Volume Title: The Roles of Immigrants and Foreign Students in US Science, Innovation, and Entrepreneurship

Volume Authors/Editors: Ina Ganguli, Shulamit Kahn, Megan MacGarvie, editors

Volume Publisher: University of Chicago Press

Volume ISBNs: 978-0-226-69562-4 (cloth); 978-0-226-69576-1 (electronic)

Volume URL: https://www.nber.org/books-and-chapters/roles-immigrants-and-foreign-students-us-science-innovation-and-entrepreneurship

Conference Date: April 27, 2018

Publication Date: February 2020

Chapter Title: Return Migrants' Self-Selection: Evidence for Indian Inventors

Chapter Author(s): Stefano Breschi, Francesco Lissoni, Ernest Miguelez

Chapter URL: https://www.nber.org/books-and-chapters/roles-immigrants-and-foreign-students-us-science-innovation-and-entrepreneurship/return-migrants-self-selection-evidence-indian-inventors

Chapter pages in book: (p. 17 – 48)

# 1

# Return Migrants' Self-Selection
## Evidence for Indian Inventors

Stefano Breschi, Francesco Lissoni, and Ernest Miguelez

### 1.1 Introduction

Return migration represents an important share of present-day total cross-border population flows. In 2008, the International Migration Outlook of the Organisation for Economic Co-operation and Development (OECD), based on indirect estimation methods, suggested that 20 percent to 50 percent of adult immigrants to advanced countries might leave within five years after their arrival, albeit with much variation due to heterogeneity of sending-receiving country pairs, years of entry, and the definition of "return migrant" itself (OECD 2008).[1]

Stefano Breschi is professor of applied economics at Bocconi University in Milan and director of the Invernizzi Center for Research on Innovation, Organization, Strategy and Entrepreneurship (ICRIOS).

Francesco Lissoni is professor of economics at GREThA, University of Bordeaux, and an affiliate of the Invernizzi Center for Research on Innovation, Organization, Strategy and Entrepreneurship (ICRIOS) at Bocconi University in Milan.

Ernest Miguelez is a research fellow at the French National Centre for Scientific Research (CNRS) attached to GREThA, University of Bordeaux, and a research affiliate at the Regional Quantitative Analysis Research Group (AQR-IREA) at Univeristat de Barcelona.

We gratefully acknowledge financial support from NBER and the French National Research Agency (TKC project—reference: ANR-17-CE26-0016). For acknowledgments, sources of research support, and disclosure of the authors' material financial relationships, if any, please see https://www.nber.org/chapters/c14104.ack.

1. In what follows, unless otherwise stated, we will adopt Dustmann and Weiss's (2007) definition of return migrants as those who settle back in their home country by their own choice after having spent several years abroad. This echoes the definition provided for statistical purposes by the United States Statistical Division of "persons returning to their country of citizenship after having been international migrants (whether short-term or long-term) in another country and who are intending to stay in their own country for at least a year" (UN 1998; as quoted by OECD 2008) but hides more complex migration patterns, such as circular and repeat migration (Constant and Zimmermann 2016).

Such high rates also affect high-skilled (highly educated) migrants. Based on a large sample of foreign recipients of a US doctorate in science and engineering, Finn (2014) calculates an average return rate—five years after graduation—of about 30 percent, with country-specific figures ranging from less than 10 percent for India and China to over 40 percent for Western European countries. In addition, evidence from questionnaires on return intentions suggests, for migrants to the United States and Germany, a U-shaped relationship between years of schooling and return rates (Dustmann and Görlach 2016)—that is, a self-selection of return migrants with respect to very low and very high educational levels. OECD (2008) estimates on actual returns conform to this pattern, especially for the United States.

High-skilled return migration is especially relevant for innovation studies. From the viewpoint of migrants' home countries, returnee scientists, engineers, and other professionals can play a role in knowledge diffusion and new business creation.[2] In this respect, high-skill return migration can act as a potential compensating mechanism for the "brain drain" suffered by sending countries (Dustmann, Fadlon, and Weiss 2011; Gibson and McKenzie 2011).

As for host countries, their policy-makers, higher education institutions, and knowledge-intensive firms fret not only about attracting but also about retaining the "best and brightest" among foreign workers and students (Hawthorne 2018; Teitelbaum 2014; Wadhwa et al. 2009). This begs the question of whether returnees self-select positively not only with respect to their immediately observable skills, such as educational level, but also with respect to harder-to-observe skills, such as inventiveness, creativity, or entrepreneurial propensity, conditional on education.

More generally, the issue of skill-based self-selection of return migrants plays a crucial role in economic theories of migration as a lifetime investment with important implications for the expected economic and social assimilation of both permanent and temporary migrants (Borjas and Bratsberg 1996; Dustmann and Görlach 2016).

Despite its relevance, return migration is an understudied topic due to a lack of data. National authorities commonly register the inflows of foreign-born and foreign nationals but not their outflows, which makes it nearly impossible to know precisely how many immigrants later leave the country and when, let alone their individual characteristics. Quantitative research then relies on longitudinal surveys or on complex manipulation of administrative panel data (Dustmann and Görlach 2016).

Most surveys, however, concern specific, often low-skilled migrant groups

2. On entrepreneurs, see Nanda and Khanna (2010), Filatotchev et al. (2011), and Luo, Lovely, and Popp (2013). On scientists, see Kahn and MacGarvie (2016), Jonkers and Cruz-Castro (2013), Trippl (2013), and Gibson and McKenzie (2014). On managers, see Nanda and Khanna (2010) and Choudhury (2016).

(such as the *gastarbaiters* of the 1960s and 1970s in the much-used German Socio-economic Panel) and/or focus on labor market determinants of return migration, such as unemployment (Bijwaard, Schluter, and Wahba 2014). Notable but rare exceptions concern academic scientists, whose return rates and individual characteristics can be obtained by combining archival and bibliometric data sources, as in Gaulé (2014) and Kahn and MacGarvie (2016).

In a recent assessment of the emerging literature on migration and innovation, Kerr (2017) states that we know very little about return migration of workers engaged in innovation and entrepreneurship except that it is rapidly growing in importance and that "clever data work to . . . quantify [it] would be most welcome" (Kerr 2017, 212). This chapter answers the call. Based on an ambitious data-linkage project joining patent data and inventors' biographical information from a web-based, professionally oriented social network, we build a large sample of US immigrant inventors of Indian origin specializing in the information and communication technologies (ICT) sector. This is a social group that both figures prominently in the recent debate on temporary work migration to the United States (most notably on the use of H-1B visas; Kerr and Lincoln 2010) and contributes significantly to international student mobility (OECD 2017).

Our data-mining strategy allows us to identify only migrants entering the United States via work and education channels, most likely associated with temporary visas. Yet we do not consider it a weak point due to two well-established stylized facts:

1. The overwhelming importance of temporary channels as a source for high-skilled immigration into the United States via the transformation of both temporary work and student visas into permanent ones (in contrast with countries such as Australia and Canada, where permanent visas for the highly skilled are more easily obtained upon entry; Koslowski 2018)
2. The remarkable innovation impact of migrant scientists and engineers entering the United States with work and student visas, as opposed to those entering through the channel of family reunions, as documented by Hunt (2013)

While subject to a number of limitations, our data set allows us to trace return migration from the United States with a degree of precision comparable to survey data, but on a much larger scale and with original information on its possible determinants. For each individual in the data set, we estimate the year of entry, the likely entry channel (work or education), and the permanence spell up to either the return to India or 2016 (right-censoring year). By means of survival analysis, we provide estimates of the probability of return migration as a function of the conditions at migration (age, education, patenting record, migration motives, and migration cohort) as well as some activities undertaken while abroad (education and patenting).

Our results, albeit exploratory, find rather different patterns for work and education migrants. Considering the former, we find that Indian inventors' return risk is positively associated with their age and education at migration, as well as their propensity to patent while in the US. As for education migrants, the return risk correlates negatively with the education level they attain. We also find some evidence of negative (positive) time-dependence for work (education) return migrants, which we interpret as indicative of negative (positive) self-selection with respect to unobservable skills acquired in the host country.

We proceed as follows. In section 1.2, we present in a rather succinct way our database-building strategy (more details in the appendix: http://www.nber.org/data-appendix/c14104/appendix.pdf), introduce our own definitions of *migrant* and *return migrant*, and propose some descriptive evidence. When necessary, we discuss some conceptual and methodological issues concerning the definition of *return migrant*. In section 1.3, we present our model specification and discuss how it serves the purpose of investigating skill-based self-selection in return migration. In section 1.4, we perform the related econometric exercise and comment on the results. Section 1.5 concludes with a special focus on further research plans and some tentative policy implications.

## 1.2 Data: Methodology and Descriptive Statistics

### 1.2.1 Methodology

Our data set originates from an ambitious data-linkage project between patent and inventor data gathered from PatentsView (http://www.patentsview.org/web/) and biographical information extracted from a large number of LinkedIn profiles. PatentsView is a data repository recently made available by the United States Patent and Trademark Office (USPTO), which provides, among other things, disambiguated data on all the inventors of patents granted by the USPTO from 1975 onward, irrespective of their country of residence. LinkedIn, a well-known professional-oriented social network, represents an unparalleled source of information on the international mobility of individuals, as the members' public profiles include information on names and (possibly) locations of their educational institutions and employers, along with graduation and recruitment years (Ge, Huang, and Png 2016; Zagheni and Weber 2015).

