

THE HEALING ARTS: ANALYSIS OF THE KNOWLEDGE FLOWS BETWEEN SUCCESSIVE GENERATIONS OF INNOVATION IN THE PHARMACEUTICAL SECTOR

Daniel K.N. Johnson

Colorado College
Chair of the Economics & Business Department
djohnson@ColoradoCollege.edu

Kristina M. Lybecker

Colorado College
Schlessman Professor of Economics
kristina.lybecker@ColoradoCollege.edu

Jeffrey Moore

Colorado College
Technical and Statistical Coordinator
jeffrey.moore@ColoradoCollege.edu

—Abstract—

This paper examines the location of innovations within pharmaceutical technology, using U.S. patent citation data to trace the knowledge flows over time. It is clear that knowledge clustering is certainly present. Our study utilizes multivariate left-censored Tobit regression analysis to control for identifiable factors, to examine whether over time the distance between successive innovators has changed. We find the distance to be increasing significantly over time, both when considering all citations and only inter-city transfers.

Key Words: *pharmaceutical, health, patent, innovation, distance, geography, citation*

JEL Classification: O3, R1, L65

1. INTRODUCTION

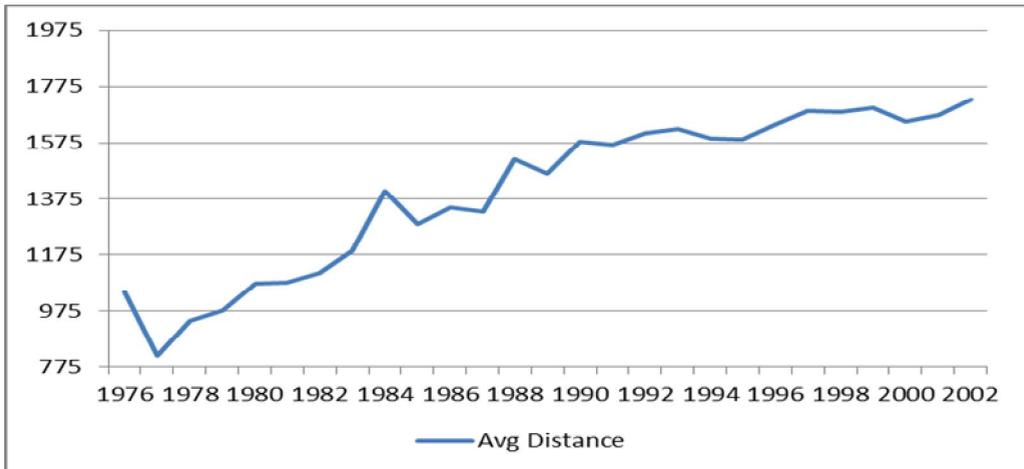
It is widely acknowledged that localization benefits frequently lead firms within an industry to cluster geographically. In the local industry, this reduces the cost of inputs to firms (Henderson, 1986; Smith and Florida, 1994), due to the rapid speed of knowledge diffusion (Caballero and Jaffe, 1993) or due to tacit learning advantages (Von Hippel, 1994). We believe this is the first study to test the

importance of these clustering forces within pharmaceutical technologies, and to trace its impact across time.

Using all pharmaceutical technology patents granted in the U.S. between 1976 and 2002, we statistically test whether there has been a trend to cite knowledge arriving from greater distances. Importantly, we analyze whether such a pattern could stem from (or be abated by) a tendency to cite other patents listing the same inventor, the same firm assignee, or the same technology class. We conclude that the geographic distance between citations has increased significantly over time, though at a decreasing rate.

As Figure 1 shows, the average distance between a citing patent and its bibliographic references has fluctuated drastically over time, with a steady increase following an initial drop in the late 1970s. The multivariate regression presented in the following section controls for other changing factors, but the same fundamental pattern remains.

