The Impact of Skilled Emigration on Innovation in South Korea

Nathan Obaid

Professor Gary Smith

Economics 190: Senior Seminar

Pomona College

I thank Professor Gary Smith for advising me on this paper. I also thank Professors Pierangelo De Pace and Minju Kim, Jae Won Saw, and Owen Deutsch and the rest of Professor Smith's Senior Seminar for helpful comments, as well as Nyema Guannu, and Professors Fernando Lozano and Bowman Cutter for technical assistance.

Abstract

When skilled workers leave a country to find better employment opportunities abroad, there are competing effects on that country's ability to innovate. On the one hand, such emigration harms domestic innovation because it reduces the number of local skilled workers. On the other hand, such emigration helps domestic innovation to the extent skilled expatriates remit the financial and intellectual gains they have gotten abroad. The crucial question is which effect dominates in practice. Following the methodology of Agrawal, Kapur, and McHale (2008), who examined the case of Indian skilled emigration, I empirically compare these competing effects using data from South Korea. Using patent data of citing-cited pairs from the United States Patent and Trademark Office to proxy for knowledge flows, I find that during the 1980s and '90s skilled emigration generally harmed South Korean innovation. This trend holds by and large when I split my sample by technological category, vintage, and lag time between the application dates of the citing and cited patents. I conclude that it is in South Korea's economic interest to make a greater effort to retain potential skilled emigrants.

The Impact of Skilled Emigration on Innovation in South Korea

I. Introduction

The migration of skilled domestic labor from poor countries has long been a subject of intense and widespread interest. Globally, the magnitude of this phenomenon is considerable, as many poor countries have nearly half of their skilled workforce living abroad (Agrawal et al., hereafter "AKM," 2008). There is no expert consensus, however, on the impact of such emigration on economic development in poor countries. One viewpoint has emphasized the "brain bank" element of skilled migration, arguing that skilled labor is wasted at home and put to more productive use abroad, in better-incentivized and higher-human capital environments that better complement skill. In this scenario global innovation increases, and skilled expatriates remit human and physical capital to their home countries (Kuhn & McAusland, 2006). On the other hand, others have emphasized the "brain drain" element, insisting that it is important for innovators to be located at home in order for domestic development to occur, because: a) knowledge spillovers, the positive externalities associated with idea sharing, may be localized; b) rich-country innovation may not be relevant for poor countries; and c) domestic knowledge production may be a prerequisite for the absorption of foreign innovation (respectively: Keller, 2002; Basu & Weil, 1998; and Cohen & Levinthal, 1989).

Thus the question emerges: which of these competing effects has a greater impact on poor-country development? AKM (2008) have attempted to determine whether brain bank or brain drain matters more for innovation in India. They developed and estimated a model founded on the assumption that the access of domestic innovators to knowledge drives innovation, and they used patent citations as a proxy for knowledge flows. Their results suggest that, in general, brain drain has a more powerful effect than brain bank on innovation in India.

In this paper I apply AKM's (2008) model to the case of South Korea (hereafter "Korea"). I am motivated to replicate their study so that I can compare my results with theirs. This analysis is interesting, as India and Korea have quite different economies – for example, Korea's GDP per capita is an order of magnitude greater than India's (The World Bank, 2011).

The rest of this paper proceeds as follows. Section II reviews the relevant literature, Section III explains my empirical strategy, Section IV describes the United States Patent and Trademark Office data that I use, Section V shows and summarizes my results, Section VI discusses my results, and Section VII concludes.

II. Literature Review

In their useful overview of the brain drain situation, Commander, Kangasniemi, and Winters (2002) noted several important patterns. For one, a country's migration rate of tertiary-educated labor is inversely correlated with both its population and its GDP per capita. Additionally, since about 1990, skilled migration has accelerated, and in fact skilled migration has increased faster than non-skilled migration.

