaniello, 2000), one might expect that it will follow the same pattern. However, two factors augur against this quick conclusion. First of all, most biotechnological information is not tacit, so will be relatively easy to communicate across long distances. Second, biotechnology patenting has occurred largely during a period when international and inter-regional communication has been increasingly effective and affordable, so once again we might expect less localization of knowledge spillovers (Feldman, 1999).

Other researchers have demonstrated a geographic pattern to European patent citations. In a limited sample of Swedish patent applications, international trade flows, rather than physical distance, was the only variable that robustly explained international references (Sjoholm, 1996). In a larger study of over 100,000 patent citations between European regions, there is strong evidence of geographic clustering (Maurseth and Verspagen, 1999). Regressions show that distance between regions is an important driving factor, along with technological similarity between regions.

III. Data

Measurement issues

This paper relies exclusively on patent citations from biotechnology patents as a geographic measure of knowledge spillovers in the sector. When a patent application is submitted for approval, it is accompanied by a list of citations to other patents and literature which have been instrumental in the creation of this technology, or which delineate the legal limits of this application. The intention is twofold: to build a convincing case that this application is novel and unobvious to someone trained in the field, and to provide a legal record of materials consulted during the invention process in order to protect patent rights in the future. To this list of citations, a patent examiner may add his or her own list of citations. The result is a paper trail of knowledge creation.

Of course, patents records do not perfectly reflect the creation of technology, as some innovations are never patented and patents vary greatly in size and importance. However, within the U.S. on a state-by-state level, patents have a high correlation with other measures of innovative activity. For example, there is a 0.88 correlation between patents and R&D expenditures, 0.99 between patents and research employment records, and 0.93 between patents and a census of innovation citations in scientific and trade journals conducted by the Small Business Administration (Feldman, 1994).

Citations themselves do not perfectly reflect the transfer of knowledge, as they may be inserted for a variety of other reasons including legal protection or examiner privilege. Jaffe et al. (2000) relates survey evidence showing that only ¼ of all patent citations correspond to a clear spillover of knowledge, another ¼ have some possibility of a spillover, and the remaining ½ do not reflect knowledge transfers. However, their statistical tests indicate that overall citations can be interpreted as a signal of spillovers, albeit a noisy signal.

As a final definitional challenge, "biotechnology" definitions differ between nations and over time (see Johnson and Santaniello, 2000). Therefore, we follow the most recent published biotechnology definitions of the U.S. Patent Office (USPTO, 1998), which include portions of eleven separate classes from the U.S. patent classification system.[1]

III.B. Clustering of knowledge citations

Patent citations may also cluster for non-geographic reasons, coincidentally causing a pattern which appears geographic merely through correlation with other phenomena. For example, inventors may be more familiar with their own patents, citing them more frequently than others, which would give a biased impression of the importance of geography. The same may be true of assignees, if employees of a firm are familiar with other patents held by the firm. While inventor and assignee self-citation may drive a pattern of geographic clustering, they confuse the issue of "local knowledge spillovers," which is the primary focus of our analysis, so we describe and separate it. On the other hand, we do not wish to simply ignore self-citations as being obviously local. If an assignee firm is located in several different locations, high familiarity with other inventions by the same assignee may actually work against a geographic clustering of citations. The same may be true of an inventor who moves during his or her career. Therefore we include self-citations in the analysis but control for them separately.