Figure 1: Average citation distance in kilometers



In section 2 of the paper, we very briefly review the existing literature on technology clustering and the geographic nature of knowledge spillovers. Section 3 presents our data set, and Section 4 presents multivariate regression results, controlling for non-geographic effects in presenting the role of distance. Section 5 identifies implications for policy and further research.

2. LITERATURE REVIEW

The majority of the existing technical and economics literature finds that knowledge diffuses more readily across shorter distances. The underlying

supposition is that inventors rely more heavily on local innovations because they are more aware of (or find more use for) inventions located proximate to them. The importance of location in the spillover of knowledge from one member of an innovation network to another is confirmed by empirical evidence (Gelsing, 1992), but some research indicates that the role may differ by technology (Lundvall, 1992) with location more vital to technologies undergoing radical innovation. During technological revolutions, such as the pharmaceutical sector experienced in the period under study, we might reasonably expect some large geographic impacts on knowledge diffusion.

Geographic proximity has effectively explained R&D-intensive activities (Dorfman, 1988), but firm location may not be a good indicator of the location of innovation (Feldman, 1992; Johnson and Brown, 2004). Earlier work has documented that patents are more likely to cite proximate patents than patents by parties that are located at greater distance (Jaffe et al., 1993; Sjöholm, 1996; Maurseth and Verspagen, 1999; Johnson and Lybecker, 2012; Moore et al., 2012), an effect which is very pronounced in nuclear technology, electronics, and optics (Jaffe and Trajtenberg, 1996). However, existing empirical studies do not examine how that importance has changed over time. Importantly, some studies suggest that distance has never mattered much (Thompson and Fox-Kean, 2005), or that the change over time has been minimal, even with revolutionary changes in information and telecommunication technology (Vaileiadou and Vliegthart, 2009; Graham, 2001).

3. DATA

The application for every patent is required to include citations to any other patents critical to its creation, or which limit its legal reach. Innovators develop this citation list to establish the novelty of the patentable product or process, and the result is a traceable record of knowledge creation. Realistically, patent records do not measure innovation perfectly, as some inventions remain unpatented and patents differ greatly in importance. Nevertheless, patents are highly correlated with the location of other measures of inventive activity (Feldman, 1994). Although citations are an imperfect measure of the transfer of knowledge, they may be included for a variety of reasons. Recent evidence indicates that half of all citations trace true knowledge transfer (Graham, 2001), and if the noisiness of this signal is constant over time, it may be utilized for comparisons across time even with an implied degree of imprecision.

We follow the World Intellectual Property Organization's definition for pharmaceutical technology (Jaffe et al., 2000), and our dataset therefore includes

all patents granted between 1976 and 2002 that qualify as pharmaceutical technology, appended with all patents cited by those patents, at least those that were themselves granted between 1976 and 2002. Citing and cited patents from all non-U.S. inventors have been excluded for feasibility reasons. However, the literature indicates that international citations are growing in frequency across a large set of technologies (WIPO, 2012).

Non-geographic reasons may explain patent citations clustering as well, generating a pattern that appears geographic. Specifically, innovators (or the assignees firms which retain patent rights) may have more familiarity with their own patents, and therefore cite them with greater frequency, a pattern which would give a biased impression of the importance of geography. In order to address this, we include self-citations in the analysis but specifically identify and control for them separately.

Utilizing U.S. patent data from numerous sources (NBER website as described in Hall et al., 2001, in addition to raw data from the independent firm MicroPatent), each patent citation's endpoints (citing patent and cited patent) were geo-coded for the primary location of each listed U.S.-based inventor. We identify locations at the geographic center of the relevant city due to the fact that specific street addresses are available for less than ten percent of all patent documents. The result is a dataset of 87,585 citations from U.S.-based pharmaceutical technology patent documents to other U.S.-based patent documents. The existing literature (Johnson and Lybecker, 2012; Moore et al., 2012) indicates that each of the following factors may play some role in the distance of a citation, so our analysis measured each for every observed citation between citing patent K and cited patent k:

- whether they have the same inventor (hereafter, SI);
- whether they have the same assignee (SA);
- whether they are in the same technology (ST);
- how similar the citing and cited states are in technology types (SC);
- whether the cited patent is also classified as pharmaceutical technology (P);
- whether the assignee is a government agency (G);
- whether the assignee is an educational institution (U);

- how old the citation is, in years between citing and cited patent (A), along with its squared term (A^2) to account for the potentially nonlinear effects of age;
- and a time trend variable to proxy for the year of citation (T), along with its squared term (T^2).

