# Return migrants' self-selection:

# Evidence for Indian inventors

Stefano Breschi [1/§], Francesco Lissoni [1/2], Ernest Miguelez [2/3]

[1] *ICRIOS – Università "L.Bocconi", Milan*

[2] *GREThA UMR CNRS 5113 – Université de Bordeaux*

[3] *AQR-IREA – Universitat de Barcelona*

[§] contact author: stefano.breschi@unibocconi.it

# Appendix – Data Methodology

From the NBER book, *The Roles of Immigrants and Foreign Students in U.S. Science, Innovation, and Entrepreneurship*, edited by Ina Ganguli, Shulamit Kahn, and Megan MacGarvie, published by the University of Chicago Press, 2019. https://www.nber.org/books/gang-1

### A. Data Sources

The data set used in this paper is the result of a linkage between USPTO patent and inventor data gathered from Patentsview[1] and biographical information extracted from a large number of LinkedIn profiles. Patentsview is a data repository and data visualization tool recently made available by the USPTO, which provides disambiguated data on inventors of USPTO granted patents from 1975 onward. LinkedIn, a well-known social networking system, reports a very large number of users' public profiles that include information on the users' educational curricula and careers (name and possibly locations of education institutions and employer), thereby allowing to trace (return) migration with a scale and degree of precision unmatched by other sources of data[2].

LinkedIn data are subject to a number of limitations. First, resume information is self-reported by individuals and therefore subject to misreporting or even cheating. Second, the choice of creating an account in a professional social network might be correlated with factors affecting the propensity to move (and migrate), thus leading to biased results. Third, we used LinkedIn "public" profiles, namely those who are publicly visible on the internet without being logged into LinkedIn. Hence, our data exclude those profiles for which the account holder chose to keep the profile as "private" and thus visible only from within the system and/or for paying subscribers. In spite of these limitations, we argue that LinkedIn data represent an unparalleled source of information on the international mobility of individuals, both as students and as workers (Ge et al., 2016; Zagheni and Weber, 2015). In what follows, we describe in detail the methodology used to build our sample and we report some tests on the accuracy of information coming from LinkedIn.

### B. Sample selection

For the purposes of the present paper, we extracted all the patents granted to the 179 largest US public firms in the ICT industry, from 1975 to 2016. The definition of ICT industry follows the one provided by the OECD[3]. To select our sample of firms, we proceeded as follows. For each SIC code contained in the OECD definition, we extracted from Compustat the list of public US firms active in that SIC and we matched them to the USPTO patent data. As company names reported in patents (i.e. patent assignees) may be written in different ways, we used two sources of information in order to disambiguate them: (a) the concordance tables between Compustat GVKEY codes and patent assignees provided by the NBER patent data project website[4]; (b) the PTMT Custom Bibliographic Patent Data Extract DVD produced by the USPTO, which provides first-named assigned owner at grant as harmonized for spelling variations[5]. From the resulting sample, we dropped all firms with less than 200 patents granted and that either disappeared (because of exit or acquisition) or were delisted before 2005. It is important to stress that for this paper, we kept patents of parent companies by simply disambiguating their names, but we did not collect patents of their subsidiaries with different names from the parent company. For example, ADC Telecommunications Oy and ADC Telecommunications Inc. were considered as the same company. However, patents of Codenoll Techology Corporation, which was acquired by ADC Telecommunications in 1996, have not been collected and consolidated with those of the parent company. Moreover, each company included in our sample was considered as active from the date of foundation to the date of exit (most often because of acquisition). Thus, for example, we considered ADC Telecommunications as an active independent company from 1974 to 2010, given that it was acquired by TE Connectivity in December

[1] http://www.patentsview.org/web/
[2] LinkedIn data used in this paper were obtained in June 2016.
[3] https://www.oecd.org/sti/ieconomy/1835738.pdf
[4] https://sites.google.com/site/patentdataproject/
[5] https://www.uspto.gov/web/offices/ac/ido/oeip/taf/data/misc/data_cd.doc/custom_extract_dvd/

2010. Finally, for each of the 179 firms thus identified we selected the inventors of their patents using Patentsview, for a total of 262,849 distinct individuals. The complete list of the 179 firms considered in the paper is reported at the end of this appendix.

*C. Ethnic analysis of inventor names: identification of Indian-origin inventors*

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% of all inventors employed by the companies in our sample. It is worth noting that this share is higher than the one reported in Kerr (2008). He estimates that the share of Indian inventors residing in the US with a patent application in the period 1975-2004 in the Computers technology field (i.e. the field closer to our sample) is equal to 6.9%. The difference with our estimates might be due both to the different time span covered (our sample includes patents granted up to December 31 2016) and to the different methodologies and data sources used to assign ethnicity (i.e. IBM GNR vs. Melissa)[6]. Moreover, our sample includes also inventors that, even though patented for US companies, do not reside in the US. Yet, the difference still persists even if we restrict the attention to US residing inventors. In this case, Indian origin inventors are 19,222 out of a total 211,480 inventors (i.e. about 9% of all US residing inventors).

*D. Matching Indian inventors and LinkedIn profiles*

Indian inventors were matched with the employees of the 179 ICT firms in our sample having a LinkedIn profile. The linkage was accomplished by matching first and last name of inventors and employees, on the one hand, and employer and patent assignee names, on the other hand. In other words, for each inventor having made patents with a given company we searched for an individual with the same (or a very similar) first and last name reporting the same company as an employer in the LinkedIn resume. Given that patent assignees and employer names were unlikely to match exactly, due to spelling variations, abbreviations and so on, we implemented a Python script using fuzzy matching techniques and regular expressions. Similar techniques were used to match names of individuals appearing in patent documents and in LinkedIn profiles. To this purpose, we preliminarily standardized the names in the two sets (e.g. removing special characters, such as dots, commas, hyphenations, semicolon etc., converting UTF8 characters into latin characters, removing suffixes such as Jr, PhD, and so on). Using these standardized names, we first performed an exact match between the names of inventors and the names of employees from the LinkedIn profiles. When an exact match was not found, we computed the Jaro-Winkler similarity[7] between the full name of inventors and LinkedIn profiles and we kept only those matches with a name similarity higher than 0.85. For those cases where inventors were matched to multiple LinkedIn profiles[8] because of homonyms, we used the city, state or country reported in the LinkedIn profile and in patents (when available) to improve our matching algorithm. We dropped all cases in which we were unable to unambiguously link an inventor to a unique LinkedIn resume. This exercise yielded 10,839 inventors matched with a LinkedIn account (around 45% of the original Indian inventor sample). This preliminary matched sample was further processed to drop false positives and improve accuracy. In

[6] For further details on the methods used to identify the ethnicity of inventors, see Breschi et al. (2017).

