BRAIN DRAIN OR BRAIN BANK? THE IMPACT OF SKILLED EMIGRATION
ON POOR-COUNTRY INNOVATION

Ajay Agrawal
Devesh Kapur
John McHale

Brain Drain or Brain Bank? The Impact of Skilled Emigration on Poor-Country Innovation
Ajay Agrawal, Devesh Kapur, and John McHale
NBER Working Paper No. 14592
December 2008
JEL No. O3,O33

ABSTRACT

The development prospects of a poor country depend in part on its capacity for innovation. The productivity of its innovators depends in turn on their access to technological knowledge. The emigration of highly skilled individuals weakens local knowledge networks (brain drain), but may also help remaining innovators access valuable knowledge accumulated abroad (brain bank). We develop a model in which the size of the optimal innovator diaspora depends on the competing strengths of co-location and diaspora effects for accessing knowledge. Then, using patent citation data associated with inventions from India, we estimate the key co-location and diaspora parameters; the net effect of innovator emigration is to harm domestic knowledge access, on average. However, knowledge access conferred by the diaspora is particularly valuable in the production of India's most important inventions as measured by citations received. Thus, our findings imply that the optimal emigration level may depend, at least partly, on the relative value resulting from the most cited compared to average inventions.

Ajay Agrawal
Rotman School of Management
University of Toronto
105 St. George Street
Toronto, Ontario M5S 3E6
CANADA
and NBER
ajay.agrawal@rotman.utoronto.ca

Devesh Kapur
3600 Market Street, Suite 560
Centre for Advanced Study of India
University of Pennsylvania
dkapur@sas.upenn.edu

John McHale
Queen's School of Business
Goodes Hall
143 Union Street
Kingston, Ontario
Canada K7L 3N6
jmchale@business.queensu.ca

1. Introduction

The development impact of skilled migration from poor countries has long been a contentious issue. Scholars are even far from a consensus on the narrower question: What is the impact on innovation when a poor country loses a large fraction of its science and engineering workforce through emigration?

One school of thought argues that such talent is often wasted at home. Migration to more supportive environments raises global innovation, and some gains flow back to the poor country through the imports of products with improved technology or lower cost (Kuhn and McAusland, 2006). Furthermore, gains may flow back to the developing country via returnees with enhanced skills, personal connections, and ideas for innovation (Saxenian, 2005).

However, another school of thought focuses on the importance of domestic technology innovators. Despite their typically considerable distance from the technology frontier, domestic innovators could be important for various reasons: 1) international technology diffusion may be slow due to the localization of knowledge spillovers;[1] 2) rich-country innovation may not properly address the needs of poorer countries;[2] and 3) domestic knowledge production may be necessary to create the capacity to absorb foreign technology.[3]

However, the most important form of innovation for a poor country is likely the adoption of technologies developed elsewhere (World Bank, 2008). In other words, the greatest opportunities for growth in a poor country lie in moving towards the international frontier rather than in pushing that frontier forward. Highly skilled domestic innovators are likely to be central to this catch-up process.

The availability of new datasets showing high and generally increasing poor- to rich-country emigration rates for tertiary-educated workers has heightened concern about

---

[1] Keller (2002) presents evidence on international technology diffusion. Also, Jaffe, Trajtenberg, and Henderson (1993, hereafter "JTH") document the localization of knowledge spillovers. Thomson and Fox Keane (2005) provide important refinements to the JTH approach.

[2] Basu and Weil (1998) present a model in which the appropriate technology is specific to a country's available inputs.

[3] Cohen and Levinthal (1989) argue that R&D has the indirect benefit of increasing a firm's capacity to absorb technology being developed elsewhere. Caselli and Coleman (2001) show that importing technology embodied in computers is positively related domestic human capital stocks.

the "brain drain" (Docquier and Marfouk, 2005; Dumont and Lemaitre, 2005).[4] These rates are extremely high for many small, poor countries. For example, Docquier and Marfouk estimate that 41 percent of those with a tertiary education and born in a Caribbean country now live in an OECD country.[5]

At the same time, substantial flows of financial remittances also highlight the many benefits to the country of origin from international migration, extending not just to money but also to the flows of ideas and technologies from its diaspora. The latter raises the possibility that the migration of skilled human capital from poor countries may not just be a negative "brain drain"; it could also have more a positive effect as a "brain bank," accumulating knowledge abroad and facilitating its transfer back to domestic inventors (Kerr, 2008).

In this paper we develop and estimate a model in which the *access* of domestic innovators to knowledge drives innovation. This contrasts with Paul Romer's classic model of innovation and growth, where the *existence* of new ideas that might be built upon is the basis of innovation and "anyone engaged in research has free access to the entire stock of knowledge" (Romer, 1990, p. S83). For a poor country the degree of access to the existing stock of knowledge is likely of particular importance, warranting the shift in emphasis.[6]

The main building block of our model is the Knowledge Flow Production Function (KFPF). For any domestic innovator, the KFPF gives the probability of receiving knowledge from any other innovator based on structural aspects of their

---

[4] These rates measure the absence of tertiary-educated nationals from the economy. In many cases, inventors acquired their education abroad, and so the rates are not actually measures of the outflow of individuals who were trained domestically (the usual connotation of the term "brain drain").

[5] Although tertiary emigration rates tend to be considerably lower for larger developing countries, emigration rates for the most educated and talented are much higher (Kapur and McHale, 2005). To take the example of India, the overall tertiary emigration rate is estimated to be about four percent, while the emigration rates from the elite Indian Institutes of Technology (IITs) is substantially higher. An analysis of the brain drain from the graduates of IIT-Mumbai in the 1970s revealed that 31 percent of its graduates settled abroad, while the estimated migration rate of engineers more generally was 7.3 percent (Sukhatme and Mahadevan, 1987). Recent alumni data in the case of IIT-Kharagpur found 4,007 registered alumni in India, 3,480 in the U.S., and another 739 spread over 59 countries. See http://www.iitfoundation.org/directory/stats/ Accessed September 3, 2004.

[6] Klenow and Rodriguez-Clare (2004) argue that international technology spillovers explain some of the basic facts about cross-country income levels and growth rates. Using a calibrated endogenous growth model, they show that relatively small barriers to international technology diffusion — knowledge access, in the language of our model — can lead to large cross-country differences in income levels.

relationship. We focus in particular on whether innovators are co-located in the domestic economy, share a diaspora connection, or are unconnected by location *or* nationality. We assume a domestic innovator's output depends on her overall access to knowledge from domestic, diaspora, and foreign sources. The total innovation output of the national economy is then simply the sum of the innovation outputs of domestic inventors.

Hence the central tradeoff in the model: The emigration of a domestic innovator leads to a direct reduction in domestic innovator stock and weakens the network of co-located innovators but, on the other hand, it can also lead to new access to foreign-produced knowledge through the diaspora. The latter effect will be stronger where there are enduring connections to the diaspora and where emigrant innovators increase their knowledge stock by moving to environments with better resources, colleagues, and incentives to innovate.