On the theoretical side, models of a beneficial brain drain have begun to gain prominence in the past decade or so. These models all start with wages for a given skill set being higher abroad than at home (Commander et al., 2002). Then most assume that there is uncertainty about the ability to migrate: if N individuals acquire education in the interest of getting a high-paying job abroad, only pN ($p<1$) are actually able to emigrate.

(Crucially, it is also assumed that p is exogenously given; that is, accurate screening of emigrants by foreign firms does not happen. This keeps the incentive to gain education intact for those who will ultimately fail to emigrate.) The result is that educated individuals who intend but fail to emigrate generate greater domestic human capital formation under brain drain, *ceteris paribus*, than there would be without skilled emigration.

Beine, Docquier, and Rapaport (2001) tested such a model empirically and found that the growth effect of skilled emigration on a poor-country domestic economy depends on the country's population and preexisting emigration rate. Small countries, which tend to have relatively high skilled emigration rates, often had faster growth with lower levels of skilled emigration. On the other hand, they argued that a few large countries with relatively low skilled emigration rates (like Brazil, China, and India) might grow faster if they had more skilled emigration. Thus, there appear to be diminishing marginal growth returns to the magnitude of a country's skilled emigration.

AKM (2008) found that in India during the 1980s and '90s, the brain drain effect was generally significantly larger than the brain bank effect. In other words, Indian innovation would have been more vibrant if India had had a lower level of skilled emigration. However, they found that for "high-value" patents (citations received proxies for a patent's value) the brain bank effect easily outweighed the brain drain effect. Thus, the small portion of patents for which the brain bank effect is strongest may actually represent a large fraction of innovation-based productivity gains.

Commander et al. (2002) concluded that, overall, skilled emigration is likely to impede development in poor countries. However, they qualified this statement in multiple ways. First, they acknowledged that domestically beneficial skilled emigration is theoretically possible, especially in the case of high-population countries with a low skilled emigration rate. Second, they emphasized that the effects of skilled emigration differ by sector. Finally, they stressed that not all the facts are in: brain drain is an empirically complex topic, and more research is needed to draw any final conclusions about the economic effects of skilled migration on poor countries.

There is also a sizable literature on the methodology of using patent data. Hall, Jaffe & Trajtenberg (2001) have emphasized that the procedure is imperfect. First, not all inventions are patented. Second, not all inventions meet the USPTO’s patentability criteria, which stipulate that to be patented, an invention must be novel, non-trivial, and have commercial application. Finally, inventors often make a strategic decision not to patent if they might think it makes better financial sense to rely on secrecy for market control, in light of the time-intensive patent application process.

However, a survey conducted by Jaffe, Trajtenberg, and Fogarty (2000) suggests that patent citations are generally a good proxy for knowledge spillovers. Assuming that their survey responses are accurate and externally valid, their results indicate that the likelihood of a knowledge spillover between two inventors conditional on the observation of a patent citation is significantly greater, both statistically and quantitatively, than the unconditional likelihood. At the same time, Jaffe et al. (2000) qualified their conclusions, as their estimates showed that roughly one-half of citations did not correspond to any knowledge spillover. Thus, patent citations seem to be a good but imperfect proxy for knowledge spillovers.

The logic behind Jaffe et al.’s (2000) findings is fairly simple. The incentives for inventors worldwide to file for a U.S. patent are conducive to transparent information submission. A U.S. patent grants an inventor monopoly rights for her invention in U.S. markets for roughly 20 years,[12] and presumably reporting citations (and other information) fully and accurately increases an inventor's probability of receiving a patent, so she has a strong financial incentive to disclose complete and correct information (Hall, Jaffe & Trajtenberg, 2001).

III. Empirical Strategy

Following AKM (2008), I use patent citations as a proxy for knowledge flows. My methodology for employing patent data in this way follows the empirical strategy developed by Jaffe, Trajtenberg, and Henderson (1993) and adapted to uses more similar to mine by AKM (2008). I start by identifying all patents where every inventor credited on the patent is listed as being located in Korea – I call this sample the Focal Patents. The typical Focal Patent cites a handful of older patents, and this set of cited patents becomes my sample of Cited Patents.