[1] Specifically, the definition includes U.S. Patent Classes 47/1.1-47/1.4, 47/57.6-47758, 424/9.1-424/9.2, 424/9.34-424/9.81, 424/85.1-424/94.67, 424/130.1-424/283.1, 424/520-424/583, 424/800-424/832, 435/1.1-435/7.95, 435/40.5-435/261, 435/317.1-435/975, 436/500-436/829, 514/2-514/22, 514/44,

All patent citations are reviewed, revised and potentially appended by examiners at the U.S. Patent Office. Due to the nature of patent records, it is impossible to verify whether a given citation was originally submitted by the applicant or added by an examiner, so we must treat examiners as another potential source of geographic clustering. While examiners may have less geographic concentration in their knowledge, they may feel more familiarity with patents that they have examined than with patents that others have examined. This potentially introduces a bias through differences in the geographic zones of examiner caseloads. Since applicants do not know which examiner will be assigned to their case, it is unlikely that applicants will include a large number of citations to any particular examiner. Thus, we can infer an "examiner self-citation effect" to distinguish it from any geographic pattern we may observe.

Using U.S. patent data from a combination of sources (NBER website as described in Hall et al., 2001 in addition to raw data collected by the independent firm MicroPatent), we collected citations from all biotechnology patents granted between 1975 and 1994. We then traced all self-citations by inventors, allowing for some flexibility in name spellings (since the USPTO does not standardize name format). These include not only first inventors, but all inventors listed for each patent. We found that self-citation accounted for almost precisely one percent of all citations from biotechnology patents, suggesting that while some self-citation is present, there are strong inter-inventor knowledge spillovers. Unlike academic citations, there is very little reason here to self-cite as a means of advertising, so we can be fairly sure that self-citations are indicators of useful capital or legal protection.

514/783, 530/300-530/427, 530/800-530/868, 536/1.11-536/23.74, 536/25.1-536/25.2, 800, 930, 935. We

To add to this measure, we investigated citations between assignees. Using the same biotechnology patents, we found that nine percent of all citations were to the same assignee firm or person, again checking rigorously for variations of firm names. This share varied from nearly 14 percent in 1975-79 to a low of less than 8 percent in 1985-89, with no obvious trend. This intra-firm pattern is obviously an important component of the flow of knowledge, but they still imply that 9 of 10 citations are to other firms.

Conceivably, those citations to other firms and inventors were simply added by patent examiners, with no relationship to knowledge flows between inventors. We found that five percent of all biotechnology citations were made to other patents sharing the same examiner, with considerable variation between examiners. In fact, for one examiner over seventy percent of the citations made to patents he examined hail from other patents he reviewed. Naturally, citations should be made to similar technologies, and as a result biotechnology examiners should cite one another frequently, and conceivably, should cite their own relevant reviews most of all. However, this may also be evidence that some examiners are inserting or confirming references to material with which they are personally familiar. While primarily of interest to the U.S. Patent Office, it also offers insight into a weakness of using citations as measures of knowledge flow.

IV. Model and estimation

Our regression analysis extends the seminal work of Jaffe and his co-authors (Caballero and Jaffe, 1993; Jaffe and Trajtenberg, 1996), building upon a basic model with extensions that has been used successfully for other purposes. We add to the analysis by allowing coefficients to vary over time, and by directly controlling for the

exclude class PLT (plant patents) due to data limitations on these documents.

potential impact of technological similarity between regions and self-citation by inventors, assignees and examiners. The model is a recognition that the likelihood that a patent k granted in year t will be cited by a subsequent patent, K, granted in year T, is at least in part a function of the attributes of patents k and K. Using exponential rates of decay and diffusion to model the flow of knowledge over time, that probability can be written as:

$$p(k,K) = \alpha(k,K)\delta(k,K)\exp[-\beta_1(\sum_{s=t_A}^{T} P_s)][1-\exp(-\beta_2(T-t)] \qquad \textbf{(1)}$$

where $\alpha(k,K)$ represents the non-geographic attributes of patents k and K that affect the probability of citation, while $\delta(k,K)$ represents the relevant geographic attributes. β_1 represents the decay rate of knowledge, permitting the possibility that older patents are cited less frequently simply because they are older, and so may no longer be at the leading edge of their art (we follow the literature in modeling decay based on the knowledge stock available, or the number of patents P granted between the application date t_A of the cited patent and the grant date T of the citing patent). β_2 represents the rate of diffusion of knowledge, incorporating the possibility that it takes some time for new innovations to be recognized as noteworthy, (again, we follow the literature and base this on the time lag between citing and cited patents, measured in years). Both exponential terms naturally depend (directly or indirectly) upon the time elapsed between granting of the cited and citing patents. In addition, there is ample evidence of "citation inflation" over time (e.g. Johnson and Popp 2003), with recent patents recording longer citation lists than earlier patents do, a fact which time-based dummy variables will capture in our analysis.