These conflicting effects lead to the idea of the *optimal diaspora* — the emigrant stock that maximizes national knowledge access. We show that the optimal diaspora depends on the relative size of the co-location and diaspora effects. We also examine extensions to the model that allow for circulation between the home economy and the diaspora, non-random selection of emigrants and returnees, and heterogeneous KFPFs based on the importance of the innovation.

The empirical challenge is to identify the co-location and diaspora effects in the KFPF. To accomplish this, we construct a novel sample from patent data linked with Indian last name data and then build on a widely-used method that employs patent citations as a proxy for knowledge flows between inventors and "matched citations" to control for the underlying distribution of inventive activity across geographic and ethnic space. This allows us to isolate the causal impacts of location and diaspora connections on the probability of a knowledge flow.

Our empirical focus is on the knowledge access of frontier innovators in a poor country. This focus allows us to take advantage of the rare instance of a "paper trail" for national and international knowledge flows afforded by the recording of citations on a patent (Jaffe et al., 1993). We stress again that frontier innovation will typically be of second order importance for growth in poor countries. However, to the extent that networks for knowledge access operate similarly for frontier- and implementation-based

innovation, the findings on the drivers of knowledge flows at the frontier should provide a valuable clue to the relative importance of local versus diaspora knowledge networks, and thus the likely impact of skilled emigration on poor-country knowledge access and innovation.

The rest of the paper is organized as follows. In the next section, we model an optimal innovator diaspora. In section 3 we describe our empirical strategy for identifying the causal effects of co-location and diaspora membership on knowledge flows. In section 4 we describe our patent-citation and Indian-name data, presenting our results in section 5. In section 6 we discuss the implications of our findings.

2. The Optimal Diaspora

2.1 Permanent migration

We first develop a simple model of an optimal innovator diaspora, abstracting initially from the possibility of return, innovator heterogeneity, and differences in the KFPF related to the value of innovations. Our focus is on knowledge production in a relatively poor country, which we call India without loss of generality. The essential idea is that the productivity of India-residing innovators depends on their access to knowledge. This access in turn depends on their relationships to other innovators and also on the productivity of those innovators. We allow connectivity to be affected by co-location and co-nationality and also for the possibility that innovators are more productive abroad because of better incentive structures and resources. The emigration of an innovator results in a direct loss to the stock of Indian innovators, thinning domestic knowledge networks, but could actually increase total knowledge access if the diasporic linkages and productivity gains are large enough. The model's goal is to identify the size of the diaspora that maximizes the access to knowledge of India-residing innovators.

The KFPF captures the probability of a knowledge flow between any pair of innovators (at least one of whom is a resident in India) based on certain structural relationships between those innovators. The probability of a knowledge flow to a given Indian innovator, i, from another innovator, j, is given by:

(1) $\qquad K_{ij} = f + \alpha_{ij}\gamma f + \beta_{ij}\,\delta f,$

where f is the (base-case) probability of a knowledge flow if the other innovator is neither a resident in India nor a member of the Indian diaspora, α_{ij} is a dummy variable that takes the value of 1 if innovator j is also a resident in India, γ is the proportionate knowledge-flow premium from being co-located, β_{ij} is a dummy variable that takes the value of 1 if j is a member of the Indian diaspora, and δ is the proportionate premium for being in the diaspora. Note that the value of γ reflects the combined effects of co-location and the (possibly negative) relative productivity effect of doing science in India, whereas the value of δ reflects the effect of the diaspora connection and any productivity gap that might exist between members of the diaspora and foreigners. Denoting the total number of Indian innovators (both India-based and emigrant) as N, the total size of the Indian scientific diaspora as D, and the total number of foreign innovators as Z, the total (expected) knowledge flow to i is given by this knowledge access equation:

$$(2) \qquad K_i = Zf + (N - D - 1)(1 + \gamma)f + D(1 + \delta)f.$$

The aggregate knowledge access of India-residing innovators is found by multiplying both sides of (2) by the total number of such innovators:

$$(3) \qquad K = (N - D)K_i = (N - D)Zf + (N - D)(N - D - 1)(1 + \gamma)f + (N - D)D(1 + \delta)f.$$

Innovation is assumed to depend on both the access to knowledge and the absorptive capacity to turn that knowledge into valuable economic output. In this paper, we focus only on knowledge access and assume that greater knowledge access is associated with greater output: $I_i = I(K_i)$; $\dfrac{\partial I_i}{\partial K_i} > 0$. Of course, the knowledge access to innovation will be country specific and will depend, *inter alia*, on the available capital stock, the presence of complementary human capital, and security of property rights.

The diaspora size, D^*, that maximizes national knowledge access (and thus innovation) is found from the first-order condition:

(4)
$$\frac{\partial K}{\partial D} = 2D^{*}(\gamma - \delta) - Z - N(1 + 2\gamma - \delta) + (1 + \gamma) = 0.$$

Rearranging (4), we obtain an expression for the optimal diaspora as a fraction of the total stock of Indian innovators:

(5)
$$\frac{D^{*}}{N} = \left(\frac{1 + 2\gamma - \delta}{2(\gamma - \delta)}\right) + \left(\frac{1}{2(\gamma - \delta)}\right)\left(\frac{Z}{N}\right) - \left(\frac{1 + \gamma}{2(\gamma - \delta)}\right)\left(\frac{1}{N}\right).$$

Equations (3) through (5) allow us to characterize the conditions under which a diaspora is beneficial for knowledge access and innovation. We do this in two steps. First, an examination of equations (3) and (4) reveals that, for this first-order condition to identify a maximum, we require from the second-order condition that δ is greater than γ:

(6)
$$\frac{\partial^{2} K}{\partial D^{2}} = 2(\gamma - \delta) < 0 \quad \Rightarrow \quad \delta > \gamma.$$

Otherwise, the national knowledge access will decline monotonically with the size of the diaspora (see the first equality in equation (4)). We first assume that this condition does not hold. A positive diaspora is never beneficial in this case. This necessary condition can be given a more intuitive explanation. Suppose in the extreme that the potential emigrant contributes nothing directly to domestic innovation while at home. Their only contribution comes indirectly from the knowledge that flows from them to other domestic innovators. Whether their absence helps or harms, in that case, depends simply on whether domestic innovators access more knowledge from them when at home or abroad — i.e. on the relative magnitudes of δ and γ.