Next, I generate a sample of Control Patents. There is a Control Patent for every Cited Patent. Each Control Patent is a patent *not* cited by any of the Focal Patents that was applied for in the same year as its corresponding Cited Patent, and also is in the same

---

[1] More specifically, patent terms differ slightly depending on whether the application was filed before June 8, 1995. For applications pending or in force before that date, the patent term is either 17 years from the issue date or 20 years from the filing date of the earliest U.S. or international application, the longer term applying. For applications filed on or after that date, the patent term is 20 years from the filing date of the earliest U.S. application (U.S. Department of Commerce, 2008).

[2] The prize of temporary monopoly rights in U.S. markets is globally sought-after: currently, about half of U.S. patents are granted to non-U.S. residents (Rausch, 1999).

technological class as its Corresponding Cited Patent.[3] The set of Controls provides a sort of null hypothesis with which to compare the results from the Cited Patents, enabling us to answer the question, Does knowing that there is a co-location or diaspora connection between a random patent selected from the Control-Cited pool and the corresponding Focal Patent make it more (or less) likely that we've selected a Cited Patent and not a Control? The fact that I'm controlling for technology class and vintage allows me to conclude that any answer I obtain for this question is *not* due to aggregate time effects (e.g., U.S. patents get relatively easier to obtain from Korea) or pre-existing concentrations of technological activity (e.g., the fact that Korea specializes in, say, an industry whose products are relatively easily to patent).

Thus my regression, identical to AKM's (2008), is:

$$Citation = \mathrm{a}_0 + \mathrm{a}_1 CoLocation + \mathrm{a}_2 Diaspora + u_i,\ u_i \sim iid(0, \sigma^2).$$

Citation is a dummy variable equal to 1 if the patent from the Control-Cited pool is from the Cited sample, and equal to 0 if it is from the Control sample. *CoLocation* is a dummy variable equal to 1 if at least one of the inventors on the patent is located in Korea (and 0 otherwise). *Diaspora* is a dummy variable equal to 1 if at least one of the inventors has a Korean surname and none of the inventors is located in Korea (and 0 otherwise).[4]

If we randomly choose a patent from the Control-Cited pool that we know has a

---

[3] If there are multiple potential Controls that match a Cited Patent's technological class (i.e., they have the same six-digit primary U.S. technology classification), then the potential Control with the application date closest to that of the Cited Patent becomes the Control. If a Cited Patent has no potential Controls in the succeeding three patents, then it is dropped from the sample.

[4] Note that while the *CoLocation* and *Diaspora* variables might seem to exhibit perfect multicollinearity, they do not. If the value of one is 0, the value of the other is not necessarily 1: both variables might equal 0 if the patent's inventors all have non-Korean surnames and are located outside of Korea.

value of 0 for both *CoLocation* and *Diaspora*, then an estimate of the probability that the observation is a citation is given by a_0. But if we know that one of the inventors on the patent is located in Korea, then the estimate of the probability that the observation is a citation increases to $a_0 + a_1$. Thus the proportionate increase in the probability that the observation is an actual citation is a_1 / a_0, which we interpret as the proportionate increase in the probability of a knowledge flow due to co-location – I call this the "co-location premium," denoted by γ. Similarly, a_2 / a_0 is an estimate for the "diaspora premium," denoted by δ.

IV. Data

I look at a dataset consisting of the information included on patents issued by the United States Patent and Trademark Office (USPTO), to which inventors from around the world apply for patents. I restrict my sample to those patents granted by 2004 that were applied for during the period 1981-1999. This is virtually the same period examined by AKM (2008),[5] so by restricting my sample to these years I can compare my results with theirs. My full sample consists of roughly three million Focal-Cited pairs.