To start, we traced all self-citations, allowing for some flexibility in spellings of the names (since the USPTO, United States Patent and Trademark Office, does not standardize name format). These include both the first inventors, as well as all additional inventors identified by each patent. Self-citation by inventors accounted for approximately ten percent of all citations. While some self-citation is present, this suggests that very strong inter-inventor knowledge spillovers are also present. Alternatively, self-citation by assignees was more frequent, at about fifteen percent of all citations in the dataset, a smaller percentage than found in other sectors like biotechnology (Johnson and Lybecker, 2012) and traditional energy (Moore et al., 2012), indicating that knowledge transfers between individuals or firms are less frequent in pharmaceutical technology. In contrast to academic citations, there is very little reason here to self-cite as a means of advertising, so it is reasonable to assume that self-citations are genuine indicators of useful knowledge or legal protection. Self-citation was coded as a binary variable (SI) for each citation within the dataset.

Beyond location, it is also possible that patents closer in technological content may cite each other with more frequency. Our data are coded such that a binary variable, ST, indicates whether the International Patent Classification (IPC) system classifies both citing and cited patents in the same technology class at the 4-digit level. In global use since 1975, the IPC system (Jaffe et al., 2000) is the standard by which all patents are categorized (and accordingly assigned to examiners for processing, or searched by inventors and lawyers to establish claims). The system consists of 634 clusters at the 4-digit level, so identification of the patents that share a class is a powerful signal of technological similarity, and a significant indicator that they were both processed by patent examiners with very similar scientific training. In our data, approximately twenty-nine percent of all citations saw citing and cited patents sharing a technology class.

The technological correlation between citing and cited states (SC), is also employed for like reasons. Calculating the share of patent activity within each of the 634 IPC technology classes provided each state's technological profile. Pair-wise correlations between state vectors then measure the extent of technological commonality between locations. Controlling for this technological similarity

across locations reduces the likelihood that the data will reveal an importance of geography that may superficially be the result of two regions sharing the same technological portfolio and therefore attracting citation flows. Our analysis reveals an average correlation of 0.87 between cited and citing state technology profiles.

The regressions also include an indicator of whether the cited patent is classified as pharmaceutical technology (P). Obviously, all citing patents have been defined as pharmaceutical technology. Accordingly, there should be a higher probability for them to cite other pharmaceutical technology patents than to cite a random other technology group.

Given that government (G) and university (U) patents may employ distinct conventions for knowledge citations, relative to private sector patents, we include those indicators as controls as well, but only two percent fall in the government category while twenty-five percent under university for all patents in the dataset.

We include linear and squared age terms in order to capture the potential nonlinear effects for older knowledge. The average citation is just roughly 7.3 years from cited to citing document.

Finally, it is essential to incorporate a time trend (and its square, to permit nonlinearities) or to include indicator variables for each time period since the goal of the analysis is to test whether distance changes over time. We utilize a time trend and its square in this analysis, but test the robustness of the results by using year-specific indicator variables as well.