Second, we use (7) to identify the necessary and sufficient condition for a strictly positive diaspora to be beneficial:

$$(7) \qquad \frac{D^*}{N} > 0 \quad \Leftrightarrow \quad \delta > 1 + 2\gamma + \frac{Z}{N} - \frac{1+\gamma}{N}.$$

This condition is quite stringent. Even in the extreme case where N is sufficiently large enough that we can ignore the last two terms and where there is no co-location premium (i.e. $\gamma = 0$), the diaspora premium must be greater than 100 percent for a diaspora to be beneficial for the total knowledge flow to India-residing innovators.[7]

2.2 Circulatory migration

The model with permanent migration abstracts from one potentially important element — the return of emigrant innovators. Such returnees are likely to have developed connections with foreign innovators while away, connections that may endure on their return to facilitate ongoing knowledge flows.[8] To explore the implications of return, we examine the steady state of a simple extension of the model that allows for circulation.

At any point in time the change in the diaspora share mechanically depends on the emigration rate (e), the return rate (r), the growth rate of new Indian scientists (n), and the initial diaspora share[9]:

---

[7] From (5) we can see that the optimal diaspora share converges to one half as δ approaches infinity. In other words, it will never be optimal for a country to have more than half its innovators abroad Although in reality we expect the optimal diaspora share to be well below one half, this finding is of interest because there are several countries for which the number of tertiary-educated nationals residing abroad is greater than the number residing at home (Docquier and Marfouk, 2005). These general emigrant shares are likely to underestimate the share of innovators, given the tendency for emigrant shares from poor countries to rise with education level. The model suggests that this is detrimental to knowledge production no matter how large the productivity gains are from emigrating and no matter how strong the diasporic connections are. This result implies that countries must have a sufficient number of innovators at home to reap the benefits of emigrant-related productivity gains and diasporic connections.

[8] Agrawal, Cockburn, and McHale (2006) provide evidence of the impact of enduring social capital acquired during past co-location on subsequent knowledge flows.

[9] The emigration rate is the fraction of the stock of India-residing innovators ($N - D$) who emigrate each period; the return rate is the fraction of the innovator diaspora (D) who return each period; the new innovator growth rate is the proportionate growth in the total stock of Indian innovators (N).

$$d\left(\frac{D}{N}\right) = \frac{1}{N}dD - \frac{D}{N^2}dN$$

(8)

$$= \frac{1}{N}\left(e(N-D) - rD\right) - \frac{D}{N}n$$

$$= e - (e+r+n)\frac{D}{N}.$$

Setting (8) equal to zero, we have an expression for the *steady-state* diaspora share:

(9) $$\left(\frac{D}{N}\right)^{ss} = \frac{e}{e+r+n}.$$

For a given steady-state diaspora share and a given *n*, the steady state is consistent with an infinite number of (*e*, *r*) pairs. One possibility is that a given disapora share is observed with very low emigration and return rates, such that the diaspora and the stock of scientists remaining in India have the character of "stagnant pools." However, the same diaspora share could be observed with much higher emigration and return rates, such that the diaspora and India-residing stocks have more the character of "circulating pools" — whom Saxenian (2006) calls the "New Argonauts" after the Greeks who sailed with Jason in search of the Golden Fleece. The nature of the India-residing stock is likely to have implications for the strength of their connections to domestic, diaspora, and foreign scientists, with the relative strength of connections to innovators abroad increasing with the propensity to circulate.

Given perpetual circulation, the expected fraction of time that any Indian innovator will spend in the diaspora will converge to the steady-state disapora share for any strictly positive return rate. Looked at from the viewpoint of innovators currently residing in India, the expected fraction of time spent abroad in the past is therefore increasing in the steady-state diaspora share. An implication is that with a positive return rate a higher diaspora share is likely to be associated with stronger connections to foreign

innovators.[10] This suggests a potential problem with inferences about optimal diaspora size based on the static model.

The static model is developed on the premise of proportional co-location and diaspora premiums that are independent of the size of the diaspora itself.   This independence would allow us to estimate these premiums and then make inferences about the optimal size of the diaspora. However, if a larger diaspora share is associated with stronger connections to innovators abroad, then it is likely that the proportional co-location and diaspora premiums will be affected by the size of the diaspora. But when these premiums depend on the size of the diaspora, we face the problem that we cannot use estimates of these premiums (based on a time period with a given diaspora) to infer the size of the optimal diaspora. We outline our method for identifying the importance of return in the empirical strategy section below.

2.3 Heterogeneous innovators and non-random selection

We have assumed that all innovators are equally productive. However, we can weaken this assumption without affecting the results if we assume that emigrants and returnees are random selections from the stocks of India-residing innovators and the diaspora, respectively. The results are obviously affected, however, if emigrants and returnees are non-random selections from their respective pools.

Suppose, for example, that the most productive innovators have a higher probability of emigrating (possibly because they have a higher probability of qualifying for a visa such as the U.S. H-1B). This positive selection will tend to augment the absence-related loss to India, suggesting an even lower optimal diaspora.

Suppose further that returnees are a positive selection of the already positively selected diaspora. It is possible that a few truly outstanding returnees — coming back with significantly enhanced productivity due to their time spent abroad — could have a major impact on Indian innovation. In this case, our model would give a misleading picture of the long-run effect of migration. We outline our tests for non-random selection in the empirical strategy section below.

---

[10] When the return rate is zero, such that the current India-residing stock has spent no time abroad, the strength of the connection to foreign scientists is independent of the size of the diaspora.

2.4 Knowledge access and the value of an innovation

A core idea of the model is that knowledge access drives innovation. To keep the model as simple as possible, we have made the restrictive assumption that the way relationships facilitate knowledge access is the same for all innovators. One obvious concern is that the KFPF differs systematically based on the value of the innovation. For example, high-value innovations may draw relatively more on frontier knowledge through the diaspora.[11] We outline our method for testing for systematic differences in the KFPF in the empirical strategy section below.

3. Empirical Strategy

To empirically implement the model we follow the well-established approach of using patent citations as (noisy) indicators of knowledge flows between inventors.[12] Building on the technique developed in Agrawal, Kapur, and McHale (2007), we choose a control patent to match every cited patent by a patenting Indian inventor. The controls are chosen to match the technology class and timing of each of the cited patents as closely as possible.

Assuming this matching procedure is successful, the cited and control patents will have the same geographic distribution even where inventive activity is geographically concentrated within narrow technological specializations. Thus, if inventor co-location and co-membership in an ethnic diaspora play no role in facilitating knowledge flows, knowing that the inventor on the focal patent and the inventor on the cited patent have a location or a diaspora connection should be of no help in distinguishing an actual citation from a control. On the other hand, if co-location and diaspora membership are disproportionately associated with actual citations, we can use the estimated premiums as measures of the causal effects of location and diaspora connections on knowledge flows.

The model points to the central empirical task: the identification of δ and γ parameters. If we find that δ is less than γ, then emigration is detrimental to knowledge

---

[11] As another example of how the KFPF may be context specific, Nandra and Khanna (2007) find that diaspora connections are more important for Indian software entrepreneurs operating in weak institutional environments.

[12] See Jaffe and Trajtenberg (2002) for key developments in the use of patent citation data to track knowledge flows.

flows. Even if δ is greater than γ, the gap will have to be large for a diaspora to be beneficial.