As a pilot project, we focus on a subset of high-skilled migrants in the United States—namely, Indian inventors with ICT patents. This is a distinctive social group due to both its inventive contribution (Kerr and Lincoln 2010; Breschi, Lissoni, and Miguelez 2017) and its implication in two important temporary migration channels—namely, highly qualified temporary work (most notably through the H-1B visa system; Kerr et al. 2015; Kapur and McHale 2005) and education (Finn 2014; Kapur and McHale 2005). It is also a highly represented group on LinkedIn, which in 2016 registered well over 100 million members in the US and over 30 million in India, with the two countries standing at the top of LinkedIn world rankings for both membership and traffic.[3]

We extracted from PatentsView all the patents granted to the 179 largest US public firms in the ICT industry from 1975 to 2016 and the relative inventors for a total of 262,847 distinct individuals.[4] We then proceeded to the ethnic analysis of such inventors' names and surnames based on Global Name Recognition, a name search technology produced by IBM (from now on, IBM-GNR) and adapted to our purposes by Breschi et al. (2017). This allowed us to identify inventors of presumed Indian origin (from now on, Indian inventors) for a total of 24,017 individuals representing 9.1 percent of all inventors employed by the companies in our sample. Each Indian-named inventor was then matched to one LinkedIn profile based on name and company matching with extensive manual checking. This exercise yielded 10,839 inventors with valid LinkedIn accounts (around 45 percent of the original sample). For details, see sections C and D of the appendix.

We then proceeded to codify three major sets of variables, concerning education, employment, and patent records. On that basis, we also estimated the inventors' years of birth as well as their migrant, nonmigrant, and return migrant status.

We coded information on education according to the 2011 version of UNESCO's International Standard Classification of Education (ISCED) for educational levels from 3 (upper secondary) to 8 (doctoral or equivalent).[5] After jointly treating ISCED levels 5 and 6 (respectively, short-cycle tertiary and bachelor's) and distinguishing between master's of arts and/or science and MBAs, we ended up with the following classification: upper secondary education, bachelor's, master's, MBA, and PhD, plus a residual unclassified category. We then geolocalized as many education institutions as possible at the country level by means of Google Maps and obtained at least one geolocalization per inventor. (For full details, see section E of the appendix.)

As for employment, we recorded the start and end years of each related employment spell as well as the employer's name. We geolocalized the latter, at the country level, only on the basis of the information provided by the

3. Unofficial statistics are from https://www.statista.com/statistics/272783/linkedins-membership-worldwide-by-country/ (last visited April 2018).

4. The definition of an ICT industry follows the one provided by the OECD (https://www.oecd.org/sti/ieconomy/1835738.pdf). More details in section B of the appendix: https://www.nber.org/data-appendix/c14104/appendix.pdf.

5. See http://ec.europa.eu/eurostat/statistics-explained/index.php/International_Standard_Classification_of_Education_(ISCED) (last visited March 2018).

LinkedIn profile, with no further attempt to use GoogleMaps, which would prove useless for multinationals with several branches and affiliates in multiple countries. Thus our estimates on migration and return migration for work reasons have to be considered extremely conservative. In section F of the appendix, we discuss some possible ways to improve them by capturing more return moves based on a more sophisticated treatment of LinkedIn information.

As for the inventive activities of each inventor, we geolocalized them at the country level on the basis of the inventor's address as reported on his or her various patents and dated them on the basis of the patent's priority year (De Rassenfosse et al. 2013).[6] Based on the unique inventor ID provided by PatentsView, we could then calculate the number of patents signed by each inventor each year either in India or abroad.

Coming to the inventor's year of birth, our preferred option was to estimate it on the basis of education information, with reference to the lowest-level education achievement among those reported in the LinkedIn profile, the year of completion, and the presumed age at start (see section G in the appendix; see also Gaulé 2014). For the inventors whose profiles did not report any information on the timing or level of education, we estimated the year of birth based on the average age of the other inventors in the same patent cohort (i.e., the inventors who filed their first patents in the same year). In most cases, the age so calculated is around 32, which is close to general estimates by Jones (2009).

After dropping the inventors whose LinkedIn profiles did not provide sufficient information for estimating either the educational level or the year of birth, 8,982 observations remained (see table 1A.5 in the appendix). For these, we estimated the accuracy of our PatentsView-LinkedIn match based on around 1,000 LinkedIn profiles of Indian ICT professionals that report patent information. Based on around 800 "true positives" (successful matches of a LinkedIn profile to an inventor in PatentsView with coherent patent information) and 30 "false positives" (successful matches, but with discordant patent information), we calculated a 96.4 percent precision rate and a 77 percent recall rate. The high precision rate suggests that the education, employment, and age information in our data set are rather accurate (i.e., it is unlikely that they refer to the wrong inventor). However, the low recall rate suggests that our sample possibly suffers from truncation problems, to the extent that the excluded inventors may share some characteristics associated with the phenomenon of our interest (return migration).[7]

6. We obtain the priority year of the patent from its priority date—namely, the date of filing of the first USPTO application or, in case of patents extended to the United States but first filed abroad, the first application worldwide.

7. In section 1 of the appendix, we further investigate the properties of our sample of 8,982 inventors. We first compare their patent records to those of other Indian-named inventors in

We finally proceeded to identify migrant and return migrant inventors to/from the United States, as described in detail by figures 1A.6 and 1A.7 in the appendix. We proceeded by elimination, first dropping from our sample all the inventors without any education, employment, or patenting records within India, who may include second-generation migrants or members of the historical Indian diaspora in the former British Commonwealth. Second, we dropped all those we consider nonmigrants—namely, the inventors without any education, employment, or patenting records outside India. Among the remaining inventors, we considered as "education migrants" all the inventors whose LinkedIn profiles report enrolment in foreign higher education institutions as the first events in their lives taking place outside India and occurring earlier than any patenting activity abroad. Similarly, we considered as "work migrants" all the inventors whose LinkedIn profiles report recruitment by foreign-based companies as the first events taking place outside India and/or who have at least one patent abroad dating to before any enrolment in foreign higher education institutions.

Finally, we restricted our attention to migrants whose first moves outside India occurred in the United States. This left us with 3,943 "education migrants" and 1,589 "work migrants" from India to the United States for a total of 5,532 individuals. For the former, we considered as the migration year the starting year of the first education program undertaken in the United States. For the latter, we similarly defined the migration year as the beginning year of the first working spell in the United States or, alternatively, the priority year of the first patent. When distinguishing between "work" and "education" migrant inventors, it is important to keep in mind that the distinction refers only to the individuals' condition at migration time. Nothing impedes a work migrant from entering a master's or PhD program in the United States or impedes an education migrant from starting to work there. Indeed, the first case is rather frequent, and the second is very frequent.

Coming to return migration, we record as a returnee every migrant report-

our initial data set and find no significant differences for what concerns the average number of patents granted, conditional on the year of the first patent. However, based on the year of the first patent, inventors with a LinkedIn profile appear to be younger than those without one. Second, we compare the inventors for whom we found a LinkedIn profile (whether complete or not) to all others and find that inventors who patent exclusively in India have a significantly higher probability of being matched with a LinkedIn profile than inventors who patent exclusively in the US or both in India and in the US. These diagnostics suggest that, based on our data, we may underestimate migration from India, especially for more recent calendar years, due to the relative overrepresentation of India-based inventors versus US-only- and US-plus-India-based ones, the former group being more likely to contain nonmigrants and the latter more likely to contain migrants. Reasoning along similar lines, we may risk overestimating return migration, since the propensity to have a LinkedIn profile is higher for US-plus-India-based inventors than for US-only-based inventors. More generally, the younger the inventors, the more representative our sample.

ing an employment or a patent in India after having moved to the United States. We do not record return events related to further education in India, but we suspect these to be very few. However, we record employment in higher education. As for the return year, this coincides with either the start of the relevant employment spell or the priority year of the relevant patent. All migrants for whom we do not observe any return event are considered as still living in the United States in 2016, our final year of observation. For the sake of simplicity, at this stage of our research, we do not code any event following the first return to India. Similarly, we ignore any move from the United States to a country other than India. For example, we will treat an Indian student in the United States who leaves for the United Kingdom after graduation as if he or she was staying in the United States. This implies that we ignore circular migration. A cursory look at our data, however, reveals very few instances of this type.

Albeit imperfect, our coding of return events (and, in consequence, permanence abroad) does not compare unfavorably with similar coding one can find in the literature. Borjas's (1989) classic study based on the 1972–1978 Survey of Scientists and Engineers simply recorded as returnees all foreign respondents to the 1972 questionnaire who had left the sample by 1978. Gaulé (2014), who relied on several editions of the Directory of Graduate Research of the American Chemical Society from 1993 to 2007, first identified as potential returnees all foreign faculty and postdocs who appear at least once in the directory and then disappear. He then looked manually in bibliographic and web resources for information on the likely motives for the disappearance (to distinguish between return to the home country while not ceasing the academic career and moves to industry or third countries, and deaths). To our knowledge, the only accurate survey of return moves is provided by Gibson and McKenzie (2014), but for a very small sample.[8]

Even much-used resources for studying low-skill return emigration, such as the German Socio-economic Panel (GSOEP), are far from faultless. In some cases, they resort to measuring return intentions rather than actual moves.[9]

8. Gibson and McKenzie (2014) survey around 800 high-achieving secondary school graduates from New Zealand, Tonga, and Papua New Guinea, 200 of whom undertook academic careers. In this subgroup, 78 percent moved abroad, with a 25 to 30 percent return rate.

