WORKING PAPER

Ain't No Sunshine When You're Gone: Analysis of the Knowledge Flows Between Successive Generations of Solar Innovations

by
Daniel K.N Johnson, Jeffrey Moore
& Kristina M. Lybecker

Colorado College Working Paper 2012-09
June, 2012

Department of Economics and Business
Colorado College
Colorado Springs, Colorado 80903-3298
www.coloradocollege.edu/dept/EC

AIN'T NO SUNSHINE WHEN YOU'RE GONE: ANALYSIS OF THE KNOWLEDGE FLOWS BETWEEN SUCCESSIVE GENERATIONS OF SOLAR INNOVATIONS

Daniel K.N Johnson
Colorado College
Dept of Economics
& Business
14 East Cache La Poudre
Colorado Springs, 80903
djohnson@
ColoradoCollege.edu

Jeffrey Moore
Colorado College
Dept of Economics
& Business
14 East Cache La Poudre
Colorado Springs, 80903
Jeffrey.Moore@
ColoradoCollege.edu

Kristina M. Lybecker
Colorado College
Dept of Economics
& Business
14 East Cache La Poudre
Colorado Springs, 80903
Kristina.Lybecker@
ColoradoCollege.edu

ABSTRACT

This paper examines the location of innovations within solar technology, using U.S. patent citation data to trace their diffusion over time. Knowledge clustering is clearly present. We employ multivariate left-censored Tobit regression analysis to control for identifiable factors, to examine whether the distance between successive innovators has changed over time. We find the distance to be increasing slightly over time, both when considering all citations and only inter-city transfers.

1. INTRODUCTION

It is well recognized that localization benefits frequently lead firms within an industry often cluster geographically. This reduces the cost of inputs to firms in the local industry [1,2], due to the rapid speed of knowledge diffusion [3] or due to tacit learning advantages [4]. We are unaware of any study which has tested the importance of these clustering forces within solar technologies, nor traced its impact across time.

Using all solar 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. Moreover, we examine whether such a pattern could arise 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 slightly 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 very significant jump in the late 1970s. The multivariate regression which presented in the following section controls for other changing factors, but the same fundamental pattern remains.

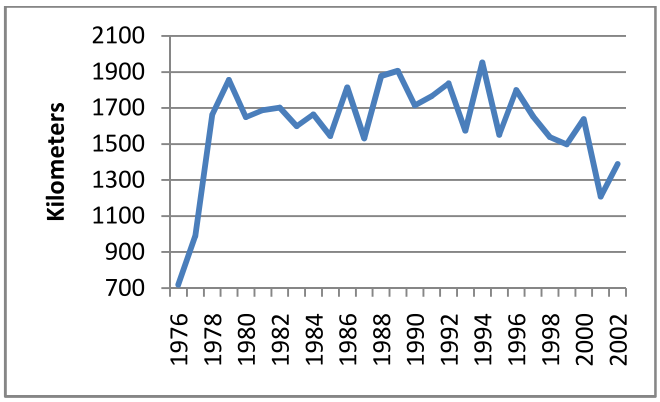

Fig 1. Average citation distance in kilometers

In section 2 of the paper, we concisely review the current literature on technology clustering and the geographic nature of knowledge spillovers. Section 3 presents our data set, designed for compatibility with the recent literature, and Section 4 presents multivariate regression analysis that controls for non-geographic effects in presenting the role of distance. Section 5 then concludes, identifying implications for policy and further research.

2. LITERATURE REVIEW

Most technical and economics literature indicates that knowledge diffuses more readily across short distances. The underlying supposition is that inventors build more heavily on local inventions because they more aware of (or find more use for) inventions located close to them.

Empirical evidence confirms the importance of location in the spillover of knowledge from one member of an innovation network to another [5], but some research points out that the role may differ by technology [6] with location more critical for technologies undergoing radical innovation. During technological revolutions, such as solar technology experienced in the period under study, we might reasonably expect some large geographic impacts on knowledge diffusion.