We run the following regression to identity the key parameters:

(8) $\quad Citation = a_0 + a_1 CoLocation + a_2 Diaspora + u_i, \qquad u_i \sim iid(0, \sigma^2).$

The dependent variable throughout our analyses is *Citation*, which is a dummy variable assigned a value of 1 if the "citation" is an actual citation, thus reflecting a knowledge flow, or 0 if it is a control. We use two main explanatory variables. *Co-location* is a dummy variable assigned a value of 1 if at least one of the inventors on the cited patent is located in India (and thus is co-located in the same country as the inventors of the focal patent who are all located in India, by construction) and 0 otherwise. *Diaspora* is a dummy variable assigned a value of 1 if at least one of the inventors has an Indian last name and none of the inventors are located in India.

If we randomly choose a cited/control patent for which we know that both *Co-location* and *Diaspora* equal 0, then an estimate of the probability that the observation is an actual citation is given by $\hat{a}_0$. However, if we know that the inventors are co-located, the estimate of the probability that the observation is an actual citation is given by $\hat{a}_0 + \hat{a}_1$. The proportionate increase in the probability that the observation is an actual cited patent is then $\dfrac{\hat{a}_1}{\hat{a}_0}$, which we take to identify the proportionate increase in the probability of a knowledge flow caused by co-location — that is, an estimate of γ. Similarly, $\dfrac{\hat{a}_2}{\hat{a}_0}$ provides an estimate of δ.

Co-location and diaspora membership are unlikely to be equally important for all knowledge flows. Thus, we examine: 1) differences based on elapsed time between the focal patent and the cited patent (we expect that relationships are less important the longer the invention is in the public domain); 2) differences based on whether the knowledge is flowing across or within technological boundaries (we expect that relationships based on location and co-ethnicity are more important for inventors who do

not share a technology specialization);[13] 3) differences based on broad technology class (for example, owing to differences in the importance of non-codifiable knowledge, knowledge exchange in computing research might be less dependent on proximity than knowledge exchange in medical research);[14] and 4) differences between "vintages," by comparing the co-location and disapora parameters for earlier versus later focal patents.[15]

We examine the importance of return in two ways. First, we simply measure how many of the India-residing inventors are actually returnees. A finding that returns are rare will provide support for the constant parameters assumption. Second, we determine whether the co-location and diaspora premiums are systematically different for returnees compared with inventors who never emigrated. Even if return is a significant phenomenon, a finding that the KFPFs are not significantly different for returnees and non-returnees will also provide support for the constant parameter model.

To determine the importance of non-random selection we use the number of forward citations to an invention as a proxy for the quality of the inventor (rather than the invention). To determine if emigrants are differentially selected, we look forward from the application dates of each focal patent to see if the inventors subsequently emigrated. We then compare the "quality" of the patents of non-emigrants to those of emigrants. Similarly, we compare the "quality" of the patents of returnees and non-returnees to make inferences about returnee selectivity.

Finally, we look for differences in the KFPF based on the value of an innovation. Our indicator of innovation value will be the number of forward citations to a patent (Trajtenberg, 1990). Differentiating by the value of an innovation, we then test whether the co-location and diaspora effects — i.e. the KFPFs — are systematically different for higher value innovations.

---

[13] Technology co-specialization is measured by the focal patent and the cited/citing patent sharing the same NBER two-digit technology classification.

[14] We divide focal patents in broad technological classes based on NBER one-digit technology classifications.

[15] The motivation for testing for such effects is that advances in communications technology may have changed the relative value of location-based and diaspora-based relationships.

4. Data

We use patent citations as a proxy for knowledge flows.[16] As such, focal-cited patent pairs are the unit of analysis. (Cited patents are listed as references on the focal patent.) First, we identify all patents issued by the United States Patent and Trademark Office by 2004 (inclusive) that were applied for during the period 1981-2000 (inclusive) where all inventors are located in India.[17, 18] There are 831 such patents. These are our focal patents. On average, they cite 6.7 patents, generating 5,527 focal-cited patent pairs.

Next, we identify control patents that match the cited patents on two dimensions: vintage and technology area. Specifically, control patents must match cited patents on application year and the full six-digit primary U.S. technology classification. If we cannot identify a suitable control patent, then we drop the observation.

If we identify more than one suitable control, then we select the patent that matches as many full secondary six-digit classifications as possible. If more than one potential control patent with "the best" match on technology classifications exists, then we select the one with the application date closest to that of the cited patent. Based on these criteria, we find control patents for 86 percent of our cited patents. Thus, our sample consists of 9,520 observations of which, by construction, half are focal-cited patent pairs and the other half are focal-control patent pairs. Approximately two percent

---

[16] It is important to note that although we use citations as a proxy for knowledge flows we recognize that citations are not straightforward to interpret. Patents cite other patents as prior art, with citations serving to delineate the property rights conferred. Some citations are supplied by the applicant, others by the patent examiner (Alcacer and Gittelman (2006); Hegde and Sampat (2005)), and some patents may be cited more frequently than others because they are more salient in terms of satisfying legal definitions of prior art rather than because they have greater technological significance. Cockburn et al. (2002) report, for example, that some examiners have "favorite" patents that they cite preferentially because they teach the art particularly well. Nonetheless, we are of the opinion that even examiner-added citations may reflect a knowledge flow. Jaffe et al. (2002) surveyed cited and citing inventors to explore the meaning of patent citations and found that approximately one-quarter of the survey responses correspond to a "fairly clear spillover," approximately one-half indicate no spillover, and the remaining quarter indicate some possibility of a spillover. Based on their survey data, the authors conclude that "these results are consistent with the notion that citations are a noisy signal of the presence of spillovers. This implies that aggregate citation flows can be used as proxies for knowledge-spillover intensity, for example, between categories of organizations or between geographic regions" (p. 400).

[17] We use information from the inventor country address field, not the assignee field.

[18] Since we focus on knowledge flows proxied by citations, we also impose the restriction that focal patents make at least one citation. The majority of patents (84 percent) meet this criterion. Those that don't, either because they make no citations or because the citations they make are to patents issued before 1976 and are thus not in our database, are dropped from the sample.

of the cited/control patents are co-located with the focal patent (table 1); approximately four percent of the cited/control patents are invented by the diaspora (table 1).