9. As explained by Bönisch, Gaffert, and Wilde (2013), the basic information on return migration provided by GSOEP consists of nonresponse items accompanied by the "moved abroad" motivation. This amounts to underreporting, as observed by Constant and Massey (2002), who find that a much larger number of individuals in the panel leave for one or more years without providing a motivation explicitly related to a move back home and hence resort to code as returnees all absentees for three or more years. Kirdar (2009) reports similar problems for more recent issues of the survey. As in many surveys of low-skilled migrants, the GSOEP collects information on return intentions. Similar information for the highly skilled is collected by Baruffaldi and Landoni (2012). While useful for testing theoretical models of temporary migration, return intentions may be different from de facto choices. For example, the 2000–2013

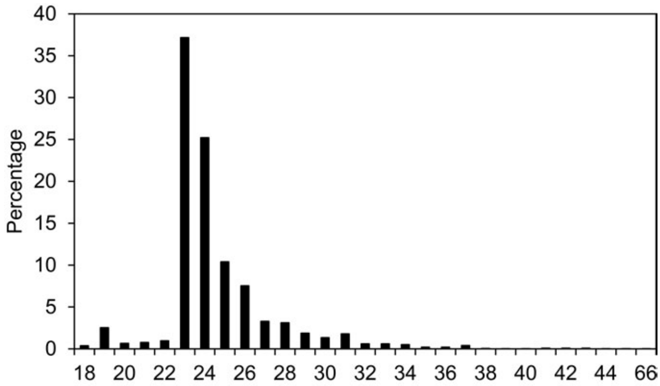


**Fig. 1.1a Estimated age at migration, education channel (percentage distribution of all education migrants to the United States)**

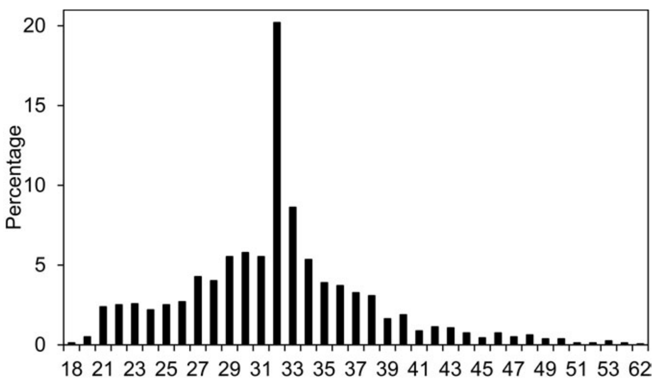


**Fig. 1.1b Estimated age at migration of work migrants (percentage distribution of work migrants to the United States)**

### 1.2.2 Descriptive Statistics

In what follows, we produce a number of descriptive statistics that serve the dual purpose of checking the information contents and quality of our data and providing some basic evidence on the phenomenon under study.

Figures 1.1a and 1.1b report the distribution of the age at migration for education and work education migrants, respectively. We notice that the overwhelming majority of the former move to the US at 23 or 24 years of

trends for return migration and return intentions calculated by Finn (2014) for a longitudinal cross-section of foreign doctoral graduates in the US are markedly different.

age, which is compatible with the age for starting a master's course or possibly a PhD. The very sparse observations for ages less than 19 are due to either errors in our calculation of migrants' years of birth or the very few Indian migrants who move to the US for bachelor's studies. As for the very few apparently moving at older ages, especially over 30, they may be mature postgraduate students or professionals taking MBA courses. Figure 1A.12a in the appendix reports the distribution of age at migration for all Indian immigrants in science, technology, engineering, and mathematics (STEM) occupations for the migration cohorts 1990s and 2000s based on data from the American Community Surveys 2000 and 2010 (pooled samples; IPUMS US). Figure 1A.12b reports the same calculations for the subset of STEM-employed Indian immigrants with education at the college, master's, or PhD level. The modal values are, respectively, 24 and 25, which is slightly higher than for our educational sample. Moreover, the age at migration distribution taken from IPUMS data is flatter. Again, as our sample is composed of educational migrants, this concentration in early ages of migration is expected.

For comparison purposes, we also look at the number of H-1B petition filings by age for the years 2007–2017 from the US Citizenship and Immigration Services (figure 1A.13 in the appendix). The most numerous group is the one at ages 25–34, followed by the 35–44 group. Again, this is slightly different from our education sample. In this regard, figure 1.1b shows the distribution of the age at migration as way more skewed to the right. However, the figure shows a high peak at 32, which also differs from figures 1A.12 and 1A.13 and seems to be a statistical artifact that results from the inclusion in this migrants category of many inventors with two characteristics. First, for want of better information, we estimate their ages based on the priority year of their first patents. Second, they appear on these patents with US addresses, and this is the earliest evidence we have of their migration. Yet we notice that the age distribution is rather symmetric around 32. This is compatible with migrants in this group moving abroad after completing their education in India and starting their careers there, as happens with many H-1B visa holders, as well as being employees of Indian firms temporarily detached to the United States. When excluding from the work migrants all inventors whose ages were determined by the year of the first patent, the shape of the distribution does not change much, since the modal value remains at 32 and the symmetry is preserved (figure 1A.15 in the appendix). However, the percentage of people migrating at 32 goes from 20.2 percent to 13 percent in figure 1A.15, which is significant. In any case, we should be cautious when interpreting the estimates of the effect of age at migration on return decisions (for work migrants).

Table 1.1 provides a breakdown of our data set by migration motives and cohorts (decades during which migration occurred). Two features emerge.

**Table 1.1** **Migrants to the United States by cohort and channel**

| Channel | 1960 | 1970 | 1980 | 1990 | 2000 | 2010 | Total |
|---|---|---|---|---|---|---|---|
| Education | 19 | 102 | 697 | 1,739 | 1,315 | 71 | 3,943 |
| % column | 100 | 95.3 | 95.2 | 85.9 | 56.3 | 22.8 | 71.3 |
| Work | 0 | 5 | 35 | 286 | 1,022 | 241 | 1,589 |
| % column | 0.0 | 4.7 | 4.8 | 14.1 | 43.7 | 77.2 | 28.7 |
| All channels | 19 | 107 | 732 | 2,025 | 2,337 | 312 | 5,532 |
| % column | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| % row | 0.3 | 1.9 | 13.2 | 36.6 | 42.2 | 5.6 | 100 |

First, most migrants in our sample belong to the 1990s and 2000s cohorts. This is broadly compatible with historical records of high-skilled Indian migration to the United States (Desai, Kapur, and McHale 2005) but also possibly emphasized by the characteristics of our LinkedIn records—namely, right truncation at 2016 and underreporting for the earlier cohorts (the older an individual, the less likely he or she is to maintain a LinkedIn profile).

Second, the importance of the education channel relative to the work channel is both evident for early cohorts and declining over time. This trend again is broadly compatible with the history of graduate and postgraduate education in India since the 1960s, whose offer and quality were extremely limited until the 1990s (so an early Indian migrant seeking a job in science or engineering usually obtained a graduate education in the host country; Kapur 2010). But it may be accentuated, once again, by underreporting for early cohorts and its correlation with educational levels (the more likely an individual is to have migrated through the work channel, which is associated with a lower education level, the less likely he or she is to maintain a LinkedIn profile, especially in the case of an older individual). These observations suggest that our data are more reliable for the 1990s and 2000s cohorts, which concern 4,362 individuals—namely, 79 percent of migrants in our database.

Figures 1.2a and 1.2b provide further details on the education levels of both education and work migrants. We first remark on how the overwhelming majority of the former and the relative majority of the latter hold masters' degrees. This suggests that PhD holders and academic scientists, for which Finn (2014), Gaulé (2014), and Kahn and MacGarvie (2016) have provided some evidence, are not a representative sample of migrant inventors in the ICT industry. We also notice that the share of doctorate holders is higher for education-based migrants, while the share of bachelor holders is higher for work-based ones, which is in line with our selection criteria for the two categories.

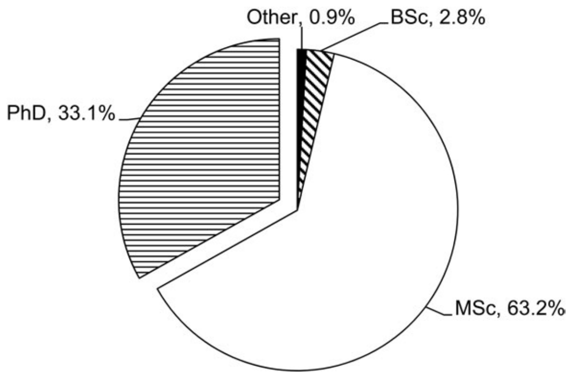


**Fig. 1.2a Highest educational attainment, percentage distribution: education migrants**

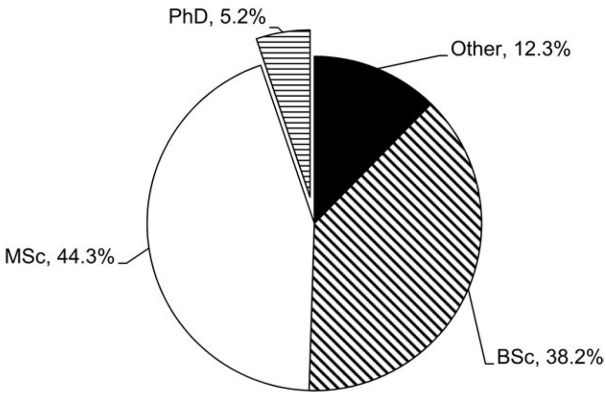


**Fig. 1.2b Highest educational attainment, percentage distribution of work migrants**

Figure 1.3 reports the total return rates for all migrants in our sample (irrespective of length of stay) by migration channel. For comparative purposes, the return rates are calculated both according to the definition of returnee we adopted earlier (first job or patent back in India, as per LinkedIn profile) and to a purely patent-based definition (first patent back in India, irrespective of other information). The latter corresponds to that found in most of the available literature on the international mobility of inventors, which relies exclusively on patent data and can observe a cross-border move only for inventors with at least two patents in as many different countries (e.g., Oettl and Agrawal 2008). We notice immediately that this definition severely underestimates return rates (black bars) compared to the one also based on job information (white bars), whatever migration channel we consider. In fact, the latter also includes among the returnees the inventors