R&D-intensive activities have been effectively explained using geographic proximity [7], but firm location may not be a good indicator of the location of innovation [8,9]. It has been firmly established that patents are more likely to cite proximate patents than patents by parties that are located farther away [10-14], an effect which is very pronounced in electronics, optics, and nuclear technology [15]. However, existing studies do not examine how that importance has changed over time.

However, there are studies suggesting that distance has never mattered much [16], or that there has been minimal change over time, even with revolutionary changes in information and telecommunication technology [17,18].

3. DATA

Every patent application must include citations to any other patents critical to its creation, or which limit its legal reach. Inventors develop this citation list to prove the novelty of the patentable product or process, and the result is a traceable record of knowledge creation. Of course, patent records do not measure innovation perfectly, as some inventions remain unpatented and patents differ greatly in importance. However, patents are highly correlated with the location of other measures of innovative activity [8]. While citations do not perfectly reflect the transfer of knowledge, as they may be included for a variety of reasons, evidence indicates that half trace true knowledge transfer [18], and if the noisiness of this signal is constant over the years, we can use it to compare across time even with an implied degree of imprecision.

We follow the World Intellectual Property Organization's definition for solar technology [19], and our dataset therefore includes all patents granted between 1976 and 2002 that qualify as solar technology, appended with all patents cited by those patents, at least those that were themselves granted between 1976 and 2002. Due to feasibility issues, citing and cited patents from all non-U.S. inventors have been excluded. However, there is evidence in the literature that international citations are growing in frequency across a large set of technologies [20].

Patent citations may cluster for non-geographic reasons as well, generating a pattern that appears geographic. For example, inventors (or the assignees firms which retain patent rights) may have greater familiarity with their own patents, and therefore cite them frequently, a pattern which would give a biased impression of the importance of geography. Given this, we include self-citations in the analysis but specifically identify and control for them separately.

Using U.S. patent data from a variety of sources (NBER website as described in [21], 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 innovator. Given that specific street addresses are available for less than ten percent of all patent documents, we identified locations at the geographic center of the relevant city.

The result is a dataset of 10,997 citations from U.S.-based solar technology patent documents to other U.S.-based patent documents. The existing literature [13,14] indicates that each of the following characteristics may play some role in the distance of a citation, so this study 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 solar technology (S);
- 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).

First, 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 inventors listed for each patent. Self-citation by inventors accounted for between one and ten percent of all citations, depending on the subsector (with alternative energy technologies accounting for the most self-citations, and most technologies within the range of three to five percent). This suggests that while some self-citation is present, very strong inter-inventor knowledge spillovers are also present. On the other hand, self-citation by assignees was very frequent, at about twenty-five percent of all citations in the dataset, a greater percentage than found in other sectors like biotechnology [13] and traditional energy [14], suggesting that knowledge transfers between individuals or firms are less frequent in solar technology. 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 genuine indicators of useful knowledge or legal protection. Self-citation was coded as a binary variable (SI) for each citation within the dataset.

Regardless of location, it is also possible that patents closer in technological content may cite each other more frequently. The data are coded so that a binary variable, ST, indicates whether the International Patent Classification (IPC) system identifies both citing and cited patents in the same technology class at the 4-digit level. In global use since 1975, this system [19] is the standard by which all patents are categorized (and thus assigned to examiners for processing, or searched by inventors and lawyers to establish claims). There are 634 clusters at the 4-digit level, so identification that the patents share a class is a significant signal of technological similarity, and a powerful indicator that they were both processed by patent examiners with very similar scientific training. Within this dataset, approximately twenty percent of all citations saw citing and cited patents sharing a technology class.

The technological correlation between citing and cited states (SC), is also utilized for a similar reason. Each state's technological profile was calculated as the share of patent activity within each of the 634 IPC technology classes. Pair-wise correlations between state vectors then establish the extent of technological similarity between locations. Controlling for technological similarity across locations will reduce the likelihood of the data to showing an importance of geography that may superficially be the result of two regions sharing the same technological portfolio and hence attracting citation flows. Our calculations reveal an average correlation of 0.87 between cited and citing state technology profiles.