We identify inventors as being members of the Indian diaspora based on their last names.[19] When there are multiple inventors we define a diaspora patent as one where at least one inventor has an Indian last name and none of the inventors reside in India. We generate Indian name data from a list of 213,622 unique last names compiled by merging the phone directories of four of the six largest cities in India.[20] We then code these names based on their likelihood of being Indian.[21] The list of names we use in this study includes only the 6,885 last names that were coded as "extremely likely to be Indian."[22]

Although we construct our dataset from focal patents applied for during the period 1981-2000, the mean application year is 1997 (table 1). These data are skewed towards

---

[19] There is reason to believe that the Indian diaspora are reasonably likely to stay connected to individuals in their home country. For example, members of the U.S. resident Indian diaspora identify strongly with their ethnicity, perhaps partly because many are of a recent vintage. Of the 2001 Indian-American population residing in the U.S., those born in the U.S. were fewer than those born in India (0.7 million versus one million; source: U.S. Census Bureau, Current Population Survey, March Supplement, various years.) Furthermore, more than one third of the Indian-born came after 1996 and more than half after 1990. The Indian-born population in the U.S. numbered only 12,296 in the 1960 census. The population has grown dramatically in the last four decades, reaching 51,000 in 1970, 206,087 in 1980, 450,406 in 1990, and 1,022,552 in 2000. H-1B visas provided a major route of legal access to the U.S. labor market in the 1990s for highly skilled individuals with job offers. Highly skilled Indians, especially those working in the computer industry, have been by far the largest beneficiaries of H-1B visas. In fiscal year 2001, Indian-born individuals received almost half of all H-1Bs issued, 58% of which were in computer-related fields. Moreover, survey evidence underlines the strong ethnic identification of the diaspora in America: 53% visit India at least once every two years, 97% watch Indian TV channels several times a week, 94% view Indian Internet sites several times a week, 92% read an Indian newspaper or magazine several times a week, and 90% have an Indian meal several times a week (Kapur, 2004).
[20] The cities are Bangalore, Delhi, Mumbai (Bombay) and Hyderabad.
[21] Of the 213,622 last names identified from the phone books 38,386 names appeared with a frequency of five or more. Of these, 13,418 matched a proprietary database of U.S. consumers (prepared by InfoUSA). One of the authors and an outside expert coded each of these names as: 1) extremely likely to be Indian, 2) extremely unlikely to be Indian, or 3) could be either.
[22] We do not expect the frequency of false positives in our name data to be large. In a random phone survey (N=2,256), 97 percent of the individuals with last names from our sample list responded "yes" to the question: "Are you of Indian origin?" (Kapur, 2004). Nor do we expect the frequency of false negatives to be large. Although we constructed our name set from the phone books of large metropolitan cities, the vast majority of Indian overseas migration to the United States is an urban phenomenon; the likelihood of an urban household in India having a family member in the U.S. is more than an order of magnitude greater than a rural household. A different problem arises when people change their last name after migration. This is more likely with Indian women due to marriage. However, even among second-generation Asian-Americans, Indian-American women are least likely to marry outside the ethnic group (62.5 percent marry within the ethnic group (Le, 2004). To the extent that noise exists in our name data, it will bias our result downwards.

the end of our study period due to the significant growth of patenting in India at that time. The average lag between the focal patent and the preceding cited patent is eight years.[23]

We compare various types of knowledge flows in terms of the degree to which they are mediated by co-location and diaspora effects. These comparisons include: 1) flows within versus across fields, 2) flows associated with more important versus less important inventions, 3) flows associated with returnees (individuals who patented an invention outside of India and then returned to patent within India) versus those who show no evidence of ever having left India, and 4) flows associated with future emigrants (individuals who patent in India and later patent abroad) versus others.

Table 1 shows that more than half (62 percent) of the focal-cited pairs represent within-field knowledge flows.[24] In terms of the importance of the focal invention, the mean number of citations received by focal patents is approximately three (table 1). However, we define important patents as those in the 90th percentile or above and as such delineate between focal patents receiving six or more citations versus others. In terms of "circulation," returnees invent approximately 2.5 percent of the focal patents in our data. Finally, in terms of future emigrants, individuals who later leave India invent approximately three percent of the focal patents in our data.

5. Results

Table 2 reports the OLS results for the full sample.[25] Focusing first on specification (1), we find evidence of a large and statistically significant co-location effect and a much smaller (though still statistically significant) diaspora effect. The difference between the two effects is also positive and statistically significant at the one percent level. The implied estimate of the proportionate co-location premium is (0.388 / 0.491) = 0.792, whereas the implied estimate of the proportionate diaspora premium is just (0.062 / 0.491) = 0.127. Interpreted through the lens of the model, the much larger co-location premium implies that the total access of India-residing inventors to

---

[23] Recall that the lag between focal and control patents is precisely the same, by construction.
[24] Again, the fraction of focal-control pairs that represent within-field knowledge flows is the same, by construction.
[25] We find identical conditional probability estimates using a logit specification. We concentrate on the OLS results due to their ease of interpretability.

knowledge is harmed by the absence of fellow Indian inventors. Furthermore, the very large co-location premium confirms the importance of localized knowledge flows.

The other specifications in table 2 allow for the co-location and diaspora effects to vary by the citation lag and also by whether the citation occurs within or across NBER two-digit classifications. By construction, we find no direct effect of lags and sub-category matches since we choose the control citations to match the actual citations based on both timing and technology class. It is possible, however, that our relationship variables and the lag and/or match variables interact. The results reported in specifications (2) through (4) suggest that the only important interaction is the one between co-location and the lag between the application dates of the citing and cited patent. However, the sign of this interaction is opposite to our prior, with the co-location effect being stronger for older cited patents.

Table 3 shows the results for our base specification for five of the six NBER one-digit classifications (we leave out the sixth category, "Others," due to the very small number of observations.) We find the previously identified pattern of large co-location effects and small diaspora effects in most categories. We also find relatively large diaspora effects for both Electrical & Electronic and Mechanical, but only the former is statistically significant at the 10 percent level. The single exception is Computers & Communications, which has no co-location effect. Perhaps India's international competitiveness in this sector, particularly information technology, involves drawing from a more global knowledge base, which is reflected in this finding.

In table 4, we examine whether vintage mediates the co-location and diaspora effects on knowledge flows. We take 1995 as the cutoff, but the results are not sensitive to this choice. The gap between the co-location and diaspora parameters is somewhat greater for the more recent focal patents (both because the co-location effect has risen and the diaspora effect has fallen), but the gap remains large even for older vintage focal patents.

As outlined in sections 2.2 through 2.4, the interpretation of these results is made more complicated by return migration, non-random selection, and heterogeneous valued innovations. To analyze the impact of returnees, the first two columns of table 5 split the sample into returnees and non-returnees. The first thing to note is that returnees account

for just 2.3 percent of our sample of focal patents. We are concerned that this may be an undercount of the true number of returnees since the identification of a returnee in the patent database requires that the individual previously patented abroad. Thus, we also examine if the returnees we have identified look any different from others in terms of the co-location and diaspora effects. The co-location effect is lower for returnees, suggesting a weaker link to other India-residing inventors,[26] but the diaspora is similar. Most importantly, the gap between the co-location and diaspora effects remains large.