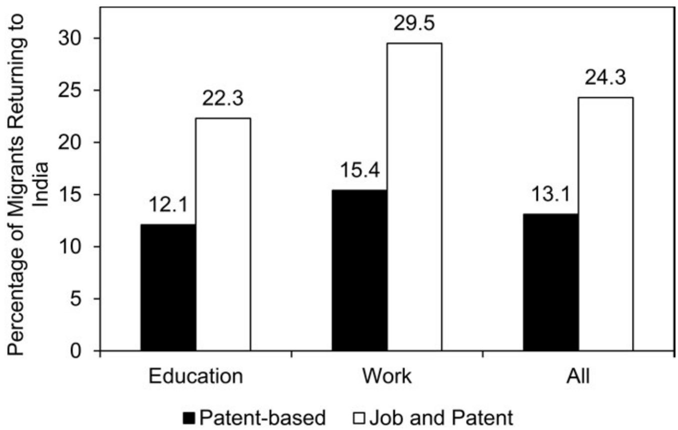


**Fig. 1.3 Total return rates by migration channel (irrespective of the length of stay in the United States)**

with no more than one patent in their careers (either in the United States or in India) but education or employment in a different country than the one where that only patent was signed. More generally, it also counts as returnees the inventors whose entire patent production occurred in one country but whose education or career took place elsewhere.

When comparing migration channels, figure 1.3 reports a seven-point difference in the return rate of work-migrant inventors compared to education ones. This may be due to the different types of visas used to enter the United States in terms of both initial validity length and renewal ease and also different efforts that work and education migrants may make to convert their temporary visas into permanent ones. Different types of migrants may also be differently exposed to opportunities to establish social ties in the United States, which may influence their propensity to return at each point in time.

Figures 1.4a and 1.4b report the total return rates (based on both patent and job information) for different cohorts of migration to the United States. The return rates for education migrants appear to be increasing, and this is despite the longer observation interval for older cohorts (which intuitively should lead to more accumulated returns). However, for cohorts before 1990, the number of observations is rather limited, and as discussed in the previous subsection, the probability of underreporting by return migrants is rather high. As for the 2010 cohort, once again we are faced with very few observations, which makes the very high return rate figure extremely unreliable. Once again, we can trust only the data for the 1990 and 2000 cohorts, which still exhibit different return rates.

Contrary to education migrants, the return rates of work migrants appear rather stable, especially for recent cohorts.

As discussed in the introductory part of the chapter, the return rates found

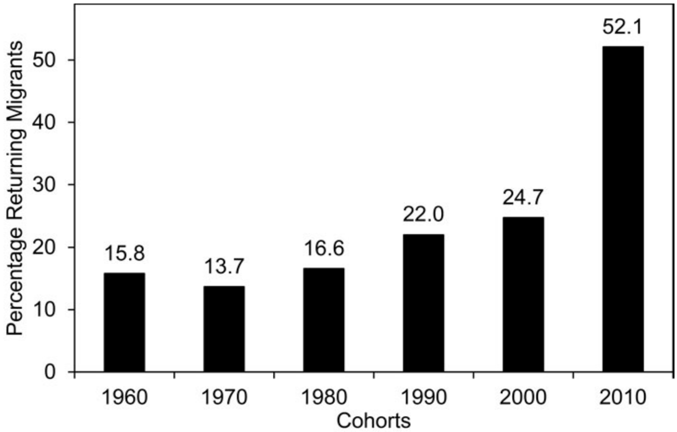


**Fig. 1.4a Percentage of education migrants returning to India by cohort (irrespective of the length of stay)**

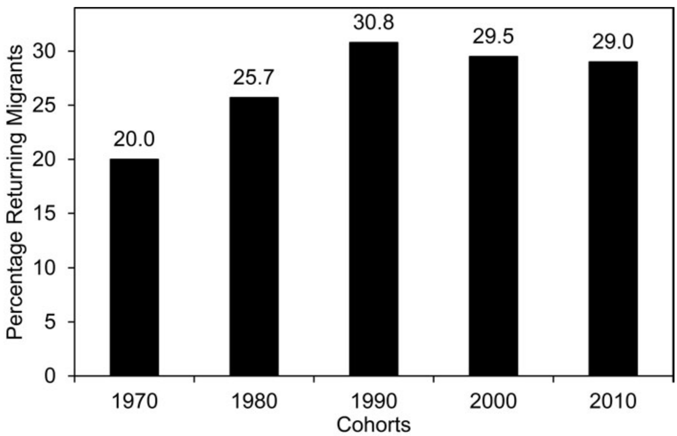


**Fig. 1.4b Percentage of work migrants returning to India by cohort (irrespective of length of stay)**

in the literature vary considerably depending on the sample and countries analyzed. While Finn (2014) calculates a return rate just after graduation of about 10 percent for India (up to around 15 percent in more recent estimates; Finn and Pennington 2018), other studies report return rates of around 40 percent both for Indian H-1B visa holders (Lowell 2000) and for Indian PhD or master's students (Wadhwa 2009).

Figure 1.5 reports the Kaplan-Meier estimators for work and education migrants from the 1990 and 2000 cohorts, with time measured yearly. We notice that the survival (stay) rate for work migrants is both lower and more

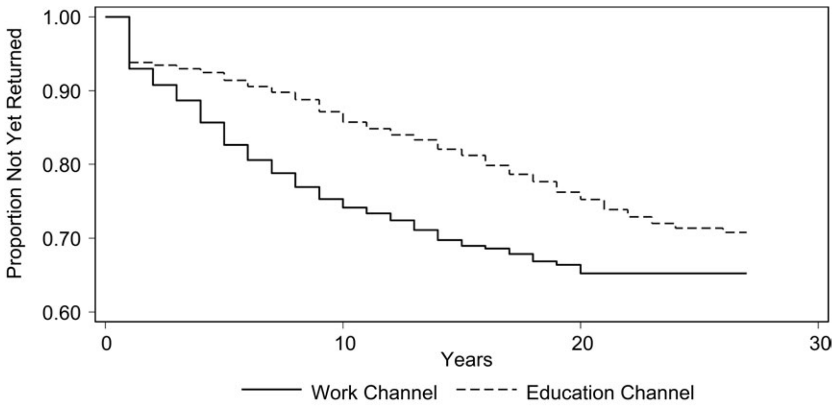


**Fig. 1.5 Stay rates over time (years since migration) by migration channels (1990 and 2000 cohorts)**

rapidly decreasing over time than for education ones. We also notice that the stay rate after 10 years since migration for education migrants (slightly less than 90 percent) is very close to what is reported by Finn (2014) for Indian PhD graduates in the United States. We take this as a sign of the reliability of our data.

Table 1.2 provides detailed information on the return time for migrant inventors in the 1990 and 2000 cohorts. Returnees in the first cohort leave the United States, on average, 11 years after their arrival. The minimal return time is zero (which implies a return to India less than a year after entry into the United States), and the value of the first quartile is 5.5. This indicates that 25 percent of the returnees in the 1990 cohort go back to India either in the same year of their arrival or not later than 5.5 years afterward. An additional 25 percent leave between 5.5 and 11 years after their arrival, followed by 25 percent more who leave between 11 and 16 years. The maximum stay, for returnees, is 25 years. When splitting the 1990 cohort between work and education returnee migrants, the former exhibit shorter stay periods both on average and according to the quartile distribution. The 2000 cohort exhibits, on average, shorter stays than the 1990 one (which may be due to shorter exposure to the return risk) but also less striking differences between work and education migrants.

## 1.3 Specification

We exploit our data to explore the extent of skill-based self-selection in return migration of the highly skilled. Skill-based self-selection was first investigated by Borjas (1989) in order to provide an explanation for two common stylized facts concerning the education and income levels of migrants.

**Table 1.2** **Average time to return by cohort**

| | All channels | Education | Work |
|---|---|---|---|
| | *Cohort 1990* | | |
| # of inventors in cohort | 2,025 | 1,739 | 286 |
| # of returnees | 471 | 383 | 88 |
| Mean | 10.58811 | 11.22193 | 7.829545 |
| Std | 6.643841 | 6.71642 | 5.560992 |
| Min | 0 | 0 | 0 |
| 25% | 5.5 | 7 | 3 |
| 50% | 11 | 12 | 8 |
| 75% | 16 | 16 | 13 |
| Max | 25 | 25 | 19 |
| | | *t*-test 4.952 (*p*-value 0.000) | |
| | *Cohort 2000* | | |
| # of inventors in cohort | 2,337 | 1,315 | 1,022 |
| # of returnees | 626 | 325 | 301 |
| Mean | 3.889776 | 4.132308 | 3.627907 |
| Std | 3.897362 | 4.316674 | 3.374574 |
| Min | 0 | 0 | 0 |
| 25% | 0 | 0 | 0 |
| 50% | 3 | 3 | 3 |
| 75% | 7 | 8 | 6 |
| Max | 16 | 16 | 15 |
| | | *t*-test 1.635 (*p*-value 0.103) | |

*Note:* Cohort 1990 includes inventors who migrated to the US between 1990 and 1999; cohort includes inventors who migrated to the US between 2000 and 2009.