I identify inventors as being part of the Korean diaspora based on their surname. Specifically, a diaspora patent is one where at least one of the inventors has a Korean surname, and none of the inventors is located in Korea. This identification-via-surname procedure is theoretically susceptible to Type I and Type II error, i.e., false positives and false negatives. A false positive has occurred if I have identified an individual as a member of the Korean diaspora, but she/he is not actually a member. Such an error would

[5] AKM (2008) looked at 1981-2000.

happen for two possible reasons. First, the individual may have a surname that, while spelled the same way as a Korean surname, is not linguistically Korean. Examples of this include the surnames “Lee” (also English) and “Jung” (also German). I drop observations (*not* entire patents) where the (co-)inventor of the Cited/Control patent has such a surname.

Second, even if the individual’s surname is linguistically Korean, she/he may only be Korean by ancestry. For example, perhaps the individual’s great-great-great-great-great-great-great-grandfather emigrated from Korea to Japan, passing his surname down to her/him. In this case, while the individual has a linguistically Korean name, she/he would not in any relevant sense be part of the Korean diaspora: the population that I want to identify as part of the diaspora consists of inventors who have emigrated recently enough that they are potentially still in communication with family, friends, and possibly innovative partners in Korea.[6] I am not too concerned with this issue, however, because the Korean diaspora is mostly a recent phenomenon. As of 2001, roughly 88% of the estimated 5.7-million-person Korean diaspora had left Korea since 1972.

Although I am confident that my identification procedure results in few false positives, my estimated diaspora premium will be biased downward to the extent my analysis suffers from such errors. The false positive observations would consist of inventors who almost certainly have less of a professional connection to Korea than do actual (recent) Korean emigrants. Thus, for these inventors compared with recent Korean emigrants, I assume that the conditional likelihood that a random selection from the

[6] I ignore the odd possibility of an inventor (or someone in the inventor’s preceding familial lineage) arbitrarily changing surnames to a Korean surname. I assume that such incidences are exceedingly rare.

Cited-Control pool is a Cited Patent is closer to the unconditional likelihood, 0.5. Thus, the presence of such "polluting" observations would drive down a_2 and, as a result, decrease the diaspora premium.

A false negative has occurred if I have failed to identify a member of the Korean diaspora as such. Specifically, for my identification procedure I use a list of the 50 most common Korean surnames that encompasses about 72% of the 2000 Korean populace (Korean National Statistical Office, 2000). Korean diaspora members with surnames not on this list are not identified. These omissions only bias my results if Korean emigrants with common surnames are different from Korean emigrants with uncommon surnames. Seeing no reason to think this is the case, I am not worried about false negatives here.[7]

Table 1 reports descriptive statistics for my full sample. The observations of interest – those where the cited patent is either located in Korea or from the Korean diaspora – together make up roughly 1% of the observations. Also worth noting are the fairly large differences in mean application year between the full sample of cited patents, cited patents located in Korea, and cited patents from the Korean diaspora. These two latter subsamples tend to be comprised of more recent cited patents than the full sample. Although having controlled for time effects this fact will not bias my results, it is important to realize that the time frame for which my results are relevant is generally closer to 1990 than it is to 1980.

V. Results

[7] Equivalently to the false positive case, I ignore the odd possibility of an inventor (or someone in the inventor's preceding familial lineage) arbitrarily changing surnames *from* a Korean surname. I assume that such incidences are exceedingly rare.

Table 2 reports OLS results for my full sample.[8] I find evidence of a somewhat large and statistically significant co-location effect. I also obtain statistically significant results for a diaspora effect, but the magnitude of this effect is much smaller than that of the co-location effect. Specifically, the estimated co-location premium is 38.528 (0.214/0.006), and the estimated diaspora premium is just 1.138 (0.006/0.006).