[7] https://en.wikipedia.org/wiki/Jaro%E2%80%93Winkler_distance

[8] In some cases, this problem was due to the fact that the same person opened up multiple profiles. In those cases, we picked up among the different profiles opened by the same individual the one containing more information, under the assumption that this is the profile currently maintained and updated by the person.

what follows, we describe the methodology used to extract and code information from LinkedIn resumes, as well as the steps undertaken to minimize measurement errors.

*E. Classification of educational attainments and country of education*

For each matched inventor, we extracted from the LinkedIn resume the information on their educational attainments and we coded the level of education according to the ISCED standard (2011 version)[9]. In particular, we coded the following education levels:

1) ISCED level 3: upper secondary education
2) ISCED level 5 – level 6: short-cycle tertiary education, Bachelor's or equivalent
3) ISCED level 7: Master's or equivalent level
4) ISCED level 8: Doctoral or equivalent level

Given our focus on inventors (i.e. scientists and engineers), we also distinguished between Master of Sciences (generaly in electrical and electronic engineering, computer science or related fields) and Master in Business Admnistration (MBA). It must be pointed out that information on education (like most other types of information) contained in LinkedIn resumes consists of free, unstructured text fields. As a consequence, the assignment of a given educational attainment to the corresponding ISCED level must be done by implementing some type of text classification algorithm. To this purpose, we implemented a Python script, which uses regular expressions and a list of keywords, capturing possible variations in which a certain degree title is written (e.g. a Bachelor of Engineering can be found written as such, but also as BEng, B. Eng, B.E. or other similar variations). We denoted as *unclassified* all those titles which we were unable to classify in any of the ISCED levels. They include a miscellanea of diplomas (e.g. Diploma of Information Technology) and professional certifications (e.g. Certificate IV Web Design, Project Management Professional PMP, and so on) that do not easily fit into any of the ISCED categories. Moreover, some of the matched LinkedIn resumes did not report any information on the educational attainment.

For each education level, we also coded the starting and end year and the country of the school where the education title was achieved. In few cases, resumes reported only the starting year of education. In those cases, we estimated the end year by using the average duration of the corresponding education level (e.g. four years in the case of Bachelor)[10].

Regarding the country of education, this was found by geocoding school names. To this purpose, we implemented a simple Python script which fed the name of the school into Google Maps, using the Google Maps Geocoding API. In few cases, Google returned more than one country match. We manually cleaned and checked those cases. Still in other cases, Google was unable to return a valid address, as information contained in the school name was not sufficiently detailed to allow accurate geocoding. We did not make any further check in those cases and we considered as missing the information on the school country. Overall, it is important to stress that this exercise might be prone to some (possibly limited) measurement errors. Given that the full address and city of schools is unknown and the only information we can provide is the school name, the accuracy of Google geocoding is high whenever the school name is sufficiently unique and distinctive (e.g. Bocconi University, Insead and so on), but it is likely to be lower for school names, such as St. James or St. Joseph School. Since geocoding errors are less likely to occur for university names, we manually

[9] http://uis.unesco.org/en/topic/international-standard-classification-education-isced

[10] More specifically, the average duration in years of the different educational levels for the inventors in our sample is: 5 (High school), 2.5 (short-cycle tertiary education), 4 (Bachelor), 3 (Master), 2 (MBA), 5 (PhD).

checked all geocoding results related to ISCED level 3 educational institutions (i.e. secondary schools).

*F. Employment history and employment country*

Similarly to what done in the case of education, we extracted the start and end year of each employment spell as well as the name of the employer as reported in the resume. Given that our interest is on the mobility of inventors across countries, particularly from India to the US and return, and not across firms, we did not disambiguate employers' names appearing in the LinkedIn resumes. Rather, we focused on the job location of each employment spell. In this respect, it is important to note that reporting the job location is not compulsory when filling the employment history of a LinkedIn resume. To illustrate this issue, we report below the employment history of two different inventors in our sample as they are reported in their respective LinkedIn resumes (see Tables A1 and A2). Both inventors report to have worked for Broadcom Corp. at some time during their working career. However, whereas one resume reports the job location at Broadcom in Bangalore, the other does not report any information on the job location.

In order to track the mobility of inventors from India to the US and return, we took the job location (if not missing) «self-reported» by the inventor in her resume and we coded whether the location was in the US (e.g. San Jose, Bay Area) or in India (e.g. Bangalore). Out of a total 35,456 employment spells recorded by the inventors in our sample, 7,743 reported the job location, namely around 22% of all job spells.

**Table A1: Inventor A, resume reporting job location in employment history**

| Job title | Employer | Job location | Period |
|---|---|---|---|
| Sr. System Engineer | Motorola Solutions | Dallas/Fort Worth Area | 1997-2000 |
| Member of Technical Staff | Iospan Wireless | San Jose, Bay Area | 2000-2002 |
| Student | University of Texas at Austin | Austin/Texas Area | 2003-2005 |
| Member of Technical Staff | Texas Instruments | Dallas/Fort Worth Area | 2005-2006 |
| Director (Technical ATD) | Broadcom | Bangalore | 2006-Present |

**Table A2: Inventor B, resume not reporting job location in employment history**

| Job title | Employer | Job location | Period |
|---|---|---|---|
| Mixed Signal Design Engineer | Crystal Semiconductor | . | 1997-1999 |
| Staff Design Engineer | Level One Communications (Intel Corp) | . | 1999-2000 |
| Director of Engineering, Broadcom Distinguished Engineer | Broadcom Corporation | . | 2000-2014 |
| Director, Touch and Sensing Hardware | Apple | . | 2014-Present |

In this paper, *we only considered «self-reported» job locations in assessing mobility and migration events*. When job location was missing, we did not consider the corresponding employment spell in assessing inventors' mobility from India to the US and return.

As illustrated above, only slightly less than a quarter of all employment spells reports the job location in the resume. Hence, our approach is fairly accurate (under the assumption that the location reported in the resume corresponds to the actual job location), but it is likely to under-estimate the extent of mobility and migration.

In principle, one might improve upon this method at the cost of somewhat lower accuracy. In particular, when the information on the job location is missing, one can estimate the likelihood of the job location to be in India (or more generally in a certain country) by exploiting information on other LinkedIn profiles reporting the same employer *and* the job location[11]. Given a certain employment spell whose job location is unknown, one can compute the fraction of all its employees with a LinkedIn profile (not just inventors, but any LinkedIn profile holders) who associated such employer to an Indian address (or to the address of a focal country).