First, stock data on foreign-born versus native populations recurrently show that the former are, on average, better educated than the latter for most traditional destination countries. Second, when observing a cohort of foreign-born over time through successive censuses, it is often found that starting from a lower average wage or income level, migrants catch up relatively quickly. Regardless of whether migrants are positively self-selected at entry, with respect to their education and/or unobservable skills, negative self-selection may help explain this evidence to the extent that return migrants escape successive censuses, therefore leaving behind them, in the host country, only the best and brightest of their respective immigration cohorts.

Borjas and Bratsberg (1996) provide a classic treatment of the topic, in which they show that different remuneration levels of skills in the host and home countries jointly determine whether migrants will be positively (negatively) self-selected upon arrival and, conversely, negatively (positively) selected upon return. In other words, return migration is expected to reinforce the sign of skill-based self-selection at entry. Dustmann and Görlach (2016) provide the last in a series of refinements of this basic idea, which describes the migrant's behavior at his or her destination (including his or her investment in the acquisition of education and skills) as resulting from the same lifetime optimization plan that determines the return decision and timing.

Other, less-dominant theories of return migration stress the fact that many migrants neither move permanently to the host country nor return home once and for all after a prolonged spell abroad. Instead, they move back and forth between the home and the host countries (or several host countries), possibly in response to economic shocks (Constant, Nottmeyer, and Zimmermann 2013). In this case, we should not expect any positive or negative self-selection, the economic shocks being orthogonal to skill levels.

Empirical studies on return migration can be categorized according to two criteria: (a) whether they observe and explain the actual duration of migration spells, from entry to return, or simply compare the characteristics of stayers and returnees; and (b) whether they focus on observed return moves or on return intentions.

With respect to (a), empirical studies fall into one or the other category depending on data availability and, to a lesser extent, on their theoretical focus. On the data side, most studies simply do not have longitudinal information on individual migrants—that is, they have no records on entry and return dates. Based on this limited information, they can only apply linear probability or logit/probit models and investigate the determinants of the probability to return, irrespective of when this occurs. When longitudinal data are available, instead, one can apply duration analysis (also known as survival or event history analysis; Allison 2014). This has two advantages over linear probability or logit/probit models. First, it is not inherently static, and therefore it allows one to consider time-varying covariates, so as to study how intervening changes in the migrant's characteristics may affect the return decision. Second, and more importantly, duration analysis allows estimating the propensity to return for those who have not yet returned, at each point in time during their entire permanence abroad, and not just the probability to return after a pre-determined spell abroad (say one, two or five years). By derivation, one can explain or predict the timing of the return decision and not just the probability of its occurrence. This also implies that by means of duration analysis, we can test whether the probability to return is time-dependent, either positively or negatively. According to Constant and Massey (2002), negative time dependence may be indicative of negative skill-based self-selection (where skills are unobservable). The longer a migrant stays in the host country, the more country-specific skills he or she accumulates, which are hard to transfer and/or are less remunerated at home, ceteris paribus. This makes return increasingly less likely. At the same time, to the extent that migrants vary in the speed at which they accumulate local skills, early returnees would necessarily be those who, at a given point in time, have accumulated fewer local skills.

Coming to the distinction between studies based on observed return moves or declared return intentions, this often boils down, once again, to data availability, with survey data being much better at recording the latter than the former (see our earlier discussion on how we record return moves). However, some recent literature suggests that data on return intentions better serve the purpose of testing lifetime income maximization models. This is because, according to such theories, most migrants leave their countries with the intention to return at a date which depends on their investment plans in education and skill acquisition while abroad.

The data structure for our regression exercises is a panel one, with each inventor $i$ being observed repeatedly since his or her immigration year until the minimum between his or her return year (when he or she exits the panel) and 2016, our last observation year. In this way, we have a large number of right-censored observations, but no left-censored ones. In what follows, we exploit this feature of our data and estimate the determinants of actual return decisions by means of discrete time duration analysis. Given the exploratory nature of our exercise, we do not put forward any claim of having established causal links. We care instead for producing much-needed evidence on return frequency and timing and its association to observable and unobservable skills (i.e., self-selection based on education, patenting activity, and time spent in the United States).

Following Jenkins (2005), we assume a proportional hazard function, which, in a discrete time setting such as ours, results in a complementary log-log (cloglog) model, as follows:

$$h(t, x)_i = 1 - \exp[-\exp(c(t) + \beta_i X_i)],$$

where $c(t)$ represents a generic inventor's baseline probability to return home after a migration spell $t$ (duration), conditional on not having yet returned, and $\beta_i X_i$ is a scaling factor depending on specific inventor $i$'s characteristics $X_{it}$ (some of which are time-variant). As for $t$, we measure it as either the number of years (plus 1) spent in the US since immigration or, for conducting robustness checks on education migrants only, the number of years since the end of their first education spell in the United States.

Concerning the baseline hazard ratio $c(t)$, we adopt two alternative specifications. First, we follow Constant and Massey (2002) and enter $t$ with a quadratic term, as follows:

$$c(t) = \alpha_1 t + \alpha_2 t^2. \tag{1}$$

This parametric specification may allow us to test for any time dependence of the hazard ratio, and its sign, in a rather immediate and intuitive way, on the basis of estimates for $\alpha_1$ and $\alpha_2$. But it comes at the cost of imposing a specific functional form to $c(t)$.

Second, we experiment with a nonparametric specification (as in Gaulé 2014) and make use of fixed effects, as follows:

$$c(t) = \eta_1 t_1 + \ldots + \eta_N t_N, \tag{2}$$

where $(t_1 \ldots t_n)$ is a set of duration dummies corresponding to migration spells lasting from 1 to N years (and N is the longest spell observed in our data). This model has the advantage of not imposing any functional form to the hazard ratio, but it produces so many estimated coefficients that in order to appreciate any time dependence of the hazard ratio, one needs a graphical representation.

Based on the evidence from figures 1.4 to 1.6, plus table 1.2, in the previous section, we expect time to affect differently the hazard ratio of work- and education-based migrant inventors. Hence we run separate regressions for the two types of migrant inventors. We also restrict our regressions to the two most populated migration cohorts in our sample—namely, the 1990s and the 2000s ones, for which data are more reliable. We also right-censor our data at 2016 as a matter of convenience. This makes the longest possible duration equal to 27 years.

Coming to our choice of regressors $X_i$, they include both a set of time-invariant variables that describe the migrant's conditions at entry in the United States and a set of time-variant ones that describe his or her activities during his or her permanence there (see table 1.3 for descriptive statistics).

As for conditions at entry, we consider the inventor's age, educational level, migration cohort, and patenting experience at migration, all of which we expect to be positively associated to the return hazard, as they may proxy for the inventor's stronger attachment or professional insertion in India and may negatively affect his or her chance to renew the initial temporary visa. We measure age in years (*Age at migration*) and education with the dummy variable *Master's or more at migration* (the reference case being that of migrants with no more than a bachelor's at migration; as for doctorate holders, they are too few to create a meaningful separate category, so we treat them as master's holders). Due to our restriction of the analysis to just two migration cohorts, we control for them with just a dummy for the 2000s one (1990s as reference). As for patenting experience, we measure it with the cumulative number of patents signed at the time of migration (*Patent stock at migration*).

As for activities in the United States, we consider the following:

- the migrant's student status (*Student*), which is a dummy taking a value of one for all the years between the start and end years of an education spell in the United States, whatever its level, and zero otherwise;

- the migrant's educational attainment while in the United States, as measured by the dummy variables *Master's* and *PhD*, which takes a value of zero before the year of completion of, respectively, the migrant's master's or doctoral studies, and one thereafter;
- the migrant's productivity as an inventor while abroad, which we measure with the *cumulative number of patents* from entry into the United States up to observation time *t*.

We expect the student status to lower return hazard, as it guarantees the migrant the renewal of his or her temporary visas. As for the educational attainment, based on the existing evidence of Indian graduates' low return rates, we also expect a negative impact on the return hazard. In other words, we expect negative self-selection based on education. As for the number of patents filed in the United States, we would expect negative self-selection, but the interpretation of this variable is complicated by the fact that not all migrants in our sample, once in the United States, pursue careers as inventors but may instead move on to management, entrepreneurship, or academia. (We come back to this issue when commenting on the results.)

## 1.4 Results

Table 1.3 reports separate descriptive statistics for the education and work migration channels. We notice some important differences between education and work migrants besides the age at migration.

First, work migrants are considerably more likely to leave India after graduating at the master's level; most education migrants move to the United States precisely to earn that same degree. As for earning a PhD, this happens almost exclusively to education migrants. In this respect, it is important to remark that this may happen on top of getting a master's but also as an alternative to it, with the latter case being the most frequent.[10]

Both education and work migrants exhibit a rather low average number of patents before moving to the United States, but the figures are higher for the latter. At a closer inspection, our data reveal that most migrants in our sample leave India without having filed any patent there. In fact, only about 1 percent of education migrants and 4 percent of work migrants have a nonnull patent record before migrating. As for the cumulative number of patents filed while in the United States, its average value is higher for work migrants than for education ones (around five against four). When looking at the underlying distribution (unreported in the table), we notice that only 2 percent of work migrants never file any patent while in the United States, while the same figure for education migrants amounts to 14 per-

10. It is very likely, however, that we largely overestimate the number of PhD holders without a master's. This is due to many LinkedIn members reporting only their highest educational achievements (such as a doctorate) and not the previous ones (such as a master's).