I also interact the co-location and diaspora variables with *Lag*, which measures the length of the period between the application dates of the citing patent and the patent that it cites, and with *Subcategory Match*, which is a dummy that takes a value of 1 if an observation's citing and cited patents are classified under the same NBER three-digit technology category (and 0 if they are not). Table 2 shows these results too. I find that the co-location and diaspora effects are stronger for patent pairs with short lags and stronger when both patents share a technological category. While these trends hold for both the co-location and the diaspora effects, the interaction effects are much stronger for co-location.

Table 3 shows the results when I regress for five of the six NBER one-digit technology classifications (I omit the "Others" classification because it is unclear how results for this category would appropriately be interpreted). Notably, Chemical and Drugs and Medicine appear to have relatively large co-location effects compared with the other classifications, while I estimate a relatively weak co-location effect for Computers and Communications. The diaspora effects for Chemical as well as Drugs and Medicine

[8] I recognize that regressing dummy variables using OLS is potentially problematic for various econometric reasons. (See Long (1997) for a discussion of these issues.) Thus, I have run my regressions in the logistic form. These results are not significantly different from my OLS results, however, and for purposes of concision and comparison with AKM's (2008) OLS results, I have reported only my OLS results.

are relatively low, as well as statistically insignificant, whereas the diaspora effect for Electrical and Electronic is quite high (an order of magnitude higher than the baseline diaspora effect).

Finally, I separate "Recent" and "Non-recent" patents (i.e., patents applied-for since and before, or during, 1996). These results are in Table 4. I find similar co-location effects, but a markedly stronger diaspora effect for Recent patents.

VI. Discussion

The first striking thing about my results is their similarity to those that AKM (2008) found for India. Such a resemblance was not expected *a priori*, since India and Korea's economies are qualitatively quite different from one another. For both India and Korea, it appears that the harmful effect of skilled emigration on domestic innovation caused by the loss of domestic collaborators easily outweighs the beneficial diaspora effect of skilled emigration (although again, there is evidence that this latter effect, though small in size, still exists). I estimate a co-location premium that is about 34 times larger than the diaspora premium (38.528/1.138). In comparison, AKM's (2008) ratio is about 6 (0.792/0.127), suggesting that the magnitude relative to brain bank of brain drain in Korea is larger than the magnitude in India.

A primary potential culprit for my lack of a meaningful diaspora effect is a failing strategy for identifying diaspora members. If my diaspora sample does not in fact represent what I think it does – that is, if it is "polluted" by the inclusion of individuals who are not recent Korean immigrants (perhaps non-Koreans, or those of Korean ancestry who have been settled outside of Korea for generations) – then the measured effect is biased downward.

It is also instructive to compare my categorical regressions with AKM's (2008). Like them, I find that for Computers and Communications, the co-location effect is relatively low, presumably because intranational collaboration is less essential in this industry than it is for others. This is intuitive: an industry characterized by communications technology should be the most apt to take advantage of that technology's powers, rendering domestic face-to-face contact, and the co-location effect, less important. Again like AKM (2008), I find a relatively strong co-location effect for Drugs and Medicine (though my result is not nearly as large as theirs). This finding is intuitive as well: the pharmaceutical industry relies on extensive face-to-face collaboration, so I would expect to see a strong co-location effect here.

It is also notable that the co-location effect is stronger for patent pairs with short lags, and stronger when both patents share a technological category. There are numerous plausible explanations for these observed phenomena, but it is especially plausible for both findings that short lag times and a shared technological category reflect personal collaboration. Short lag times are on the order of two or three years, a time span that could easily include a period of collaboration between domestic inventors, and almost certainly domestic collaboration occurs with more frequency and intensity within technological classes than it does between them.

Finally, as for the recent/non-recent split-sample regressions, I speculate that the higher diaspora effect for recent patents is due to better communications technology, and thus facilitated international innovative collaboration.