This statistical model has an illustrious history in the literature, and has proven quite effective at describing citation patterns (e.g. see Johnson and Popp 2003 for justification and history of the model). The only real novelty introduced here is the particular set of control variables and geographic variables chosen for inclusion. In particular, we include six control variables (α parameters) and one geographic (δ) variable:

- whether or not patents *k* and *K* have the same inventor (α_{SV}),
- whether or not patents *k* and *K* have the same assignee (α_{SA}),
- whether or not patents *k* and *K* have the same examiner (α_{SE}),
- whether or not patents *k* and *K* have the same technology class (α_{SI}),
- pendency lag of cited patent *k* (α_{LAG}),
- year T of citing patent *K*, to account for citation inflation (α_T), and
- distance between state origins of patents *k* and *K* (δ_D).

While most of the variables listed are binary in nature, and thus are easy to incorporate into a cohort-based estimation process (including the year T, which is incorporated as a series of dummy year variables T_i), there are two notable exceptions.

First, pendency lag is a continuous variable so potentially cited patents were grouped into four categories. Short (lag <=3 years), medium (3 < lag <=5 years), long (5 < lag <=10 years) and very long (lag>10 years) are defined to reflect the broad differences.

Second, physical distance is a continuous variable so *(k,K)* patent pairs are grouped into 100-kilometer cohorts ranging from high distance (over 2,300 kilometers between state capitals of *k* and *K*) to low distance (less than 100 kilometers), for twenty-four distance groupings.[2]

[2] Sensitivity tests performed to include 10 groupings (1-250 km, 251-500 km, etc.) produced similar results.

For the sake of computability, we group the data into five-year time periods for α_T (1975-79, 1980-84, 1985-89, and 1990-94), to allow for some change in the key relationships over time.

Therefore, we postulate the functional forms

$$\alpha(k,K) = \alpha_{SV}\alpha_{SE}\alpha_{SA}\alpha_{SI}\alpha_{LAG}\alpha_T \qquad \textbf{(2)}$$

$$\delta(k,K) = \delta^D \qquad \textbf{(3)}$$

where δ is a single parameter raised to the power D, the number of units of physical distance between citing and cited locations. Therefore, a distance of 200 km is associated with an estimated coefficient based on δ^2. Estimations based on a more flexible functional form of $\delta(k,K)$, allowing each distance to enjoy its own independent δ coefficient, produce similar but less parsimonious results.

Using these parameters, the probability of a patent *k* granted in year *t* being cited by a patent *K*, granted in year *T*, can be estimated as:

$$p_{k,K} = \alpha(k,K)\delta(k,K)\exp[-\beta_1(\sum_{s=t_A}^{T} P_s)][1-\exp(-\beta_2(T-t))] + \varepsilon_{k,K} \qquad \textbf{(4)}$$

A true geographic effect of clustering would be evidenced by a δ value less than unity (i.e. a lower probability of citation when the distance between citing and cited patent is high), since other relevant attributes of patents *k* and *K* have been controlled.

Since most patents are never cited, if we were to estimate equation **(4)** for individual pairings of citing/cited patents, the dependent variable would be zero for the vast majority of all observations. Thus, following the work of Jaffe and his coauthors, we group the patents into “cohorts” of potential citations. Cohorts are constructed as mutually exclusive and exhaustive subsections of all possible citations, and therefore include counts of simultaneous citation (one patent citing another within the same application year), self-citation, and citations to patents in the same geographical area, as well as citations to patents dramatically different than the citing patent. For example, one cohort may be defined as “all patents granted in 1976 and cited in 1978 by a patent that shares the same inventor, the same assignee, a different examiner, a different technology group, and hails from a state within 100 kilometers with a highly similar technology profile”.