**Table 1.3** **Descriptive statistics by migration channel**

| | Education channel | | | | | Work channel | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Obs | Mean | Std. dev. | Min | Max | Obs | Mean | Std. dev. | Min | Max |
| Migration cohort | 50,211 | 1,993.1 | 4.630 | 1,990 | 2,000 | 15,333 | 1996.8 | 4.648 | 1,990 | 2,000 |
| Age at migration | 50,211 | 24.32 | 2.652 | 18 | 52 | 15,333 | 31.87 | 5.956 | 18 | 62 |
| Master's or more at migration | 50,211 | 0.09 | 0.283 | 0 | 1 | 15,333 | 0.34 | 0.473 | 0 | 1 |
| Current student status | 50,211 | 0.20 | 0.403 | 0 | 1 | 15,333 | 0.04 | 0.192 | 0 | 1 |
| Master's in the US | 50,211 | 0.66 | 0.474 | 0 | 1 | 15,333 | 0.04 | 0.202 | 0 | 1 |
| PhD in the US | 50,211 | 0.20 | 0.400 | 0 | 1 | 15,333 | 0.01 | 0.097 | 0 | 1 |
| MBA in the US | 50,211 | 0.08 | 0.267 | 0 | 1 | 15,333 | 0.04 | 0.201 | 0 | 1 |
| Patents at migration | 50,211 | 0.01 | 0.114 | 0 | 5 | 15,333 | 0.03 | 0.354 | 0 | 12 |
| Cumulative # patents US | 50,211 | 3.83 | 10.64 | 0 | 261 | 15,333 | 4.71 | 9.07 | 0 | 162 |

cent (the overwhelming majority of these individuals patent only when they return to India, while a tiny minority may have patents before migrating). As for those who filed at least one patent in the United States, the differences between work and education migrants are much less striking, albeit education migrants exhibit more variability (witness the standard error reported in table 1.3). In both subsamples, over a third of migrants file just one patent while in the United States and as many file from two to five (followed by a very long tail for values higher than ten), but education migrants are slightly more likely to file just one patent, or two to five, as well as more than one hundred.

We notice an important difference between education and work migrants with respect to the number of patents filed while in the United States, which on average is higher for the latter. As for the very high maxima that we observe for this variable, they correspond to very senior principal scientists in large ICT companies.[11]

Table 1.4 reports the results of our regressions, which we run separately for education and work migrants. The first two columns refer to parametric specification (1) of the baseline hazard ratio $c(t)$, while the other two refer to the nonparametric specification (2). In both cases, we calculate the estimated odds ratios, which we read as the marginal effects of the covariates on the return hazard ratio (Jenkins 2005).

We first ask to what extent return migrants appear to be self-selected with respect to either one of their observable skills, namely, education and patenting activity. We then move on to analyze the sign of time dependence of the hazard ratio.

Concerning education, we first notice that the odds ratio for *Master's or more at migration* is greater than one in all columns of table 1.4, but it is significant in only one case (for education migrants in column 1). Hence there is evidence of return migrants being positively selected with respect to education they obtained in India, but it is rather weak. On the contrary, all return migrants appear to be negatively selected with respect to education obtained in the United States. For education migrants, both *Master's in the US* and *PhD in the US* have estimated odds ratios largely inferior to one (the reference case being migrants obtaining only a bachelor's degree or not completing their graduate studies).

However, the difference between the underlying coefficients is nonsignificant, which suggests that for individuals holding either a master's or a PhD, graduate education is all that matters, and more advanced or research-oriented degrees do not convey any particular advantage to migrants intending to stay in the United States or to those with return intentions. As for

11. These are the cases, respectively, of education migrant Durga Malladi of Qualcomm (261 patents) and work migrant Alok Srivastava, an independent consultant with activities in both India and the United States (162 patents).

**Table 1.4** **Event history analysis of return risk, discrete time analysis, by migration channel**

| | Education channel (1) | Work channel (2) | Education channel (3) | Work channel (4) |
|---|---|---|---|---|
| Time from migration | 0.881*** | 0.883*** | | |
| | (0.0201) | (0.0307) | | |
| Time from migration$^2$ | 1.005*** | 1.002 | | |
| | (0.000830) | (0.00195) | | |
| Migration cohort = 2000 | 1.779*** | 1.423*** | 1.867*** | 1.424*** |
| | (0.138) | (0.168) | (0.150) | (0.170) |
| Age at migration | 0.872*** | 0.899*** | 0.977 | 0.904*** |
| | (0.00565) | (0.00467) | (0.0159) | (0.0115) |
| Master's or more at migration | 1.623*** | 1.154 | 1.180 | 1.138 |
| | (0.227) | (0.136) | (0.176) | (0.139) |
| Current student status | 0.595*** | 0.160*** | 0.459*** | 0.173*** |
| | (0.0681) | (0.0809) | (0.0908) | (0.0884) |
| Master's in the US | 0.432*** | 0.724 | 0.568*** | 0.719 |
| | (0.0444) | (0.215) | (0.0709) | (0.216) |
| PhD in the US | 0.552*** | 1.259 | 0.585*** | 1.430 |
| | (0.0744) | (0.763) | (0.0805) | (0.835) |
| MBA in the US | 0.866 | 0.401** | 0.711** | 0.403** |
| | (0.148) | (0.169) | (0.124) | (0.171) |
| Patents at migration | 2.525*** | 1.429*** | 2.320*** | 1.431*** |
| | (0.358) | (0.0842) | (0.301) | (0.0822) |
| Cumulative # patents US | 1.001 | 1.011** | 0.999 | 1.012** |
| | (0.00429) | (0.00528) | (0.00524) | (0.00528) |
| Observations | 50,211 | 15,333 | 50,211 | 15,094 |
| Times dummies | NO | NO | YES | YES |
| # unique inventors | 3,054 | 1,308 | 3,054 | 1,308 |
| Chi2 | 11,757 | 4,625 | 11,347 | 4,604 |
| LogL | –3,623 | –1,684 | –3,442 | –1,664 |

*Note:* Inventor-level clustered standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

those holding both a master's and a PhD, however, the two effects may sum up, which reinforces the negative selection effect of education on return migrants.

As for work migrants, neither *Master's in the US* nor *PhD in the US* is significant, and what really seems to count to increase their chances of staying in the United States is getting an MBA, whose coefficient is way less than 1, although significant only at 95 percent. Notice that *MBA in the US* also appears significant in one of the regressions for education migrants, but with an odds ratio closer to one.

Coming to patenting activity, inventors who leave India with substantial patenting experience are definitely those with the higher return hazard: witness the size of the odds ratio of Patents at migration for both education and work migrants (respectively, well over 2 and close to 1.5). Whether this result can be interpreted as evidence of positive self-selection (in contradiction with the education-based negative self-selection) is doubtful. The number of individuals in our sample with at least one patent at migration is very limited, and for several of them, we may overestimate the occurrence of return.[12]

As for the patenting activity in the United States (*Cumulative # patents US*), we also find it to be positively related to the return hazard, with odds ratios barely larger than one and not significant for education migrants. However, rather than being related to positive self-selection, this result may be related to specialization. In fact, inventors in our database range from the occasional to the professional ones, the former having signed one or very few patents before or after migration, the latter displaying instead a significant patenting record, one that possibly spans several years. In the absence of information on the migration strategies adopted by individuals in our sample or on the opportunity and constraints that may shape them, we can speculate about what follows. Professional inventors are more likely to move to the United States on a strictly temporary basis and for the specific task of undertaking inventive activities there, possibly on request of their employer in India, which organizes their two-way trip. Occasional inventors instead may be a more heterogeneous group, which includes a large number of individuals moving to the United States on their own initiative, rather than their employer's, and more determined to turn an originally temporary visa into a permanent one. They will be at once more open toward different career options and less bound by the original visa arrangements. For example, they may move out of the research and development (R&D) laboratory and stop producing patents, possibly to undertake managerial functions or an entrepreneurial career, thus getting more chances to stay in the United States. This interpretation fits with the size and significance ratio of the MBA in the US variables, on which we commented above. Notice that this explanation applies better to work migrants than education ones, all of them entering the United States via a higher education program and therefore more likely to be occasional rather than professional inventors. This is coherent with the odds ratios for *Cumulative # patents US* being de facto equal to one in the regressions for education migrants.

Moving to time dependence of the hazard ratio, the estimated odds ratios in columns (1) and (2) suggest it to be negative and monotonic for work

12. Many individuals with patents at migration are considered returnees on the basis of their patenting activity, with the patent apparently marking their return ("return patent") to India closely following the event (job, education, or patent) marking their original migration to the United States. For education migrants, it may well be that the "return patent" was actually invented before the migration event but filed afterward, so we are facing a false positive case of return migration. For work migrants, besides false positives, we may face cases of inventors temporarily detached in the United States for very short periods.

migrants (the coefficient for the time-squared is not significant) but possibly nonmonotonic for education ones (the coefficient for the time-squared is significant, and the odds ratio is greater than one).

Following Constant and Massey (2002), we interpret the negative time dependence of the return hazard ratio as indicative of some negative self-selection with respect to unobservable skills the migrant acquires through experience in the host country that are not as well rewarded back at home. Admittedly, Constant and Massey's interpretation of the time-dependence of the hazard ratio is rather speculative, since other factors besides skill accumulation may intervene, such as increasing investments in real estate or social capital, both of which increase the opportunity cost of return. Still, the negative time dependence we find for work migrants is coherent with the possibility that those among them who stay longer in the United States also engage in managerial functions or undertake entrepreneurial careers. Such career moves come with developing skills for which the US-India remuneration gap may be higher than that for the skills exclusively associated with R&D-performing tasks, thus discouraging return. They may also come with job contracts for which it is easier to obtain a permanent visa than a temporary one..