Table 5 also allows us to explore the nature of selection for returnees and emigrants. Our measure of inventor "quality" is the number of forward citations to the invention.[27] The last row in the table gives the mean number of forward citations for the various sub-samples. Comparing returnees and non-returnees by this measure, we find that returnees are of higher quality on average, although the difference is relatively small. In contrast, we find evidence that emigrants are highly positively selected. For our sample of focal patents, the mean number of forward citations is a little more than two for those who do not subsequently go on to emigrate and just over 18 for those who do. Taken together, these results suggest that emigrants are positively selected and returnees are negatively selected from the resulting (select) diaspora pool. These findings on returnees and selection reinforce the inference based on the simple model: Inventor emigration harms knowledge access and domestic innovation.

In table 6, we examine whether invention quality mediates the co-location and diaspora effects on knowledge flows. It is well known that the distribution of patents in terms of their value is highly skewed (i.e., a small fraction of patents accounts for the majority of value). Following the literature, we use citations received by the focal patent as a proxy for patent value (Hall et al, 2006; Harhoff et al, 1999; Lanjouw and Schankerman, 1999).

The results are striking. Focusing on the 88th percentile and above (column 2), we see a somewhat lower co-location effect and a substantially higher diaspora effect,

---

[26] The difference is not statistically significant.
[27] We follow prior literature in using forward citations as a proxy for invention quality (Lanjouw and Schankerman, 1999).

compared to the rest of the sample (column 1) or the full sample (table 2).[28,29]   Further narrowing the sample to only the 93rd percentile and above (column 3), we see an even greater diaspora effect (almost 10 times the magnitude as that for the overall sample), and the co-location effect is no longer statistically significant. This diaspora-oriented result continues to hold when we cap the sample even further along the tail of the distribution to include only the 95th percentile and above.

These results are particularly salient since prior research has shown that the value of innovations increases nonlinearly with the number of citations (Trajtenberg, 1990).[30] Although the diaspora effect never exceeds unity (see section 2.1), it does come close when we look at the 95th percentile and above. Thus, even though our results indicate that a diaspora is not beneficial even when we limit our attention to high quality inventions, the rising diaspora effect does give us some pause in concluding that a disapora is never beneficial.

The small number of patents with even larger numbers of forward citations limited us from restricting attention to even higher quality patents. But the rising size of the disapora effect (both absolutely and relative to the co-location effect) as we restrict the sample to higher and higher quality focal patents raises the possibility that a diaspora is beneficial where the welfare effect of high quality inventions is large relative to the average invention.   In the conclusions section, we discuss further how this finding tempers our interpretation of the main findings reported in table 2.

6. Conclusions

This paper finds evidence of a large co-location premium for knowledge flows between Indian inventors associated with the "average" invention. It also finds evidence of a diaspora premium, but its size is much smaller (13 percent compared to 79 percent). Interpreted through the lens of a simple relationships-based model of knowledge access

---

[28] We also look at the number of forward citations occurring within specified time windows — three years, five years, and 10 years — and find similar results.

[29] The percentile cutoffs are not round numbers since they are dictated by the distribution of patents with certain numbers of citations received, which are discrete count values.

[30] An important caveat is that the evidence cited was taken from a single industry: computed tomography scanners.

and innovation, the difference between the effects is a sufficient condition for emigration to be harmful to the domestic economy.

While our model abstracts from the possibility of return and also of the non-random selection of emigrants and returnees, we find that returnees are quite rare in our sample of Indian innovators and that their knowledge-flow characteristics are similar to innovators who never left. Our data also indicate that emigrant innovators are a highly positively selected sub-sample of the Indian innovator population and that returnees are negatively selected from the emigrant stock. Thus, our basic conclusion is robust to returnee and selection effects.

However, we temper this conclusion drawn from our main results with our additional finding that domestic access to knowledge facilitated by the diaspora is relatively more important for high-value inventions. Given that the distribution of patents is highly skewed with respect to market value (and social value), the small fraction of patents for which the diaspora effect is particularly important might actually represent a large fraction of the productivity gains that result from innovation. Thus, to fully understand the effect of emigration on domestic innovation in a poor country, we need to better understand the relative value of very important innovations compared to others.

The central assumption of our model is that innovation output depends on access to knowledge. This focus on knowledge access allows us to incorporate a range of widely discussed, but hard to quantify, emigration-related impacts on the domestic economy, including the loss of local knowledge spillovers, the gains via diaspora connections, and the implications of knowledge-worker circulation. A limitation of our approach, however, is that we investigate innovation indirectly through our measures of knowledge access. The most important next step is to more directly measure how migration flows affect national innovation. We are currently exploring this question using detailed information on the career paths and productivity of mobile scientists.

Two additional issues in our paper need further investigation. One is whether skilled migration indeed entails a tradeoff between a smaller domestic stock of innovators and larger international networks. We have not addressed the possibility that the

domestic stock of innovation-producing talent might actually increase as a result of migration. This may occur because of two possible effects.

The first effect arises because the possibility of migration induces higher investments in education due to greater returns abroad (Beine, Docquier and Rapoport, 2001). If these additional investments in human capital are sufficiently large but only a fraction can actually leave, then it is possible that the country will end up with a greater stock of human capital. Although the basic "brain gain" story has plausibility given the clearly forward-looking nature of the demand for skills, doubts remain as to it its quantitative importance. Commander, Kangasniemi, and Winters (2004) as well as Schiff (2005) have argued, for example, that the highest-ability individuals will invest in skills regardless of the prospect of emigrating, but these individuals will be particularly prone to being recruited away when the prospect of emigration is enhanced. Thus, increased investments may provide the largest boost to the supply of more moderate-ability individuals.

A second effect could arise whereby increases in financial remittances may increase investments in education investment by easing binding liquidity constraints (Yang, 2006). Here, too, are contrary effects. For instance, if parents are absent from the household as a result of migration, there could be less parental inputs into education acquisition and greater work pressure on remaining household members. While Yang presents evidence from the Philippines where the former effect dominates, in Mexico's case the second factor appears to dominate (McKenzie and Rapoport (2006). This issue needs further investigation.

A second issue arises from the time period of the data (which ends in 2000 because of our use of forward citations) and the implications of right-censoring our data. The experiences of countries as varied as Ireland and South Korea and more recently of China point to the importance of changing domestic economic conditions in catalyzing the "brain bank" effect; diasporas have little impact on their home countries as long as their economies remain closed. India's economic liberalization and recent rapid growth rates are attracting some of its diaspora back in tandem with new multinational R&D investments. This effect may become apparent in the patent data over time.

We began this paper by noting the controversy between those who think the emigration of knowledge workers is good for the national economy as it expands the global technology pool and those who are concerned about the harm to national innovation. Overall, we find it unlikely that a poor country with a reasonably functioning economy and working hard to absorb the massive stock of available technology is actually better off if a large fraction of its scarce talent resides abroad. To this end, our main empirical results suggest that, in terms of access to knowledge, the localization effect outweighs the diaspora effect: Poor countries are better off if their highly skilled workers stay home.