The expected number of citations to a cohort with specific values for the independent variables, hereafter abbreviated *(X; t,T)*, is the likelihood of a single citation times the number of potentially citing (or cited) patents:

$$E[C_{X;t,T}] = (N_{X;t})(A_{X;T})(p_{k,K}) \tag{5}$$

where *C* is the number of citations to the cohort of patents described by the list of attached parameters, *N* is the number of (all) patents in that cohort available to be cited, and *A* is the number of potentially citing (biotechnology) patents granted in year *T*. This equation can now be rewritten to use what we know about patents in the cohort to which the pairing *(k,K)* belongs:

$$p_{k,K} = p_X = \left(\frac{C_{X;t,T}}{(N_{X;t})(A_{X;T})} \right) \tag{6}$$

and combining **(2)**, **(3)**, **(4)** and **(6)** gives us an equation:

$$\left(\frac{C_{X;t,T}}{(N_{X;t})(A_{X;T})} \right) = \alpha_{SV}\alpha_{SE}\alpha_{SA}\alpha_{SI}\alpha_{LAG}\alpha_T\delta^D$$
$$\exp[-\beta_1(\sum_{s=t_A}^{T} P_s)][1-\exp(-\beta_2(T-t))] + \varepsilon_X \tag{7}$$

which can be estimated by non-linear least squares as long as the error term, ε_X , is well behaved.[3]

As a comparison group, we omit from our estimation the coefficients reflecting simultaneous citations (where T=t), short cited patent lags, citations within the same state (distance = 0), and the last time dummy ($T_{1990\text{-}94}$).

Using this methodology now standard in the citation literature, rather than estimating coefficients based upon ten billion possible patent-to-patent citations (which actually exist and were individually tabulated), the data are summarized into 5908 group-to-group observations of citations, each weighted by the number of citations they represent. These observational "cohorts" are groups defined to be mutually exclusive and exhaustive citation frequencies between patent categories of interest, including observations of citations between similar patents (i.e. patents which cite patents similar in time, in region, in technology, in assignee, etc.) and observations of citations between heterogeneous patents. Estimated coefficients hail from the power that group characteristics display in explaining frequent citations between certain groups, compared to infrequent citations between others. For example, if we were to find that citations in cohorts containing patents physically close to each other were relatively more frequent than citations in cohorts containing patents far from each other, the regression coefficient on distance should be a negative one, as distance detracts from the propensity to cite another patent.

In particular, since we have included time as a component of our cohort

---

[3] Because the data are grouped we weight each observation by $\sqrt{(N_{X;t})(A_{X;T})}$ to avoid heteroskedasticity issues (Greene, 1993).

definition, we can compare the different impact that distance (or any other variable) has wielded over time. We are in fact statistically comparing cohorts which are identical in every way except for different timeframes, with the difference in citation frequency between those cohorts attributable to the different timeframe. It is equivalent to including a time dummy variable in a regression equation of individual observations, except that in this case each observation is a citation frequency within a cohort.

A summary table of the salient data characteristics is presented in Table 1. Notice that there is an average 0.001 probability of citation between patents (the dependent variable), with a slight downward trend over time. Although each patent cites more than prior patents did, there is an even more rapidly increasing population of patents in existence, making the probability fall over time.

Measures of distance have remained remarkably constant over time, as has the share of patent citation cohorts displaying self-citation (by the same inventor, assignee, examiner or geographic state). Therefore, any estimated effects seen in our results are not a function of biases in the underlying variables. Instead, those effects come from the distribution of citation patterns overlaying the patent population.