4. STATISTICAL ANALYSIS

Our regression analysis follows the literature, utilizing multivariate left-censored Tobit regression analysis (Petersen and Rajan, 2002) with the citation as the unit of analysis. The distance between a cited patent k granted in year t and a subsequent citing patent K granted in year T , is modeled as a function of the attributes of the two patents, k and K :

$$\delta_{k,K} = \alpha(k, K) + \varepsilon$$

where $\delta_{k,K}$ represents the distance between patents k and K , $\alpha(k,K)$ is a vector of the non-geographic elements of patents k and K that may impact the probability of citation, and ε is a randomly distributed error term. Given that the fit of the equation is improved due to the loglinear nature of the data's underlying relationship, we propose a reduced functional form, using the log of distance (or

technically the log of [distance plus one] in order to avoid taking the log of a zero distance):

$$\begin{aligned}\delta_{k,K} = & \alpha_0 + \alpha_{SA}SA + \alpha_{SI}SI + \alpha_{ST}ST + \alpha_{SC}SC + \alpha_P P \\ & + \alpha_G G + \alpha_U U + \alpha_A A + \alpha_{A2} A^2 \\ & + \alpha_T T + \alpha_{T2} T^2 + \varepsilon_K\end{aligned}$$

where the distance δ of each observed citation is explained by the characteristics of the citing and cited patents as described above. It is important to note that we use a fixed effect specific to the citing patent (ε_K), since there are likely immeasurable characteristics specific to the citing patent which might dictate a longer or shorter average citation distance.

The estimates of the multivariate Tobit regression (left-censored for intra-city citations with a distance of 0 miles) are presented in Table 1, with White-corrected errors to accommodate the presence of heteroskedasticity in the sample, using fixed effects at the level of the citing patent in which each individual citations is the unit of analysis. For simplicity, we utilize only a time trend (and its square) to measure the change due to the passage of time, after controlling for other factors. The trend coefficient is significant across all citations and also across the set of citations with distances greater than 100 kilometers. This indicates that the average distance is increasing with time, though at a decreasing rate.

An examination of the coefficients in Table 1 reveals several variables to be significant, both across the full set of citations as well as for the set of citations with distances greater than 100 kilometers. It comes as no surprise that citations with the same assignee or same inventor are more likely to be proximate than are other citations (the coefficients on SA and SI are both negative). In addition, we see that citations from states with similar technology portfolios tend to be closer together, a fact captured by the negative coefficient on that variable (SC).

Table 1: Tobit Weighted Regressions on log(distance+1) with time trend

Variable	All citations		Only citations with distance>100km	
	Coefficient	t-statistic	Coefficient	t-statistic
SA	-2.145	-49.66	-0.227	-8.14
SI	-2.107	-36.69	-0.168	-3.95
ST	-0.034	-1.3	-0.051	-3.63
SC	-8.208	-74.93	-2.999	-55.54
P	0.035	1.45	0.106	8.22
G	0.210	3.98	-0.070	-2.39
U	0.118	5.45	0.007	0.6
A	0.023	3.45	-0.005	-1.52
A ²	-8.70E-04	-2.82	-3.56E-05	-0.21
T	0.039	3.28	0.048	6.93
T ²	-0.001	-2.92	-0.001	-5.23
Constant	13.197	94.16	9.244	117.54
F-stat	2039.19		348.48	
Observations	87585		67769	

Citations that cite other pharmaceutical patents average a slightly longer distance than their peers, but only in the case of citations with distances greater than 100 kilometers. Apparently distance matters less for the transfer of purely pharmaceutical-related knowledge than it matters for the transfer of non-pharmaceutical innovations into the pharmaceutical sector.

Across the entire universe of citations, government-assigned knowledge tends to travel longer transmission distances for the knowledge they cite, a result that is reversed when we consider only long-distance (>100 km) citations. Academic patents tend to be longer than their peers as well, but that disappears among long-distance citations (>100 km). Finally, age matters; older citations travel longer distances, an effect which other studies (Johnson and Popp, 2003) have confirmed. However this effect also disappears for long-distance citations.

To permit maximum flexibility to these nonlinearities, and potential nuances in particular years, an identical analysis was conducted using separate year indicator variables. These results are available from the authors. Again, self citation, by both inventors (SI) and assignees (SA), and technological correlation between

citing and cited states (SC) are significant. Moreover, many of the year dummies are now significant.