VII. Conclusion

I find evidence that the emigration of skilled Koreans has a generally negative effect on innovation in Korea. This conclusion is based on estimations that yield a co-location premium that is statistically significant and much larger than the diaspora premium. Auxiliary regressions do not generally overturn this finding, but they add nuance to my analysis in various ways. Variables that influence the size of the co-location and diapsora premium estimations include: lag time (the length of the period between the application dates of the citing patent and the patent it cites), subcategory match (whether the citing patent and the patent it cites are classified in the same technology subcategory), the technology category of the cited patent, and the vintage of the cited patent (whether it was applied for after 1995 or not).

The longer the lag time, the lower the co-location and diaspora premiums. The presence of a subcategory match, however, increases the premiums. The co-location effect is particularly strong in the Drugs and Medicine sector and noticeably weak in the Computers and Communications sector, while the diaspora effect is especially strong in the Electrical and Electronic sector and notably weak in the Drugs and Medicine sector. Finally, both premiums are larger when the Focal Citation is post-1995 (as opposed to older).

There are unfortunately three main limitations to my analysis, which AKM (2008) also note as applying to their research. The first concerns the indirect nature of my measurement strategy, which makes two assumptions. One, I assume that I may use patent data to proxy for knowledge flows. Two, I assume that innovation is driven by access to knowledge. Although I believe they are plausible, either of these assumptions may actually be false. Perhaps future analysis should use more direct observation techniques to examine the assumptions at the heart of this paper about patent data as a proxy for knowledge flows, and access to knowledge as the driver of innovation.[9]

The second notable limitation of my paper is that it assumes skilled emigration must entail a tradeoff between having a large domestic stock of innovators and having a weak network of diaspora connections. In other words, it may be that the domestic stock of innovators increases as skilled individuals emigrate, a phenomenon that could arise through some combination of two possible effects. First, the possibility of migration could induce higher investments in education due to greater returns abroad (Beine, Docquier, Rapaport, 2001). Second, increases in financial remittances due to skilled emigration could prompt greater investments in human capital by the recipients (Yang, 2006). A final important issue with this paper arises from the implications of time-constraining my sample (ending in 2000). Essentially, it is possible that in the 11 years since my sample ends, changes in Korea's economy have affected emigration patterns. There have been no dramatic shocks to its economy, as Korea has generally maintained stable growth, hovering around 4-5% per year (apart from the global recession starting in 2007, during which period growth rates have been close to 0%). Nonetheless, it is possible that emigration patterns have changed.

In short, this paper has found that skilled emigration generally has the effect of brain drain in the case of Korea. However, the extent to which this effect dominates the brain bank effect varies within different types of subsamples. Future research should

[9] Indeed, in AKM (2008), the authors note that they are beginning to use information about the career paths and productivity of mobile scientists to see more directly how migration flows affect national innovation.

address the issues outlined above: the assumptions of my model and empirical specification, and potential changes in the estimations found here due to recent changes in the structure and state of Korea's economy. Additionally, it would be useful to see analysis of brain drain versus brain bank done for other countries.

References

Agrawal, A., Kapur, D., & McHale, J. (2008). Brain drain or brain bank? The impact of skilled emigration on poor-country innovation. *National Bureau of Economic Research Working Paper Series*, (Working Paper 14592), Retrieved from http://www.nber.org/papers/w14592

Basu, S., & Weil, D. (1998). Appropriate technology and growth. *The Quarterly Journal of Economics*, 113(4), Retrieved from http://qje.oxfordjournals.org/content/113/4/1025.full.pdf

Beine, Docquier, Rapaport, 2001 – cited by AKM 2008

Cohen, W., & Levinthal, D. (1989). Innovation and learning: the two faces of R & D. *The Economic Journal*, 99(397), Retrieved from http://www.jstor.org/stable/2233763

Hall, B. H., Jaffe, A.B., & Trajtenberg, M. (2001). The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. *National Bureau of Economic Research Working Paper Series*, (Working Paper 8498), Retrieved from http://papers.nber.org/papers/w8498.pdf

Jaffe, A., & Trajtenberg, M. (2002). *Patents, citations, and innovations: a window on the knowledge economy*. Cambridge, MA: The MIT Press.