Table 2 presents regression estimates based on the cohorts described, incorporating all ten billion possible patent-to-patent citations. Notice that significance is measured with a test of the null hypothesis that a given coefficient is unity.

The results are broadly consistent with the literature (e.g. Johnson and Popp, 2003). For example, our regression yields time-specific constants that have diminished since 1980. Although patents per year increase, and each one has a longer list of references than previous generations did, there is a decreasing probability of citation to any particular patent. Both decay and diffusion are significant and strikingly faster than evidenced elsewhere in the literature. Our results show rates averaged that are much faster to decay and diffuse than average rates across the economy, reflecting the fact that biotechnology moves at a fast pace in its creation and decay of knowledge.

There are positive and statistically significant self-citation effects by inventors and assignees. The combination clearly suggests a strong impact of applicant self-citation, but their relative size implies that inventors have a stronger propensity to cite other laboratory co-workers and not necessarily themselves, which is probably a better signal of useful knowledge acquisition than the self-promotion effect we would suspect in the case of same-inventor citations.

Furthermore, the same-examiner effect, which we introduce for the first time in analysis of this type, is strong, positive and significant, reflecting the correlation between subject material and a particular examiner. Since applicants presumably would not be able to predict the examiner in advance, and would have little reason to cite patents by the examiner even if they knew, we can perhaps conclude that examiners apparently add those citations, adding disproportionately more citations to patents that they have read themselves.

Finally, we obtain the counterintuitive result that increased physical distance makes citations more likely (δ>1), a statistically significant result. The reason is explained in the remaining columns of Table 2, where we report estimates of equation (7) separately for each time period. The ancillary regressions omit the time-specific constant for each period, and calibrate the decay and diffusion rates to the primary estimated values.

Results are similar to the primary analysis, with one key distinction. The distance coefficient δ transitions from less than unity in 1975-79, to insignificantly different than unity in 1980-84, to significantly greater than unity in 1985-89 and 1990-94. In other words, the 1970's saw physical distance as a limiting factor to citations, with an additional hundred kilometers of distance dropping the probability of citation by five percent. In the later periods, distant citations became the norm, with longer citations more common than shorter ones.

The results point unquestionably to the fact that physical distance has become less of a constraint with the passage of time. Perhaps the trend is due to the nature of the knowledge being created, but we suspect that it is more due to advances in communication, which allows easier transmission of information across great distances in the era of computerization, internet, teleconferencing and cellular communication. In short, the principles underlying the inter-firm transfer of knowledge are changing in a striking fashion, making spillovers easier and longer than ever before.

V. Conclusion and Policy Implications

We are left with a striking picture of the inter-firm transfer of biotechnological knowledge. Controlling for other factors, knowledge flows used to diminish with physical distance, but the importance of distance has been receding with time. That is, knowledge is more likely to transfer over long distances now than it was twenty years ago. Paradoxically, it may now be more likely to jump long distances (e.g. from coast to coast), than it is to travel the shorter distances to neighboring states.

The policy recommendations of this paper have therefore been heard before. In an age of more intense and distance-free communication, the conduits of knowledge transmission take on a new importance. Researchers and firms have obviously benefited tremendously from the movement to electronic patent searches and filings. In fact, that trend may have partially driven our results.

Long-distance knowledge transfers are increasingly the norm in biotechnology. The policy implications of this paper may be important not only for regions of the U.S. but for less developed nations as well. As the importance of physical distance has diminished over time, innovation has become possible at a wider array of locations, potentially drawing on a wider range of raw materials (such as agricultural germplasm) and ideas. This might imply a possibility for the deliberate fostering of non-traditional locations for biotechnology, with a prerequisite of vibrant communication with the research community elsewhere.

VI. References

Caballero, R.J. and A.B. Jaffe (1993), "How High are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth," in Olivier J. Blanchard and Stanley Fischer, eds., NBER Macroeconomics Annual 1993, MIT Press, Cambridge, MA.