To illustrate the idea, consider the employment spells of the inventor reported in Table A3. This is still another case in which the inventor did not report information on the location of jobs. For each employer reported in her resume, one can extract all LinkedIn resumes (i.e. not just inventors, but any LinkedIn profile), who meet two conditions:

i. The resume reports the same employer name (i.e. the invidual reported to have worked for the focal employer);
ii. The resume reports information on the job location.

For example, given that inventor C reported *Art of Living* as one of her employers (see Table A3) but it did not report the job location, in order to estimate the probability that the location was in India, one can extract from LinkedIn all resumes that also reported *Art of Living* as an employer *and* reported the job location in the resume.

**Table A3: Inventor C, resume not reporting job location in employment history**

| Job title | Employer | Job location | Period |
|---|---|---|---|
| Founding Engineer | Ipcell | . | 1998-2000 |
| Sr Mgr, Software Development | Cisco | . | 2000-2008 |
| Sr Product Manager, Marketing, Business Development Manager | Cisco | . | 2008-2012 |
| State Coordinator, Texas and Director, YES for Schools | *Art of Living* | . | 2002-Present |
| Senior Consultant | Context BI | . | 2012-Present |

Table A4 illustrates the cross-country distribution of all LinkedIn resumes that reported an employment spell at *Art of Living* and also specified the location of the job. Out of 164 individuals who mentioned *Art of Living* as an employer *and* also reported job location, 90 of them (55% of total) were located in India. Hence, one can take this number as a rough estimate of the likelihood that the job location of inventor C in Table A3 was actually in India. Following this approach, one can also establish progressively looser thresholds of this probability, e.g. at 100%, 90%, 70%, 50%, and so on.

[11] Differently from schools (i.e. universities and other educational institutions) who have generally a unique location in a single country, large firms have operations, plants and subsidiaries in multiple countries. As a consequence, whereas geocoding schools through their names is likely to yield a reliable unique address, this is not the case for large firms. Put it differently, one cannot estimate the missing job locations by geocoding company names, such as Broadcom, Texas Instruments and so on.

For example, in the case illustrated in Table A3, the probability that the job location of inventor C when employed at *Art of Living* was in India is higher than 50%, but lower than 70%[12].

**Table A4: Cross-country distribution of all *Art of Living* employees reporting job location in their LinkedIn resumes**

| Country | Number of employees located in country | % of total |
|---|---|---|
| India | 90 | 54.9 |
| Canada | 16 | 9.8 |
| Germany | 13 | 7.9 |
| Australia | 12 | 7.3 |
| Other countries | 33 | 20.1 |
| Total | 164 | 100.0 |

As already mentioned above, in this paper (and in this appendix), we *rely exclusively* upon «self reported» job locations, namely on geographical information regarding job location that was *explicitly* reported in the resume (i.e. as in the case of Table A1). In this respect, our results need to be considered as conservative estimates of the true return migration.

*G. Estimating age and year of birth*

Using information on the starting year of education, we were also able to estimate the year of birth of the matched inventors. To this purpose, we assumed that inventors started a given educational programme at the most typical age for the corresponding educational level. More precisely, we assumed that the starting age was:

a) 14 for ISCED level 3 (upper secondary education)
b) 19 for ISCED level 5-6 (Bachelor or equivalent)
c) 23 for ISCED level 7 (Master or equivalent[13])
d) 25 for ISCED level 8 (Doctoral or equivalent)

To estimate the age of an individual, we followed the above list in a hierarchical order. Thus, for example, for an individual who reported to have started a high school cycle in 2000, we assumed that birth year was 1986 (i.e. =2000-14), irrespective of the other attainments achieved later in the life. Similarly, for an individual who *did not* report information on high school, but reported to have started a Bachelor in 2000, we assumed that she was 19 years old in that year and therefore was born in 1981 (i.e. =2000-19).

This approach has some obvious limitations. First, although probably correct on average, the estimated year of birth is greater than the actual one for those inventors who started a formal education programme later on in their life cycle and for those inventors who did not follow the typical sequence of studies, BSc→MSc→PhD. For example, the estimated year of birth of an inventor reporting to have started a PhD in 2000 (without reporting any other information on secondary

[12] The Art of Living Foundation is a volunteer-based, humanitarian and educational non-governmental organizations (NGO). It was founded in 1981 by Ravi Shankar. The Art of Living Foundation is spread over 156 countries. Its headquarter is in Bangalore. Not surprisingly, thus, the majority of individuals with a LinkedIn account reporting job location at Art of Living in their resumes declared a job located in India. Source: https://en.wikipedia.org/wiki/Art_of_Living_Foundation.

[13] In case the only information on educational attainment was related to MBA, we assumed a starting age at 27, as this looks the most typical age of MBA applicants.

education, BSc and MSc) is 1975, given that we assume that the average age of a first year PhD student is 25. To the extent that the inventor actually started her PhD at 30, her true year of birth is 1970 and as a consequence we are under-estimating his actual age[14]. Similarly, for an individual who started a MSc in 2000 *after* obtaining a PhD in 1995, we assume that birth year is 1977 (i.e. =2000-23), whereas her actual birth year is 1970 (i.e. =1995-25).

Second, estimating the year of birth is not possible for those inventors who either did not report any education information in the resume or whose *only* education attainment is *unclassified*, given that in this case there is not an age benchmark. In our sample of 10,839 inventors with a matched LinkedIn profile, there were 1,391 inventors whose resume did not report any information on education, and 1,585 inventors whose only educational attainment was *unclassified*.

Given our focus on educational level and education country to assess the extent of self-selection in return migration, we simply dropped from our sample the 1,391 inventors whose LinkedIn resume did not report any information on education. Regarding the 1,585 inventors whose only educational attainment was unclassified, we estimated the year of birth in the following way. For each of them, we extracted the application year of their first patent at the USPTO. From the sample of inventors for which we were able to estimate the year of birth based on education, we identified all inventors whose first patent application was made in the same year and we computed the average age of those inventors. Finally, we used this average age to estimate the year of birth. For example, given an individual whose first patent was made in 2000 and whose year of birth was unknown, we extracted all inventors whose first patent was in 2000 and for which the year of birth was estimated using educational attainments. As the average age of inventors whose first patent was made in 2000 is 32, we assumed that the year of birth of the focal inventor is 1978 (i.e. 2000 - 32). Once again, although probably correct on average, this approach is likely to be prone to some measurement errors. To this purpose, section K below reports descriptive statistics on the distribution of the inventors in our final sample by year of birth and age at the first patent.