Keller, W. (2002). Geographic localization and international technology diffusion. *The American Economic Review*, 92(1), Retrieved from http://www.jstor.org/stable/3083324

Kuhn, P., & McAusland, C. (2006). The international migration of knowledge workers: when is brain drain beneficial?. *National Bureau of Economic Research Working Paper Series*, (Working Paper 12761), Retrieved from http://www.nber.org/papers/w12761

Rausch, L. M. National Science Foundation, Division of Science Resources Studies. (1999). *U.S. inventors patent technologies around the world* (NSF Publication No. 99-329). Arlington, VA: Retrieved from http://www.nsf.gov/statistics/issuebrf/sib99329.htm

U.S. Department of Commerce, Patent and Trademark Office. (2008). *2701 patent term [R-2] - 2700 patent terms and extensions* (35 U.S.C. 154). Retrieved from http://www.uspto.gov/web/offices/pac/mpep/documents/2700_2701.htm#sect271

The World Bank. (2011). *GDP per capita (current US$)*. Retrieved from http://data.worldbank.org/indicator/NY.GDP.PCAP.CD

Table 1

Descriptive Statistics

Observations (patent citing-patent cited pairs)	16,546,240
Observations where patent cited is located in Korea	48,917 (0.28%)
Observations where patent cited qualifies as "Korean diaspora"	183,514 (1.06%)
Mean application year for cited patents	1981
Mean application year for cited patents located in Korea	1986
Mean application year for "Korean diaspora" cited patents	1991

Table 2

General OLS Estimates[10]

	(1) Focal Citation	2 Focal Citation
Co-location	0.138***	0.266***
	(0.000727)	(0.00199)
Diaspora	0.00544***	0.00635***
	(0.000367)	(0.000794)
Co-location*Lag		-0.0238***
		(0.000351)
Diaspora*Lag		-0.000511***
		(8.39e-05)
Co-location*Subcategory Match		0.0515***
		(0.00191)
Diaspora*Subcategory Match		0.00533***
		(0.000790)
Constant	0.00513***	0.00559***
	(3.90e-05)	(4.27e-05)
Observations	3,668,396	3,317,448
R-squared	0.010	0.018

Standard errors in parentheses
*** p<0.01, ** p<0.05

Table 3

[10] All regressions in this paper are heteroskedasticity-robust.

OLS Estimates by NBER One-Digit Code

Dependent Variable: Focal Citation	Chemical	Computers & Communications	Drugs & Medical	Electrical & Electronic	Mechanical
Co-location	0.232***	0.192***	0.218***	0.202***	0.290***
	(0.0181)	(0.00959)	(0.0293)	(0.00737)	(0.0165)
Diaspora	0.00117	0.00358***	0.000109	0.0174***	0.00425***
	(0.000643)	(0.00135)	(0.000620)	(0.00190)	(0.00147)
Constant	0.00279***	0.0108***	0.00158***	0.0117***	0.00401***
	(6.53e-05)	(0.000154)	(7.32e-05)	(0.000144)	(7.79e-05)
Observations	661,806	461,465	299,020	569,473	661,775
R-squared	0.015	0.012	0.018	0.017	0.022

Standard errors in parentheses
*** p<0.01, ** p<0.05

Table 4

OLS Estimates by "Vintage"

	Focal Citation for Recent Patents	Focal Citation for Non-Recent Patents
Co-location	0.222***	0.197***
	(0.00672)	(0.00714)
Diaspora	0.0104***	0.00289***
	(0.00146)	(0.000527)
Constant	0.0140***	0.00420***
	(0.000148)	(4.34e-05)
Observations	647,561	2,246,996
R-squared	0.020	0.012

Standard errors in parentheses
*** p<0.01, ** p<0.05