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

Feldman, M.P. (1994) The Geography of Innovation. Kluwer Academic Publishers, Dordrecht, Netherlands.

Feldman, M.P. (1999) "The New Economics of Innovation, Spillovers and Agglomeration: A Review of Empirical Studies". Economics of Innovation and New Technology. Vol 8, p.5-25.

Gelsing, L. (1992) "Innovation and Development of Industrial Networks". National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning. B. Lundvall, editor. Pinter Publishers, London.

Greene, W.H. (1993) Econometric Analysis, Macmillan Publishing Company, New York.

Hall, B.H., A.B. Jaffe and M.Trajtenberg (2001) "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools". NBER Working Paper #8498.

Henderson, J.V. (1986) "Efficiency of Resource Usage and City Size". Journal of Urban Economics. Vol 19, p.47-70.

Jaffe, A.B., M.Trajtenberg and R.Henderson (1993) "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations". Quarterly Journal of Economics. Vol 108, p.577-598.

Jaffe, A.B. and M.Trajtenberg (1996) "Flows of Knowledge From Universities and Federal Labs: Modeling the Flow of Patent Citations Over Time and Across Institutional and Geographic Boundaries", NBER Working Paper #5712.

Jaffe, A.B., M.Trajtenberg and M.S.Fogarty (2000) “Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors”. American Economic Review. Vol 90, p.215-218.

Johnson, D.K.N. and A.Brown (2004) "How the West Has Won: Regional and Industrial Inversion in U.S. Patent Activity”. Economic Geography, 80(3): 241-260.

Johnson, D.K.N. and D.Popp (2003) "Forced Out of the Closet: The Impact of the American Inventors Protection Act on the Timing of Patent Disclosure". Rand Journal of Economics, 34(1): 96-112.

Johnson, D.K.N. and R.E.Evenson (2000) "How Far Away Is Africa? Technological Spillovers to Agriculture and Productivity". American Journal of Agricultural Economics. Vol 82, p.743-749.

Johnson, D.K.N. and V.Santaniello (2000) "Biotechnology Inventions: What Can We Learn from Patents?". Agriculture and Intellectual Property Rights: Economic, Institutional, and Implementation Issues in Biotechnology. V. Santaniello et al., editors. CABI Publishing, New York.

Knoke, D. and J.H.Kuklinski (1982) Network Analysis. Sage Publications, London.

Lundvall, B.A. (1992) "User-Producer Relationships, National Systems of Innovation and Internationalisation". National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning. B. Lundvall, editor. Pinter Publishers, London.

Maurseth, P.B. and B.Verspagen (1999) "Europe: one or several systems of innovation? An analysis based on patent citations". Economic Challenge to Europe: Adapting to Innovation-Based Growth. Edward Elgar Publishing, London.

MetLife.com (2001) Distance Calculator. Accessed August 2001.

Sjoholm, F. (1996) "International Transfer of Knowledge: The Role of International Trade and Geographic Proximity". Weltwirtschaftliches Archiv. Vol 132, p. 97-115.

Smith, D. and R.Florida (1994) "Agglomeration and Industry Location: An Econometric Analysis of Japanese-Affiliated Manufacturing Establishments in Automotive Related Industries". Journal of Urban Economics. Vol 36, p. 23-41.

USPTO Technology Assessment and Forecast Program (1998) "Technology Profile Report: Patent Examining Technology Center Groups 1630-1650, Biotechnology". United States Patent and Trademark Office report.

Valente, T.W. (1995) Network Models of the Diffusion of Innovations. Hampton Press, Inc. Cresskill, New Jersey.

Von Hippel, E. (1994) "Sticky Information and the Locus of Problem-Solving". Management Science. Vol 40, p. 429-439.