### *H. Dropping incomplete and inconsistent profiles*

An extensive follow-up checking was performed in a semi-automated way to improve the accuracy of our matching between inventors and LinkedIn profiles. In the first place, given our focus on educational attainments to assess the extent of self-selection in return migration, we dropped from the list of 10,839 matched inventors, 1,391 inventors whose LinkedIn resume did not report any information on education. Second, we dropped 279 inventors whose estimated age at the first patent was either less than 21 (i.e. the age at the completion of a short-cycle of tertiary education) or greater than 66 (i.e. age at retirement) or whose first patent was granted before the first reported education title. These cases were dropped because they are likely to be false positives, namely matched to the wrong LinkedIn profile. In addition to this, other 187 inventors were dropped from our sample as their LinkedIn resume did not report any employment history.

Table A5 summarizes the outcome of our matching exercise between USPTO inventors and LinkedIn resumes. Out of 24,017 inventors of Indian origin, we could match 10,839 unique LinkedIn profiles. For 1,857 of them (i.e. =1,391+279+187), however, the information contained in the LinkedIn resumes was either incomplete or inconsistent and the corresponding inventors were dropped from our sample. Overall, our final sample consists of 8,982 inventors, which represent 37.4% of all Indian inventors.

[14] The opposite case of individuals starting formal education at an age lower than the typical age for a certain education level is arguably less common.

**Table A5: Indian Inventors in ICT: LinkedIn matching outcome**

| | Number | % of all Indian inventors |
|---|---|---|
| Indian inventors matched with a LinkedIn profile | 10,839 | 45.13 |
| *of which:* | | |
| Profile has no info on education attainments [a)] | 1,391 | 5.79 |
| Profile has no info on employment history [b)] | 187 | 0.78 |
| Profile has inconsistent info on age [c)] | 279 | 1.16 |
| **Profile has complete info on education and employment and consistent info on age** | **8,982** | **37.40** |
| Indian inventors not matched with a LinkedIn profile | 13,178 | 54.87 |
| All Indian inventors | 24,017 | 100.00 |

[a)] the matched profile does not contain any information on the educational attainments of the inventor;
[b)] the matched profile does not contain any information on the employment history of the inventor
[c)] the age of the inventor, estimated on the basis of the educational attainment, at the time of the first patent is either lower than 21 or greater than 66, or the first patent was applied before the first reported education title.

## *I. Accuracy of match: precision and recall*

In order to assess the accuracy of our matching, we exploited the fact that some inventors report in their LinkedIn resumes information on the patents made. In particular, we could identify 1,049 cases of Indian inventors for whom the match with LinkedIn was "certain", as the inventor herself reported information on the patents made in the LinkedIn profile. Using this subset, we were able to assess the rate of errors generated by our matching algorithm. For this test, we restricted attention to the 8,982 matched inventors for whom we have complete education and employment history and consistent information on age. In particular, we computed two types of statistics.

First, we evaluated the rate of "false positives" (Type 1 error). They correspond to those cases in which an inventor is matched by our algorithm to a "false" LinkedIn profile, i.e. the algorithm assigns the inventor to a profile, which is not the correct one. More specifically, we computed the so-called "precision rate", defined as:

$$\frac{\#\text{ of true positives}}{\#\text{ of true positives} + \#\text{ of false positives}} \qquad (1)$$

Of the 1,049 "certain" matches, our matching algorithm was able to assign a LinkedIn profile to 838 cases. Of them, 808 were "true positives" (i.e. the matched profile was the correct one) and 30 were "false positives" (i.e. the matched profile is a false one). Overall, the precision rate is equal to 808/838=0.964. This means that, when our algorithm assigns a LinkedIn profile to an inventor, it does so correctly in about 96.4% of cases.

Secondly, we evaluated the rate of "false negatives" (Type 2 error). They correspond to those cases in which our algorithm fails to find a match even when there is a valid one, i.e. the inventor has a LinkedIn profile, but our algorithm is unable to match it. More specifically, we computed the so-called "recall rate", defined as:

$$\frac{\#\text{ of true positives}}{\#\text{ of true positives} + \#\text{ of false negatives}} \qquad (2)$$

Of the 1,049 "certain" matches, our matching algorithm was able to assign a correct LinkedIn profile to 808 cases (true positives), but it failed to find a valid match in 241 cases (false negatives). Overall the recall rate is equal to 808/1,049=0.77. This means that our algorithm is able to find a valid match for about 77% of all inventors who have a LinkedIn profile.

*J. Comparison between matched and unmatched inventors*

A further control needed to ensure the representativeness of our sample is comparing the inventors matched with a LinkedIn profile and the inventors not matched. To this purpose, we restricted again attention to the 8,982 matched inventors for whom we have complete education and employment history and consistent information on age[15].

In particular, we carried out three types of tests. In the first place, we tested to what extent matched and unmatched inventors differ in terms of patent productivity. To this purpose, we carried out a simple t-test on the average number of patents produced by the inventors in the two groups. Results reported in Table A6 show that patent productivity of matched and unmatched inventors does not differ in a statistically significant way.

**Table A6: Average number of patents of matched and unmatched inmventors**

| Matched | | | Unmatched | | | |
|---|---|---|---|---|---|---|
| Obs | Mean | Std.Dev. | Obs | Mean | Std.Dev. | t-test (p-value) |
| 8982 | 7.33 | 15.54 | 15035 | 7.29 | 18.11 | 0.185 (0.853) |

Second, we assessed to what extent the sample of matched inventors includes more recent cohorts, under the assumption that younger people have more incentives or simply a higher propensity to register a LinkedIn profile than relatively older people. Ideally, we would like to compare the age profile across the two subsets. However, while this can be somehow estimated from education data for the matched inventors, there is no way to retrieve this information for the unmatched ones. As a second best solution, therefore, we computed for each subset of inventors the distribution by application year of the first patent at the USPTO. As shown in Figure A1, although the two distributions appear quite similar, the sample of matched inventors seems to include individuals with a relatively more recent patenting history than the sample of unmatched inventors. A Kolmogorov-Smirnov two-sample test (0.256, p-value 0.10) allows to reject the hypothesis that the distributions of the two samples are the same. As a consequence, keeping in mind that the date of the first patent is not perfectly correlated with the age of the inventor, we can reasonably conclude that our sample of matched inventors includes relatively younger individuals.