As for the time pattern of education migrants' return hazard ratio, regression in column (1) is not very enlightening. First, it results from imposing a parametric form to $c(t)$; second, it requires one to understand whether opposite signs of the estimated coefficients for $\alpha_1$ and $\alpha_2$ imply some nonmonotonicity, which is not immediately clear in the case of nonlinear estimation methods such as cloglog. For this reason, we prefer relying on the results of the nonparametric estimation of column (3). Based on such results, figures 1.6a and 1.6b report the within-sample estimates of the total hazard ratio $h(t)$ as a function of time and for different educational levels by migration cohort.

Both figures suggest that the return hazard follows an inverted U-shaped function of time over the first 13 years of permanence in the United States. After that, we cease to observe migrants in the 2000 cohort, due to right truncation, while the return rate for the 1990 cohort starts increasing again, albeit erratically. The hazard ratios for the early years after entry, however, may be underestimated. This is because we produced the graph by setting *Current student status* equal to zero, while in reality it should be equal to one from entry in the United States until graduation (notice that the odd ratios for *Current student status* in table 1.4 are always greater than one). As a partial remedy, we have replicated regression (3) in table 1.4, but with duration $t$ counted from the end of the migrant's first student spell in the United States. Results for the estimated return hazard ratios are reported in figures 1.7a and 1.7b, which we can compare with figures 1.6a and 1.6b. We notice how the estimation of return hazard ratios with respect to time now

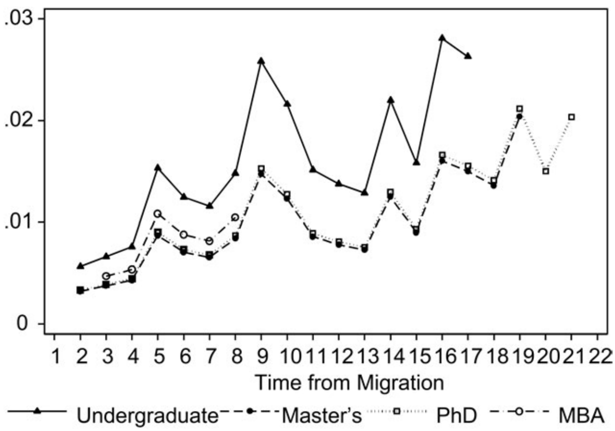


**Fig. 1.6a Estimated hazard ratios since entry into the United States by education level: education migrants, 1990 cohort**

*Note:* Within-sample estimations from regression (3) in table 1.4 for age at migration = 23 and student status = 0 (all remaining regressors at mean values).

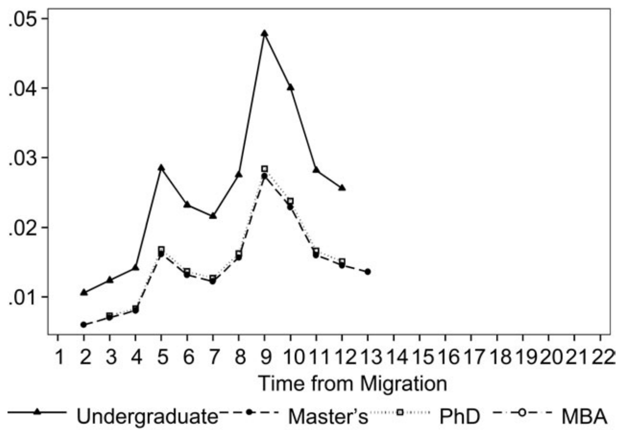


**Fig. 1.6b Estimated hazard ratios since entry into the United States, by education level: education migrants, 2000 cohort**

*Note:* Within-sample estimations from regression (3) in table 1.4 for age at migration = 23 and student status = 0 (all remaining regressors at mean values).

changes: the inverted U-shape profile we initially observed is significantly smoothed, and the return hazard ratio appears first to increase and then to flatten down.

Overall, however, we find some signs of a positive time dependence of the return hazard on time for education migrants, which may imply positive self-selection with respect to unobservable skills. We further discuss these results in the conclusions.

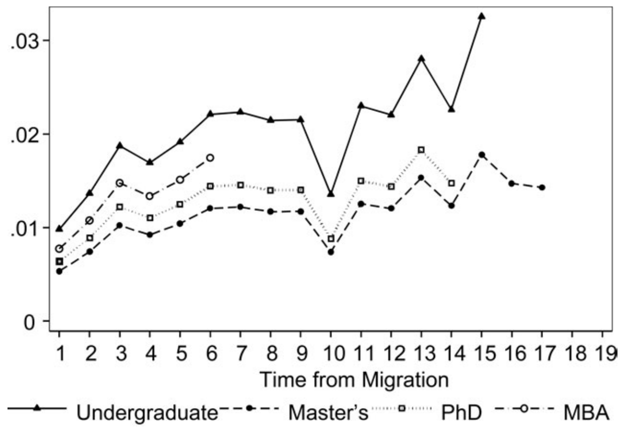


**Fig. 1.7a Estimated hazard ratios since completion of studies in the United States by education level: education migrants, 1990 cohort**

*Note:* Within-sample estimations (unreported regression) for age at migration = 23 and student status = 0 (all remaining regressors at mean values).

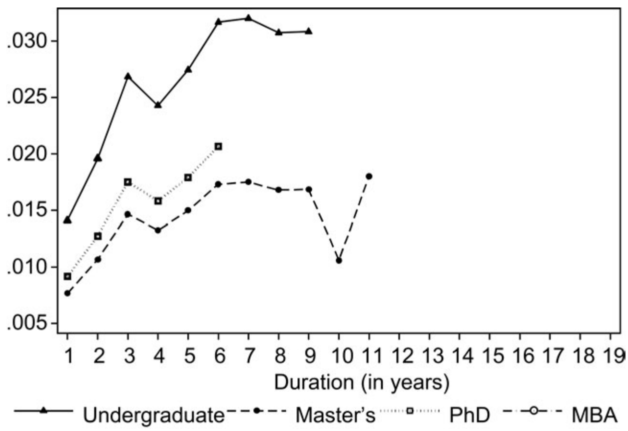


**Fig. 1.7b Estimated hazard ratios since completion of studies in the United States by education level: education migrants, 2000 cohort**

*Note:* Within-sample estimations (unreported regression) for age at migration = 23 and student status = 0 (all remaining regressors at mean values).

## 1.5 Conclusions

Return migration is a much understudied topic, especially when it comes to its implications for innovation in both the host and home countries. Lack of data is a major cause of this situation due to the virtual absence of official statistics and the technical difficulties that stand in the way of large-scale data mining.

In this chapter, we have presented the outcome of an ambitious attempt to overcome such difficulties based on linking inventor information from patent data to biographical information from an important web-based social network. We focused on Indian inventors with professional experiences of various lengths at one or more US ICT companies and obtained rather reliable data for those among them who moved to the United States in the 1990s and 2000s. Based on biographical information, we could draw a clear distinction between work and education migrants and analyze separately the related return events. In particular, we applied event history analysis and explored the issue of returnees' self-selection with respect to observable and unobservable skills.

Both the distinction between work and education migrants and the study of self-selection may contribute to evaluating the effectiveness of US migration policies, with special reference to scientists, engineers, and other innovation-relevant professional categories.

As stressed by Koslowski (2018), US immigration policies are often compared unfavorably to those of countries such as Canada and Australia, whose selective, point-based visa systems are held responsible for their records of attracting high proportions of high-skilled migrants. But the comparison is biased by its exclusive focus on migrants first entering their host countries with permanent visas, which accounts for a very limited share of entries in the United States. When considering migrants entering with temporary visas, whether work- or education-based, the United States appears the most attractive country, also in view of the large share of temporary migrants turning into permanent ones over the years. In this respect, it becomes crucial to estimate the stay rates of highly skilled permanent immigrants, which our study on Indian migrants finds rather high and in accordance with the limited evidence available in the literature, especially for education migrants.

Besides assessing the highly skilled migrants' length of stay, it is crucial to assess whether the host countries manage to retain the best and brightest among them—namely, those who can contribute most to innovation. In this respect, Wadhwa et al. (2009) give voice to widespread concerns on the difficulties supposedly met by the United States in this respect. Our results, albeit exploratory, go against such concerns for work migrants and leave room for debate on education migrants.

Concerning work migrants, Indian returnees in our sample appear to be negatively selected with respect to education as well as, most likely, to the working experience they accumulate in the United States (as inferred by the negative time dependence of their hazard ratios). Admittedly, we also find a positive relationship between the return hazard and the number of patents they produce while in the United States, but we have suggested how this may have more to do with specialization in managerial functions or entrepreneurship than with positive self-selection.

As for education migrants, Indian returnees in our sample are also negatively selected with respect to education but also appear increasingly at risk of return the longer their permanence in the United States, especially over the first 10 years after migration. This can be interpreted as positive self-selection with respect to unobservable skills, at least over the first few years after graduation. But we should bear in mind that our return migration measure does not distinguish between individuals who settle permanently back in their home countries or become engaged in circular migration patterns and/or parallel professional activities in their home and host countries.