Table 1: Summary of citation cohort data

Variable	*Mean*	*St. Dev*	*Min*	*Max*
Actual citations as share of potential citations	0.001	0.017	0	1
1975-1979	0.002	0.036	0	1
1980-1984	0.002	0.022	0	1
1985-1989	0.001	0.009	0	1
1990-1994	0.001	0.015	0	1
	Freq of 0	*Freq of 1*	*Min*	*Max*
Dummy variables:				
Same inventor			0	1
1975-1979	1735	556	0	1
1980-1984	1267	404	0	1
1985-1989	1054	270	0	1
1990-1994	486	136	0	1
Same assignee			0	1
1975-1979	1688	603	0	1
1980-1984	1238	433	0	1
1985-1989	1019	305	0	1
1990-1994	493	129	0	1
Same examiner			0	1
1975-1979	1464	827	0	1
1980-1984	1057	614	0	1
1985-1989	901	423	0	1
1990-1994	450	172	0	1

There are 5908 observations in the full dataset, due to our partitioning of all possible citations into citation cohorts. Of those, 2291 have a citing patent granted in 1975-79, 1671 are in 1980-84, 1324 are in 1985-89, and 622 are in 1990-94.

This distance variable represents observations in twenty ranges, each encompassing 100 km spans, as described in the text. The number of observations per range, in increasing distance order, is 256, 445, 300, 285, 266, 264, 267, 281, 264, 261, 258, 262, 250, 237, 247, 247, 254, 270, 234, 232, 245, 283. The distribution is very similar for each set of citing years.

Table 2: Regression results

Variable	1975-94		1975-79		1980-84		1985-89		1990-94	
Same inventor	0.20	(24.48)**	0.44	(21.65)**	4.71 x 10^{-3}	(26.47)**	1.02 x 10^{-2}	(6.68)**	6.16 x 10^{-8}	(7.45)**
Same assignee	3.10	(17.28)**	82.96	(35.06)**	1.05	(0.10)	0.11	(82.28)**	1.54	(0.82)
Same examiner	5.32	(8.84)**	0.07	(23.62)**	57.86	(39.03)**	21.78	(44.37)**	5.74	(11.75)**
Same technology	2.28 x 10^{-7}	(138.67)**	6.02 x 10^{-9}	(50.59)**	4.95 x 10^{-9}	(41.62)**	-9.67 x 10^{-5}	(71.74)**	-6.07 x 10^{-9}	(40.07)**
Pendency lags										
Medium	0.45	(2.79)**	0.34	(70.66)**	0.66	(19.55)**	0.26	(46.41)**	0.32	(59.59)**
Long	6.99 x 10^{-2}	(190.20)**	0.04	(16.99)**	0.06	(19.54)**	0.06	(59.36)**	---	---
Very Long	1.41 x 10^{-2}	(497.04)**	0.01	(11.03)**	6.30 x 10^{-3}	(65.36)**	---	---	---	---
Constant										
1975-1979	1.28	(1.58)	---	---	---	---	---	---	---	---
1980-1984	3.38	(2.13)*	---	---	---	---	---	---	---	---
1985-1989	1.62	(2.23)*	---	---	---	---	---	---	---	---
Distance	1.01	(173.87)**	0.95	(13.82)**	1.01	(0.45)	1.05	(25.75)**	1.05	(7.78)**
Decay rate^	2.34 x 10^{-2}	(14.32)**	2.34 x 10^{-2}	---	2.34 x 10^{-2}	---	2.34 x 10^{-2}	---	2.34 x 10^{-2}	---
Diffusion rate^	7.98 x 10^{-3}	(2.99)**	7.98 x 10^{-3}	---	7.98 x 10^{-3}	---	7.98 x 10^{-3}	---	7.98 x 10^{-3}	---
Adjusted R^2	0.13		0.04		0.15		0.24		0.16	
Observations	5908		2291		1671		1324		622	

Notes: ^ Decay and diffusion rates are estimated for the first column and calibrated for the last four columns. They report a standard t-test of β=0, unlike other coefficients which report a test of α, δ=1. ** indicates 99% confidence, * 95% confidence.