Third, we tested to what extent there might be a different propensity to have a LinkedIn account across groups of inventors. For instance, inventors might be more likely to sign up to keep in touch if they are away from US, or conversely more likely to do it if in the US because they need to do it for work. In order to test this type of conjectures, we split the population of Indian inventors into four mutually exclusive groups:

[15] Note that the unmatched cases include inventors who may actually have a LinkedIn profile, which for various reasons we have been unable to match. As noted above, we estimate a recall rate of 77%, meaning that we are unable to match the LinkedIn profile for about 23% of all those who actually have one. For this reason, what we are assessing here, strictly speaking, is not the probability that an inventor has or has not a Linkedin profile, but the probability that the inventor has been included in our final sample. At the same time, it is also correct to say that the majority of the inventors in the unmatched subset is composed of individuals that are truly absent from LinkedIn.

1. Inventors who patented only in India
2. Inventors who patented only in the US
3. Inventors who patented both in India and in the US (and possibly other countries)
4. Others

Table A7 reports the number of inventors in each group as well as the fraction of inventors with a matched LinkedIn profile.

**Figure A1: Distribution of LinkedIn matched and unmatched inventors by application year of the first patent at the USPTO**

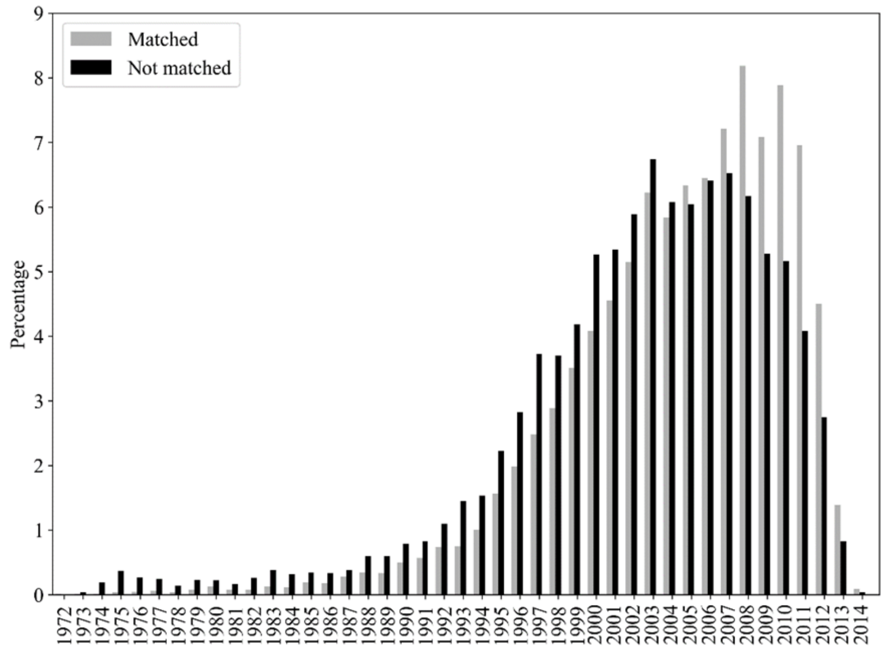



**Table A7: Fraction of inventors with a matched LinkedIn profile**

| Group | Number | Number with a matched LinkedIn profile | % with a matched LinkedIn profile |
|---|---|---|---|
| 1. Inventors who patented only in India | 4,324 | 2,003 | 46.3 |
| 2. Inventors who patented only in the US | 17,392 | 6,088 | 35.0 |
| 3. Inventors who patented both in India and in the US | 1,457 | 593 | 40.7 |
| 4. Others | 844 | 298 | 35.3 |
| All Indian inventors | 24,017 | 8,982 | 37.4 |

A simple z-test of proportions indicates that inventors that patent exclusively in India have a significantly higher probability of being matched with a LinkedIn profile than both inventors patenting exclusively in the US (z-score=13.776, p-value=0.000) and inventors that patent both in

India and in the US (z-score=3.732, p-value=0.000). This evidence is consistent with the use of LinkedIn by inventors resident in India to signal their skills and "promote" themselves in the job market. Moreover, it is also consistent with the broader pattern of LinkedIn usage by country. As a matter of fact, with 35 million accounts India is second only to the US (with 128 million profiles) in terms of registered members of Linkedin (as of the first quarter of 2016)[16].

*K. Age distribution of inventors*

In this section, we provide some descriptive statistics on the age distribution of inventors in our sample. This is once again relevant to assess the reliability of our sample, given that age was estimated based on the education attainmnents of inventors. Also in this case, we focus attention on the 8,982 inventors included in our sample.

As described above (section F), age of inventors was estimated on the basis of ISCED education levels. When the educational attainment could not be classified in any of the ISCED levels, we estimated age on the basis of the average age at the first patent. Figure A2 reports the percentage distribution of inventors according to the way in which age was estimated. For the vast majority of inventors in our sample (64%), age was estimated on the basis of the start year of the BSc, as this was the first educational attainment reported in their resume. For an additional 12% of inventors age was estimated on the basis of the start year of MSc (as this was the first educational attainment reported in their resume). For just 4% of all inventors the source of information to estimate age was the start year of the secondary school, as this is a type of information that relatively few individuals mention in their resumes. Moreover, for about 16% of all inventors whose educational attainment could not be classified in any ISCED level, we were forced to estimate age by taking the average age of inventors at the time of their first patent.

**Figure A2: Percentage distribution of inventors in the final sample by source of information used to estimate year of birth (8,982 obs.)**