The analysis also includes an indicator of whether the cited patent is classified as solar technology (S). Obviously, all citing patents have been defined as such, and there should be a higher probability for them to cite other solar technology patents than to cite a random other technology group.

Because government (G) and university (U) patents may employ distinct conventions for knowledge citations, than do private sector patents, we include those indicators as controls as well, but only four percent fall in the government category and six percent under university for all patents in the dataset.

In order to capture the potential nonlinear effects for older knowledge, linear and squared age terms are included. The average citation is just roughly eight years from cited to citing document.

Finally, it is necessary to include 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.

4. STATISTICAL ANALYSIS

Our regression analysis follows the literature [13,14] in using multivariate left-censored Tobit regression analysis [22] 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 patents k and K:

$$\delta_{k,K} = \alpha(k,K) + \varepsilon \tag{1}$$

where $\delta_{k,K}$ represents the distance between patents *k* and *K*, $\alpha(k,K)$ is a vector of the non-geographic characteristics of patents *k* and *K* that may impact the probability of citation, and ε is a randomly distributed error term. Because the fit of the equation is better 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) because the fit of the equation is better due to the loglinear nature of the data's underlying relationship:

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

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

Table 1 presents the estimates of the multivariate regression Tobit (left-censored for intra-city citations with a distance of 0 miles), with White-corrected errors to accommodate the presence of heteroskedasticity in the sample, using fixed effects at the level of the citing patent where each individual citations is the unit of analysis. For simplicity, we estimate using only a time trend (and its square) to measure the change due to the passage of time, after controlling for other factors. Here the trend coefficient is insignificant, indicating a lack of evidence that average distance either increases or decreases with time, *ceteris paribus*. These results are confirmed when considering only inter-city citations (or citations with distances greater than 100 kilometers).

An examination of the coefficients in Table 1 reveals very few variables to be significant. Only self citation (SI) and technological correlation between citing and cited states (SC) are significant both among all citations and only inter-city citations.

To permit maximum flexibility to these nonlinearities, and potential nuances in particular years, the same analysis was conducted using separate year indicator variables. These results are presented in Table 2. Again, self citation (SI) and technological correlation between citing and cited states (SC) are significant. However, in the case of inter-city transfers, 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 solar technology, inventors are not likely to move locations between self-citations.

Citations to the same assignee usually reference citing and cited patents that are closer to each other than patents with distinct assignees. Indicating that solar 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 [13, 14, 23] for an array of technologies. Interestingly, this effect is not significant among long-distance citations (>100km).

5. CONCLUSIONS

While the limited scope of this study prevents major conclusions about the nature of technological change in solar technology from this work, several themes appear relatively obvious and robust to alternative interpretations of the data.

First, in stark contrast to numerous other sectors, citation distances appear to be just barely 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 solar-innovators have been relatively unaffected by recent changes in information and communications technology.

Second, other factors may contribute to the explanation of why one patent cites another. Self-citation is not frequent, but apparently has a strong effect on patent citations. Similarly, technological similarities across states appear to correlate with more proximate citations.

Are there larger lessons here to be gleaned from the study of patent citations? Insofar as they describe the paths of knowledge diffusion, then we can identify the patterns and key actors in a technology such as solar technology. Despite the diffusion seen in other sectors, solar technology remains as localized as ever. It is disappointing that information does not appear to be more widely utilized (in a geographic sense) over time, though this may indicate that there is some level of industry stability as firms remain clustered together over the years.

At this point, we can only point to the fact that the transmission of solar technology innovation, unlike other well-documented cases, presents no distinguishable change in its historical pattern. For better or worse, recent changes in telecommunications do not appear to have impacted the ability or willingness of solar-innovators to draw inspiration from more distant locations.