However, we do not doubt that reallocation to higher productivity environments does increase global innovation and some of the fruits of that innovation surely do flow back to the poor-sending countries. Examples abound of emigrants from poor countries making great contributions to science. A few also do return to have transformative effects on their home countries, often as institution builders.[31] Furthermore, our findings suggest that access to knowledge provided by the diaspora is particularly important for the highest value inventions, those in the extreme tail of the distribution. How important is this minority share of inventions to the overall economy of poor countries? This question sets the stage for future research and contributions to this lively and important debate on the role of migration for economic growth in poor countries.

---

[31] For example, Dr. F.C. Kohli, who left India to study electrical engineering at Queen's University in Canada and then at MIT in the U.S. and who became an active member of the Institute of Electrical and Electronics Engineers headquartered in New York, returned to India to lead Tata Consultancy Services (India's largest IT firm during most of the last quarter of the 20th century). More importantly, from a social welfare perspective, many cite him more generally as the "father of the Indian software industry." He credits much of his ability to accomplish what he did to his education and social network in North America. (The IT Revolution in India, F.C. Kohli: Selected Speeches and Writings, F.C. Kohli, 2005).

References

Agrawal, Ajay, Iain Cockburn, and John McHale (2006), "Gone but Not Forgotten: Labor Flows, Knowledge Spillovers, and Enduring Social Capital," *Journal of Economic Geography*, 6(5): 571-591.

Agrawal, Ajay, Devesh Kapur, and John McHale (2007), "Birds of a Feather—Better Together? Exploring the Optimal Spatial Distribution of Ethnic Inventors," NBER Working Paper 12823.

Alcacer, Juan and Michelle Gittelman (2006), "How Do I Know What You Know? Patent Examiners and the Generation of Patent Citations," *Review of Economics and Statistics*, 88 (4), 774-9.

Beine M., F. Docquier, and H. Rapoport, (2001), "Brain drain and economic growth: theory and evidence." *Journal of Development Economics*, Volume 64, Number 1, pp. 275-289.

Basu, Susanto and David Weil, (1998), "Appropriate Technology and Growth," *The Quarterly Journal of Economics*, 113(4):1025-1054.

Caselli, Francesco and Wilbur John Coleman II, (2001), "Cross-Country Technology Diffusion: The Case of Computers," *The American Economic Review*, Papers and Proceedings, 91(2): 328-335.

Cockburn, Iain, Sam Kortum, and Scott Stern, (2002), "Are All Patent Examiners Equal? The Impact of Examiner Characteristics on Patent Statistics and Litigation Outcomes," National Bureau of Economic Research, Working Paper 8980.

Cohen, Wesley and Daniel Levinthal, (1989), "Innovation and Learning: The Two Faces of R&D," *The Economic Journal*, 99(397): 569-596.

Commander, Simon, Mari Kangasniemi, and L. Alan Winters, (2004), "The Brain Drain: Curse or Boon? A Survey of the Literature," in *Challenges to Globalization: Analyzing the Economics*, Robert E. Baldwin and L. Alan Winters, eds., University of Chicago Press, pp. 235–78.

Docquier, Frederic and Abdeslam Marfouk, (2005), "International Migration by Educational Attainment," in *International Migration, Remittances, and the Brain Drain*, Caglar Ozden and Maurice Schiff, eds., Washington D.C. and New York: The World Bank and Palgrave Macmillan.

Dumont, Jean Christophe and Georges Lemaitre, (2005), "Counting Immigrants and Expatriates in OECD Countries: A New Perspective," OECD Social, Employment, and Migration Working Papers No. 25.

Hall, Bronwyn H., Adam Jaffe, and Manuel Trajtenberg, (2006), "Market Value and Patent Citations," mimeo, University of California, Berkeley

Harhoff, Dietmar, Francis Narin, F.M. Scherer, and Katrin Vopel, (1999), "Citation Frequency and the Value of Patented Inventions," *The Review of Economics and Statistics*, 81 (3): 511-515.

Hegde, Deepak and Bhaven Sampat, (2007) "Examiner Citations, Applicant Citations, and the Private Value of Patents," working paper, Columbia University.

Jaffe, Adam and Manuel Trajtenberg, (2002), *Patents, Citations, and Innovations: A Window on the Knowledge Economy*, Cambridge: The MIT Press.

Jaffe, Adam, Manuel Trajtenberg, and Rebecca Henderson, (1993), "Geographic Localization of Knowledge Flows as Evidenced by Patent Citations," *Quarterly Journal of Economics*, CVIII: 577-598.

Kapur, Devesh and John McHale, (2005), *Give Us your Best and Brightest: The Global Hunt for Talent and its Impact on the Developing World*, Washington D.C.: Center for Global Development/Brookings Institution Press.

Kapur, Devesh, (2004), "Survey of Indian Americans in the United States (SAIUS)," mimeo, Harvard University.

Keller, Wolfgang, (2002), "Geographic Localization and International Technology Diffusion," *The American Economic Review*, 92(1): 120-142.

Kerr, William, (2008), "Ethnic Scientific Communities and International Technology Diffusion," Economics and Statistics (forthcoming).

Klenow, Peter and Andres Rodriguez-Clare, (2004), "Externalities and Growth," NBER Working Paper 11009.

Kuhn, Peter and Carol McAusland, (2006), "The International Migration of Knowledge Workers: When is Brain Drain Beneficial," NBER Working Paper 12761.

Lanjouw, Jean O. and Mark A. Schankerman, (1999), "The Quality of Ideas: Measuring Innovation with Multiple Indicators," NBER Working Paper No. W7345.

Le, C. N., (2004), "Socioeconomic Statistics & Demographics," http://www.asian-nation.org/demographics.shtml, accessed 22 July 2004. Asian-Nation: The Landscape of Asian America.

McKenzie, David and Hillel Rapoport, (2006), "Can migration reduce educational attainment? Evidence from Mexico," World Bank Policy Research Working Paper No. 3952.

Nanda, Ramana and Tarun Khanna, (2007), "Diasporas and Domestic Entrepreneurs: Evidence from the Indian Software Industry," mimeo, Harvard Business School.

Romer, Paul, (1990), "Endogenous Technical Change," *The Journal of Political Economy*, 98(5) Part 2: S71-S102.

Saxenian, AnnaLee, (2005), "From Brain Drain to Brain Circulation: Transnational Communities and Regional Upgrading in India and China," *Studies in Comparative International Development*, Vol. 20 (2), pp. 35-61.

Saxenian, AnnaLee, (2006), *The New Argonauts: Regional Advantage in a Global Economy*, Cambridge, MA: Harvard University Press.

Schiff, Maurice, (2005), "Brain Gain: Claims about its Size and Impact on Welfare are Greatly Exaggerated," in *International Migration, Remittances, and the Brain Drain*, Çaglar Özden and Maurice Schiff, eds., Washington: World Bank and Palgrave Macmillan, pp. 201–26.