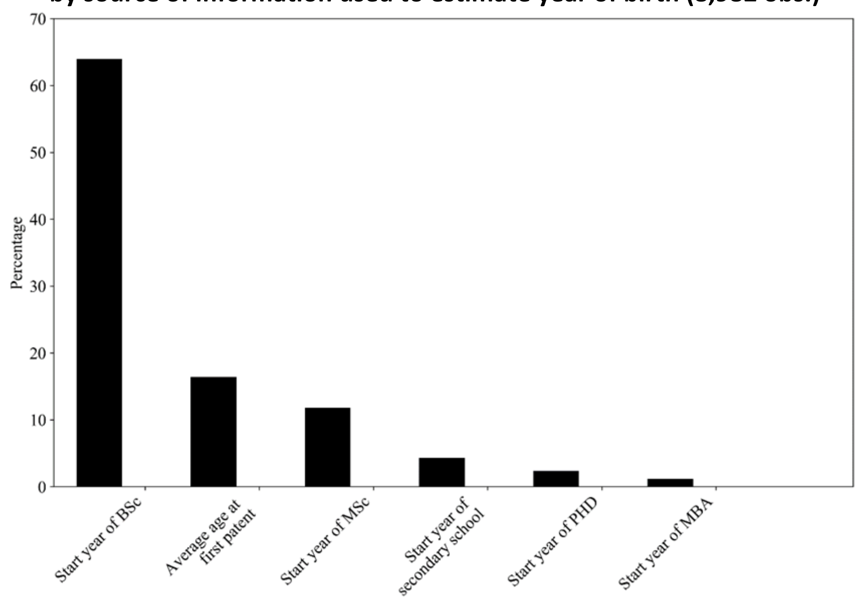



Figure A3 shows the percentage distribution of the inventors included in our final sample by estimated year of birth. The bulk of inventors are born between early 1970s and mid-1980s, with a modal value in 1978. Around 67% of all inventors in our sample are born between 1970 and 1985,

[16] https://www.statista.com/statistics/272783/linkedins-membership-worldwide-by-country/

whereas an additional 25% are born between 1960 and 1969. Overall, this evidence suggests, as already noted above, that our sample of inventors consists of relatively young individuals (i.e. the modal inventor is 40 years old in 2018).

**Figure A3: Frequency distribution of inventors in the final sample by estimated year of birth (8,982 obs.)**

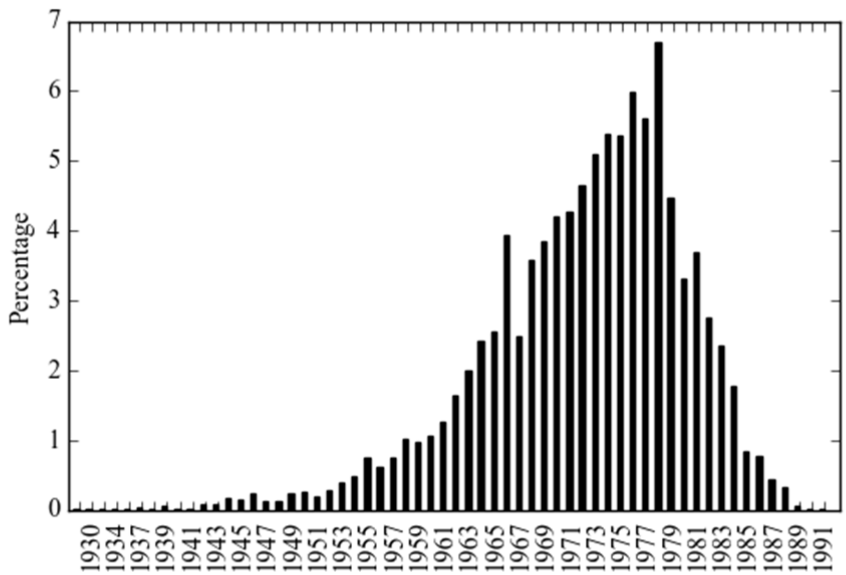



Figure A4 illustrates the percentage distribution of inventors in our sample by age at the first UPSTO patent application. The modal age is 32: 18% of all inventors made their first patent application at this age. More generally, the distribution is remarkably concentrated between 25 and 35: inventors in this age range account for 77% of all inventors in our sample[17]. These results are broadly consistent with those reported by Jones (2009).

Finally, Figure A5 shows the average age at the time of the first patent by year of the first patent. Some variation is observed for older cohorts (i.e. inventors who made their first patent in the ‘70s and in the ‘80s), yet these cohorts include very few individuals (see above Figure A1). Apart from that, no clear pattern is detectable in the data. The average age at the first patent of individuals who made their first patent in the ‘90s and ‘00s, which represent the bulk of our sample, was around 32 with no significant variation.

[17] In evaluating the peak at 32, note however that the distribution includes also 1,473 inventors for whom age was estimated on the basis of the average age at the time of the first patent, which peaks around that age for most cohorts.

**Figure A4: Percentage distribution of inventors in the final sample by age at the first patent (8,982 obs.)**

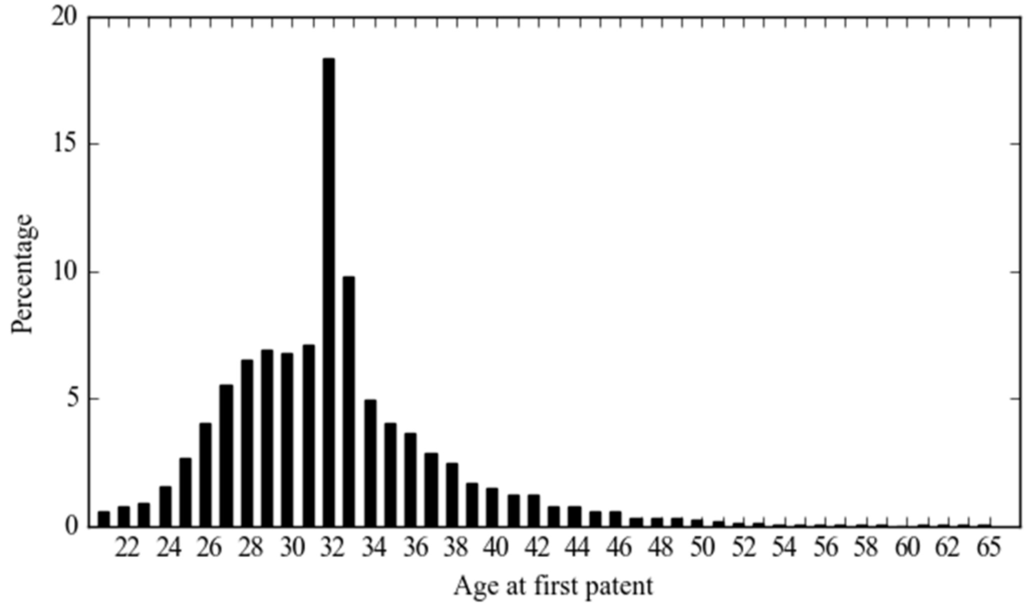
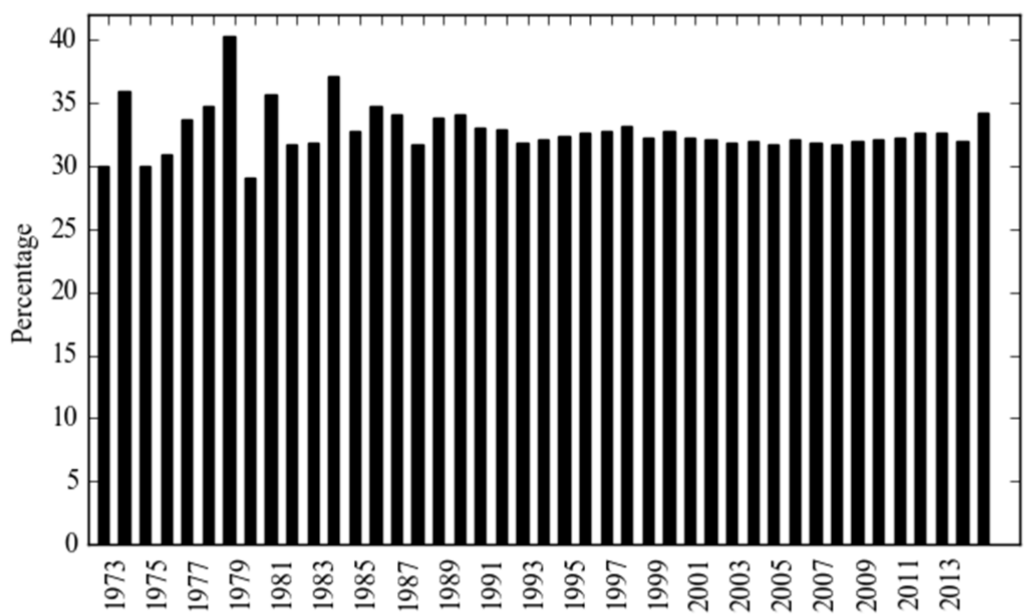