Further research is clearly needed to both assess the strength of these initial results and extend them. Further codification of the information contained in our data set will let us assess the quality and location of the educational institutions attended by migrants so as to test whether the return hazard is positively or negatively associated with the prestige of the institution and/or its links with a vibrant labor market for the highly skilled. We also plan to fully disambiguate the name of companies reported by work migrants in their LinkedIn profiles so as to distinguish between intracompany and intercompany mobility. We expect the former to generate short-term temporary migrants, not much exposed to the risk of turning permanent, while the latter should be at the origin of longer stays and more interesting phenomena of negative versus positive self-selection.

More generally, our methodology may be extended to other countries of origin of migrants besides India and to other professional categories besides those related to ICT.

While a large amount of the knowledge we may gather on highly skilled return migration will pass through the refinement and sharing of our data, we think that some ad hoc theorizing is also necessary to adapt the emerging theoretical literature on temporary and circular migration we discussed in section 1.2 to the specificities of STEM workers and students.

## References


Allison, P. D. 2014. *Event History and Survival Analysis: Regression for Longitudinal Event Data*. Vol. 46. Los Angeles: SAGE.

Baruffaldi, S. H., and P. Landoni. 2012. "Return Mobility and Scientific Productivity of Researchers Working Abroad: The Role of Home Country Linkages." *Research Policy* 41 (9): 1655–65.

Bijwaard, G. E., C. Schluter, and J. Wahba. 2014. "The Impact of Labor Market Dynamics on the Return Migration of Immigrants." *Review of Economics and Statistics* 96 (3): 483–94.

Borjas, G. J. 1989. "Immigrant and Emigrant Earnings: A Longitudinal Study." *Economic Inquiry* 27 (1): 21–37.

Bönisch, P., P. Gaffert, and J. Wilde. 2013. "The Impact of Skills on Remigration Flows." *Applied Economics* 45 (4): 511–24.

Borjas, G. J., and B. Bratsberg. 1996. “Who Leaves? The Outmigration of the Foreign-Born.” *Review of Economics and Statistics* 78 (1): 165–76.
Breschi, S., F. Lissoni, and E. Miguelez. 2017. “Foreign-Origin Inventors in the US: Testing for Diaspora and Brain Gain Effects.” *Journal of Economic Geography* 17 (5): 1009–38.
Choudhury, P. 2016. “Return Migration and Geography of Innovation in MNEs: A Natural Experiment of Knowledge Production by Local Workers Reporting to Return Migrants.” *Journal of Economic Geography* 16 (3): 585–610.
Constant, A., and D. S. Massey. 2002. “Return Migration by German Guestworkers: Neoclassical versus New Economic Theories.” *International Migration* 40 (4): 5–38.
Constant, A. F., O. Nottmeyer, and K. F. Zimmermann. 2013. “The Economics of Circular Migration.” In *International Handbook on the Economics of Migration*, edited by A. F. Constant and K. F. Zimmermann, 55–74. Cheltenham: Edward Elgar.
Constant, A. F., and K. F. Zimmermann. 2016. “Diaspora Economics: New Perspectives.” *International Journal of Manpower* 37 (7): 1110–35.
De Rassenfosse, G., H. Dernis, D. Guellec, L. Picci, and B. v. P. de la Potterie. 2013. “The Worldwide Count of Priority Patents: A New Indicator of Inventive Activity.” *Research Policy* 42 (3): 720–37.
Desai, M., D. Kapur, and J. McHale. 2005. “The Fiscal Impact of High Skilled Emigration: Flows of Indians to the US.” Mimeo, Harvard University.
Dustmann, C., I. Fadlon, and Y. Weiss, Y. 2011. “Return Migration, Human Capital Accumulation and the Brain Drain.” *Journal of Development Economics* 95 (1): 58–67.
Dustmann, C., and J.-S. Görlach. 2016. “The Economics of Temporary Migrations.” *Journal of Economic Literature* 54 (1): 98–136.
Dustmann, C., and Y. Weiss. 2007. “Return Migration: Theory and Empirical Evidence from the UK.” *British Journal of Industrial Relations* 45 (2): 236–56.
Filatotchev, I., X. Liu, J. Lu, and M. Wright, M. 2011. “Knowledge Spillovers through Human Mobility across National Borders: Evidence from Zhongguancun Science Park in China.” *Research Policy* 40 (3): 453–462.
Finn, M. G. 2014. “Stay Rates of Foreign Doctorate Recipients from US Universities, 2011.” Oak Ridge, TN: Oak Ridge Institute for Science and Education (ORISE).
Finn, M. G., and L. A. Pennington. 2018. “Stay Rates of Foreign Doctorate Recipients from US Universities, 2013.” Oak Ridge, TN: Oak Ridge Institute for Science and Education (ORISE).
Gaulé, P. 2014. “Who Comes Back and When? Return Migration Decisions of Academic Scientists.” *Economics Letters* 124 (3): 461–64.
Ge, C., K. W. Huang, and I. P. L. Png. 2016. “Engineer/Scientist Careers: Patents, Online Profiles, and Misclassification Bias.” *Strategic Management Journal* 37 (1): 232–53.
Gibson, J., and D. McKenzie, D. 2014. “Scientific Mobility and Knowledge Networks in High Emigration Countries: Evidence from the Pacific.” *Research Policy* 43 (9): 1486–95.
Hawthorne, L. 2018. “International Student Mobility: Sending Country Determinants and Policies.” In *High-Skilled Migration: Drivers and Policies*, edited by M. Czaika, chap. 11. Oxford: Oxford University Press.
Hunt, J. 2011. “Which Immigrants Are Most Innovative and Entrepreneurial? Distinctions by Entry Visa.” *Journal of Labor Economics* 29:417–57.
Hunt, J., 2013. “Are Immigrants the Best and Brightest U.S. Engineers?” NBER

Working Paper no. 18696. Cambridge, MA: National Bureau of Economic Research.
Jenkins, S. P. 2005. "Survival Analysis." Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK.
Jones, B. F. 2009. "The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation Getting Harder?" *Review of Economic Studies* 76 (1): 283–317.
Jonkers, K., and L. Cruz-Castro. 2013. "Research upon Return: The Effect of International Mobility on Scientific Ties, Production and Impact." *Research Policy* 42 (8): 1366–77.
Kahn, S., and M. MacGarvie. 2016. "Do Return Requirements Increase International Knowledge Diffusion? Evidence from the Fulbright Program." *Research Policy* 45 (6): 1304–22.
Kapur, D. 2010. "Indian Higher Education." In *American Universities in a Global Market*, edited by C. T. Clotfelter, 305–34. Chicago: University of Chicago Press.
Kapur, D., and J. McHale. 2005. *Give Us Your Best and Brightest: The Global Hunt for Talent and Its Impact on the Developing World*. Washington, DC: Center for Global Development.
Kerr, W. R. 2017. "US High-Skilled Immigration, Innovation, and Entrepreneurship: Empirical Approaches and Evidence." In *The International Mobility of Talent and Innovation: New Evidence and Policy Implications*, edited by C. Fink and E. Miguelez, 193–221. Cambridge: Cambridge University Press.
Kerr, S. P., W. R. Kerr, and W. F. Lincoln. 2015. "Skilled Immigration and the Employment Structures of US Firms." *Journal of Labor Economics* 33 (S1): S147–S186.
Kerr, W. R., and W. F. Lincoln. 2010. "The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention." *Journal of Labor Economics* 28 (3): 473–508.
Kirdar, M. G. 2009. "Labor Market Outcomes, Savings Accumulation, and Return Migration." *Labour Economics* 16 (4): 418–28.
Koslowski, R. 2018. "Shifts in Selective Migration Policy Models." In *High-Skilled Migration: Drivers and Policies*, edited by M. Czaika, chap. 6. Oxford: Oxford University Press.
Lowell, B. L. 2000. "H-1B Temporary Workers: Estimating the Population." UC San Diego Working Papers. Working Paper no. 12.
Luo, S., M. E. Lovely, and D. Popp. 2013. "Intellectual Returnees as Drivers of Indigenous Innovation: Evidence from the Chinese Photovoltaic Industry." NBER Working Paper no. 19518. Cambridge, MA: National Bureau of Economic Research.
Nanda, R., and T. Khanna. 2010. "Diasporas and Domestic Entrepreneurs: Evidence from the Indian Software Industry." *Journal of Economics and Management Strategy* 19 (4): 991–1012.
OECD. 2008. *International Migration Outlook*. Paris: Organisation for Economic Co-operation and Development.
OECD. 2017. *International Migration Outlook*. Paris: Organisation for Economic Co-operation and Development.
Oettl, A., and A. Agrawal. 2008. "International Labor Mobility and Knowledge Flow Externalities." *Journal of International Business Studies* 39 (8): 1242–60.
Teitelbaum, M. S. 2014. *Falling Behind? Boom, Bust, and the Global Race for Scientific Talent*. Princeton, NJ: Princeton University Press.
Trippl, M. 2013. "Scientific Mobility and Knowledge Transfer at the Interregional and Intraregional Level." *Regional Studies* 47 (10), 1653–67.
UN. 1998. *Recommendations on Statistics of International Migration, Revision 1*. New York: United Nations Statistics Division.

Wadhwa, V. 2009. "A Reverse Brain Drain." *Issues in Science and Technology* 25 (3): 45–52.
Wadhwa, V., A. Saxenian, R. B. Freeman, and A. Salkever. 2009. "Losing the World's Best and Brightest: America's New Immigrant Entrepreneurs, Part 5." Mimeo. http://ssrn.com/abstract=1362012.
Zagheni, E., and I. Weber. 2015. "Demographic Research with Non-representative Internet Data." *International Journal of Manpower* 36 (1): 13–25.