Sukhatme, S and I. Mahadevan, (1987), *Pilot study on magnitude and nature of the brain drain of graduates of the Indian Institute of Technology, Bombay*. Bombay: Indian Institute of Technology.

Thomson, Peter and Melanie Fox-Kean, (2005), "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment," *The American Economic Review*, 95(1): 450-460.

Trajtenberg, Manuel, (1990), "A Penny for Your Quotes: Patent Citations and the Value of Innovations," *RAND Journal of Economics*, 21 (1): 172-187.

Yang, Dean, (2006), "International Migration, Remittances, and Household Investment: Evidence from Philippine Migrants' Exchange Rate Shocks," NBER Working Paper 12325.

World Bank (2008), *Global Economic Prospects 2008: Technology Diffusion in the Developing World*, Washington D.C.: The World Bank.

Table 1
Descriptive Statistics

	Mean	Std. Dev.	Min.	Max.
Cited patent is co-located with focal	.0192227	.1373141	0	1
Cited patent is by diaspora member	.0389706	.1935351	0	1
Application year of focal patent	1997.285	3.506532	1981	2000
Lag*	7.933613	5.9179	0	27
Within-field knowledge flow**	.6241597	.4843646	0	1
Importance of focal patent***	2.881092	7.362347	0	112
Focal patent is by Returnee	.0245798	.1548489	0	1
Focal patent is by Future Emigrant	.0327731	.1780516	0	1

N = 9520 observations
* Years between the application date of the focal versus the cited patent
** Probability the focal and cited patent are both assigned to the same two-digit NBER technology subcategory
*** Number of citations *received* by the focal patent

Table 2
OLS Estimates of the KFPF

Dependent Variable = *Citation*

	(1)	(2)	(3)	(4)
Co-location	0.388***	0.330***	0.356***	0.303***
	(0.024)	(0.036)	(0.048)	(0.054)
Diaspora	0.062**	0.037	0.062	0.036
	(0.031)	(0.041)	(0.048)	(0.057)
Co-location×Lag		0.019***		0.019***
		(0.007)		(0.007)
Diaspora×Lag		0.004		0.004
		(0.006)		(0.006)
Co-location×Sub-Category Match			0.043	0.038
			(0.054)	(0.052)
Diaspora×Sub-Category Match			0.000	0.002
			(0.055)	(0.056)
Constant	0.490***	0.490***	0.490***	0.490***
	(0.002)	(0.002)	(0.002)	(0.002)
R^2	0.012	0.013	0.012	0.013
# Observations (actual & control citations)	9,520	9,520	9,520	9,520
# Clusters (focal patents)	793	793	793	793

Standard errors are in parentheses
Standard errors are robust to focal patent clustering effects
* Significance at 10 percent level
** Significance at 5 percent level
***Significance at 1 percent level

Table 3
OLS Estimates of the KFPF By NBER One-Digit Code

Dependent Variable = *Citation*

	(1) Chemical	(2) Computers &Comm.	(3) Drugs& Medical	(4) Electrical& Electronic	(5) Mechanical
Co-location	0.418***	0.017	0.433***	0.491***	0.511***
	(0.035)	(0.153)	(0.031)	(0.012)	(0.005)
Diaspora	0.033	0.078	0.048	0.169*	0.178
	(0.059)	(0.055)	(0.062)	(0.097)	(0.160)
Constant	0.489***	0.495***	0.485***	0.491***	0.489***
	(0.002)	(0.004)	(0.003)	(0.003)	(0.005)
R^2	0.016	0.002	0.022	0.011	0.019
# Observations (actual & control citations)	2,826	1,682	2,734	1,170	282
# Clusters (focal patents)	389	105	298	107	77

Standard errors are in parentheses
Standard errors are robust to focal patent clustering effects
*   Significance at 10 percent level
** Significance at 5 percent level
***Significance at 1 percent level

Table 4
OLS Estimates of the KFPF By "Vintage"

Dependent Variable = *Citation*

	Vintage	
	Recent	Early
	(1)	(2)
	Application Year for Focal Patent > 1995	Application Year for Focal Patent ≤ 1995
Co-location	0.394***	0.364***
	(0.026)	(0.056)
Diaspora	0.058*	0.099
	(0.034)	(0.073)
Constant	0.490***	0.492***
	(0.002)	(0.002
R^2	0.013	0.010
# Observations (actual & control citations)	7,524	1,996
# Clusters (focal patents)	588	205

Standard errors are in parentheses
Standard errors are robust to focal patent clustering effects
* Significance at 10 percent level
** Significance at 5 percent level
***Significance at 1 percent level

Table 5
OLS Estimates of the KFPF By Returnee / Future Emigrant Status

Dependent Variable = *Citation*

	(1) Non-Returnees	(2) Returnees[a]	(3) Non-Future Emigrants	(4) Future Emigrants[b]
Co-location	0.391***	0.296**	0.389***	0.369**
	(0.024)	(0.143)	(0.024)	(0.171)
Diaspora	0.062*	0.077	0.060*	0.262
	(0.032)	(0.115)	(0.031)	(0.225)
Constant	0.490***	0.485***	0.490***	0.488***
	(0.002)	(0.134)	(0.002)	(0.009)
R^2	0.012	0.013	0.012	0.015
# Observations (actual & control citations)	9,286	234	9,208	312
# Clusters (focal patents)	775	18	771	22
Mean forward cites to focal patent	2.883	3.598	2.366	18.083

Standard errors are in parentheses
Standard errors are robust to focal patent clustering effects
* Significance at 10 percent level
** Significance at 5 percent level
***Significance at 1 percent level
a/ Returnees are identified as inventors who are observed to have previously patented outside of India
b/ Future emigrants are identified as inventors who are subsequently observed to patent outside of India at a later date

Table 6
OLS Estimates of the KFPF By "Quality" of Focal Patents

Dependent Variable = *Citation*

	Quality			
	Slightly Below Average	High	Very High	Extremely High
	(1)	(2)	(3)	(4)
	Total Cites to Focal Patent < Less than 6 (Below 88th percentile)	Total Cites to Focal Patent ≥ 6 (88th percentile and above)	Total Cites to Focal Patent ≥ 9 (93rd percentile and above)	Total Cites to Focal Patent ≥ 12 (95th percentile and above)
Co-location	0.396***	0.307***	0.198	0.093
	(0.024)	(0.105)	(0.156)	(0.199)
Diaspora	0.053	0.229**	0.479***	0.474***
	(0.032)	(0.121)	(0.034)	(0.074)
Constant	0.490***	0.491***	0.492***	0.495***
	(0.002)	(0.003)	(0.003)	(0.003)
R^2	0.013	0.010	0.013	0.009
# Observations (actual & control citations)	8,448	1,072	650	372
# Clusters (focal patents)	699	94	59	38

Standard errors are in parentheses
Standard errors are robust to focal patent clustering effects
* Significance at 10 percent level
** Significance at 5 percent level
***Significance at 1 percent level