**Figure A5: Average age at the first patent by year of first patent application (8,982 obs.)**




*L. Coding migration events and migrant inventors*

In this section, we describe the methodology used to code migration events and to identify migrant inventors. To this purpose, we exploited three types of information on location of inventors:

1. Country of address reported in patents (from USPTO)
2. Country of job locations reported in the resume (from LinkedIn)
3. Country of educational institutions where education was attained (from LinkedIn)

As far as 2. and 3. are concerned, we already illustrated above the way in which location was extracted from LinkedIn records. Information on 1. was obtained from Patentsview.

In order to identify migrant inventors, we proceeded as follows. In the first place, we split the sample of 8,982 inventors in two mutually exclusive subsets. The first subset includes inventors who, *at any time* in their career, either made a patent, were educated or «self reported» a job location in India. The second subset includes inventors who never made patents, were educated or reported a «self reported» job location in India. Note that, as our sample consists of inventors, the second subset includes Indian-origin inventors that for sure made patents in other countries, but India.

The first subset comprises potential migrants, whereas we label inventors in the second subset as «false positives» (with respect to migration). The reason for this labelling is the following: these inventors have an Indian origin, have made patents outside India, but did not leave any trace of activity in India, particularly with respect to education. Even though they might include true migrants, they might also consist of second-generation Indians born and educated outside India. Out of 8,982 inventors, we labelled 1,445 of them as «false positives» and we dropped them from our sample. As argued above, some of these inventors might be true migrants and not second-generation Indian inventors. Yet, we cannot discriminate the former from the latter on the basis of available information. For example, an inventor who in her resume reported only a PhD attained in the US, did not report any job location in India, and never made patents in India is considered as a «false positive», even though she might have achieved a BSc in India without reporting such educational attainment in her resume.

With this caveat in mind, our sample of potential migrants, after dropping 1,445 «false positives», is reduced to 7,537 inventors. This sample was further split into two mutually exclusive subsets. The first subset includes inventors who never reported in their career an educational attainment, an employment (i.e. job location) or a patent made in a country different from India. The second subset is defined in a complementary way and it includes inventors who, *at any time* in their career, either made a patent, were educated or reported a job location outside India. We label the first subset as «non migrants» to indicate that on the basis of available information these inventors were active only in India and did not migrate during their career. Out of 7,537 potential migrants, 1,672 were labeled as «non migrants» and, given our focus on return migration, were dropped from the sample. Our sample of migrants thus consists of 5,865 inventors.

Our sample of «migrant» inventors was further split into two mutually exclusive groups reflecting the motives for which individuals migrated. In particular, we distinguished two major reasons for migration: education and work. Accordingly, the first subset includes inventors whose *first event* outside India was the attainment of an educational title. Similarly, the second subset includes inventors whose *first event* outside India was either a patent or a job in a country different from India. Overall, out of 5,865 migrant inventors, we identified 4,161 «education migrants» and 1,704 «work migrants».

Figure A6 summarizes the process followed to identify the sample of migrant inventors.