References

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

2. 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

3. 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

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

5. 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

6. 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

7. 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

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

9. 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

10. 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

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

12. 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

13. 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

14.Moore, J.A., D.K.N. Johnson and K.M. Lybecker, "Sure, But Who Has the Energy? The Importance of Location for Knowledge Transfer in the Energy Sector". Colorado College Department of Economics and Business Working Paper 2011-14

15. 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

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

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

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

19. 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

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

20. Johnson, D.K.N. and K.A. Sneed (2009) "Are Many Heads Better Than Two? Recent Changes in International Technological Collaboration". Journal of Law, Ethics and Intellectual Property 3(1): 1-9

21. 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

22. 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

23. Johnson, D.K.N. and D. Popp, '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. Previously NBER Working Paper 8374.

TABLE 1: TOBIT WEIGHTED REGRESSIONS ON LOG(DISTANCE+1), TIME TREND

	All citations		**Only citations with distance>100km**	
Variable	**Coefficient**	**t-statistic**	**Coefficient**	**t-statistic**
SA	-1.24103	-12.59	-0.0811782	-1.68
SI	-3.01409	-14.05	-0.2966353	-2.06
ST	-0.0052	-0.08	0.0074822	0.21
SC	-8.8597	-27.56	-2.2797010	-15.64
S	0.055855	0.66	0.0094207	0.22
G	0.035903	0.3	-0.0539090	-0.79
U	0.089249	0.86	0.0060484	0.09
A	0.064333	2.99	0.0006483	0.06
A^2	-0.00124	-1.29	-0.0001945	-0.39
T	0.030619	1.14	0.0122291	0.82
T^2	-0.00145	-1.82	-0.0002069	-0.47
Constant	13.85614	42.65	9.0990030	52.68
F-stat	156.63		24.37	
Observations	7275		5476	

TABLE 2: TOBIT WEIGHTED REGRESSIONS ON LOG(DISTANCE+1)

	All citations		**Only citations with distance>100 km**	
Variable	**coefficient**	**t-statistic**	**coefficient**	**t-statistic**
SA	-1.253097	-12.8	-0.0919297	-1.92
SI	-2.978057	-13.81	-0.2975982	-2.05
ST	-0.0172369	-0.26	-0.0050387	-0.14
SC	-8.835647	-27.34	-2.250812	-15.36
S	0.0694775	0.82	0.000962	0.02
G	0.039804	0.33	-0.0188992	-0.27
U	0.0982451	0.97	0.0029136	0.04
citingyear77	-0.5085765	-0.59	0.1206566	0.21
citingyear78	0.3011388	0.39	0.9745374	2.03
citingyear79	0.6557474	0.91	0.8633348	1.84
citingyear80	0.3302404	0.46	0.9249412	1.97
citingyear81	0.3992135	0.56	0.7928389	1.71
citingyear82	0.3635106	0.51	0.6742532	1.44
citingyear83	0.3614746	0.5	0.6774156	1.44
citingyear84	0.2749228	0.38	0.7014251	1.49
citingyear85	0.3025387	0.42	0.4360577	0.93
citingyear86	0.5767055	0.82	0.8555372	1.86
citingyear87	0.1234487	0.17	0.5657875	1.22
citingyear88	0.3676643	0.51	0.8383999	1.78
citingyear89	0.4892412	0.7	0.8890818	1.92
citingyear90	0.3762426	0.53	0.6613748	1.43
citingyear91	0.2917013	0.41	0.8578649	1.83
citingyear92	0.5106411	0.71	0.9636189	2.09
citingyear93	0.0623288	0.09	0.6477427	1.4
citingyear94	0.3847774	0.55	0.7671445	1.64
citingyear95	0.1955091	0.28	0.9379707	2.03
citingyear96	0.1590613	0.23	0.874433	1.9
citingyear97	0.2662705	0.38	0.9289408	2.01
citingyear98	-0.0485908	-0.07	0.9533797	2.06
citingyear99	0.312702	0.45	0.8634637	1.87
citingyear00	0.2191028	0.31	0.7970408	1.72
citingyear01	-0.0290474	-0.04	0.7280178	1.57
citingyear02	0.0066041	0.01	0.8237827	1.79
A	0.0623391	2.85	0.001768	0.15
A^2	-0.0011411	-1.17	-0.0002562	-0.51
Constant	13.62037	18.61	8.410842	18.18
F-stat	51.14		9.93	
Observations	7275		5476	