Turning our attention to the significant elements of the regression results, we notice that patents in states that have similar technology sets in their innovative profiles tend to be close together, a fact captured by the negative coefficient on that variable. Unsurprisingly, citations with the same inventor are more likely to be proximate than are other citations. This suggests that within pharmaceutical technology, inventors are not likely to move locations between self-citations.

In addition, citations to the same assignee usually reference citing and cited patents that are closer to each other than patents with distinct assignees. Indicating that pharmaceutical firms may have well-developed knowledge transfer between branches or between the main office and their local innovators.

The age of the cited patent matters when examining the full set of citations. Older citations travel longer distances, presumably because it takes time for knowledge to travel, an effect confirmed by other studies (Johnson and Popp, 2003; Johnson and Lybecker, 2012; Moore et al., 2012) for a variety of technologies. Interestingly, again this effect is not significant among long-distance citations (>100km).

5. CONCLUSIONS

Admittedly the limited scope of this study prevents wide-reaching conclusions about the nature of technological change in pharmaceutical technology. Nevertheless, several themes appear relatively obvious and robust to alternative interpretations of the data.

First, citation distances appear to be significantly increasing over time, whether we model those distances simply as a function of time or as a more complicated function of the attributes of the underlying patents. Thus, with the exception of the late 1970s, it appears that knowledge flows between pharmaceutical-innovators have been increasing over time, clearly indicating the diminished importance of distance over time, perhaps due to recent changes in information and communications technology.

Second, other characteristics may contribute to the explanation of why one patent cites another. Self-citation is not frequent, but this analysis demonstrates that it has a strong effect on patent citations. In like manner, technological similarities across states appear to correlate with more proximate citations.

What conclusions may we take from the study of pharmaceutical patent citations? Insofar as they describe the paths of knowledge spillovers, then we can identify the patterns and key actors in pharmaceutical technology. Over time it appears that distance has come to matter less and less in the diffusion of pharmaceutical innovation. Long-distance knowledge transfers are increasingly common in pharmaceutical technology. Knowledge flows are possible across a wider array of locations, potentially drawing on a wider range of raw materials and ideas. This suggests a potential for the deliberate fostering of non-traditional locations for pharmaceutical technology, with a prerequisite of vibrant communication with the research community elsewhere.

BIBLIOGRAPHY

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.

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.

Graham, A. (2001) "The Assessment: Economics of the Internet". Oxford Review of Economic Policy 17(2): 145-158.

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, 19: 47-70.

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 R.Henderson (1993) "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations". *Quarterly Journal of Economics*, 108: 577-598.

Jaffe, A.B., M.Trajtenberg and M.S.Fogarty (2000) "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors". *American Economic Review* 90: 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 K.M. Lybecker (2012) "Does distance matter less now? The changing role of geography in biotechnological innovation". *Review of Industrial Organization* 40(1): 21-35.

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.

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.

Moore, J.A., D.K.N. Johnson and K.M. Lybecker (2012), "Sure, But Who Has the Energy? The Importance of Location for Knowledge Transfer in the Energy Sector". *World Renewable Energy Forum Conference Proceedings*, 2012.

Petersen, M.A. and R.G. Rajan (2002) "Does Distance Still Matter? The Information Revolution in Small Business Lending". *The Journal of Finance* 57(6): 2533-2570.

Sjoholm, F. (1996) "International Transfer of Knowledge: The Role of International Trade and Geographic Proximity". *Weltwirtschaftliches Archiv*, 132: 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*, 36: 23-41.

Thompson, P. and M. Fox-Kean (2005), "Patent Citations and the Geography of Knowledge Spillovers: A reassessment". *American Economic Review* 95: 450-460.

Vasileiadou, E. and R. Vliegthart (2009) "Research Productivity in the Era of the Internet Revisited". <http://ssrn.com/abstract=1431457>.

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

World Intellectual Property Organization (2012) "IPC Green Inventory". Website describing definitions <http://www.wipo.int/classifications/ipc/en/est/>, accessed March 16, 2012.