**Figure A6: Identification of migrant inventors**

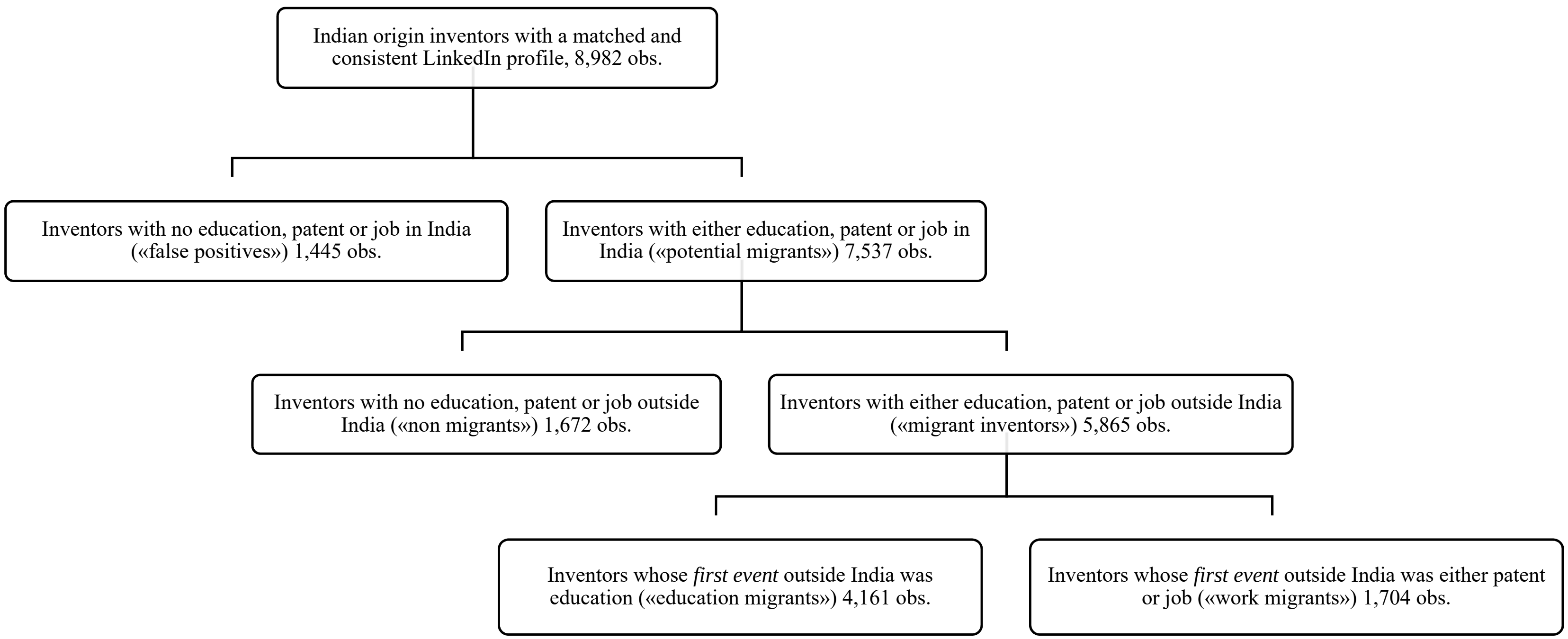
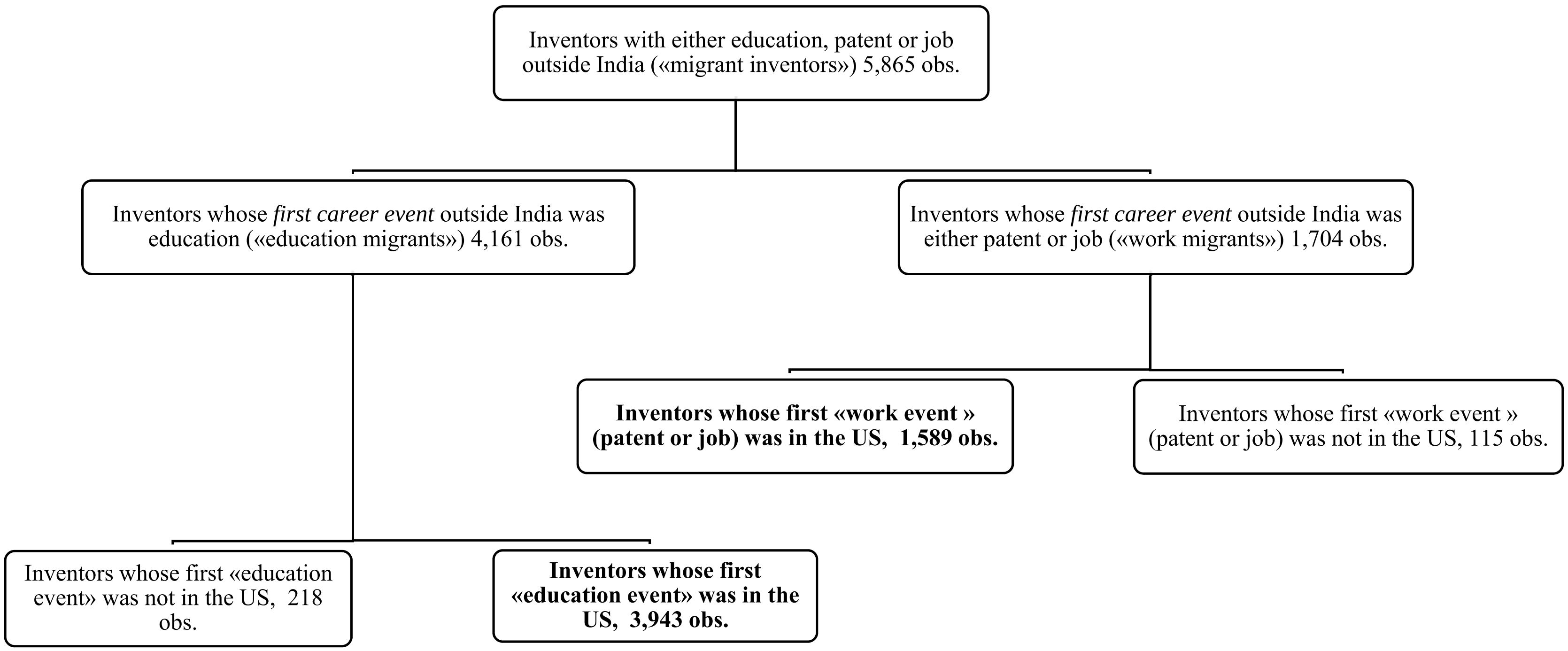


**Figure A7: Identification of migrants to the US**



«Education migrants» can be further split into two distinct categories: i) migrant inventors that never made any patent outside India (317); ii) migrant inventors that patented outside India *after* being educated abroad (3,844). Similarly, «work migrants» can be further split into two distinct categories: i) migrant inventors that did not take any education outside India (1,253); ii) migrant inventors that took education outside India *after* either patenting or taking a job outside India (451).

As a final step, we further split the sample of 5,865 migrant inventors into two mutually exclusive groups. The first group comprises migrant inventors whose first event outside India was in the US. The second group consists of migrant inventors whose first event outside India was in a country different from the US. Figure A7 illustrates this further selection step. Out of 5,865 migrant inventors, 5,532 (i.e. 94% of all migrant inventors) are defined as «migrants to the US», whereas 333 are defined as «migrants to other countries». Of the 5,532 migrants to the US, 3,943 migrated for education motives, whereas 1,589 migrated for work reasons. In what follows, we focus on the subset of «migrants to the US».

### *M. Coding migration year*

Once coded migration events and identified migrant inventors (to the US), we defined the year in which migration took place. The identification of the year of migration differs according to the migration motive. For inventors whose migration motive was education, we assumed that migration occurred at the beginning of the first education programme undertaken by the inventor in the US. For example, for an inventor whose first event outside India was a MSc in the US started in 1981, migration year was set equal to 1981.

For inventors whose migration motive was work, the migration year was similarly defined as the date of the first event occurring outside India. As for «work migrants» two possible events, i.e. patent or employment, can mark the starting of migration, the year of migration was defined accordingly. Thus, for inventors whose first event outside India was a patent made in the US, migration year was set equal to the application year of the first patent in the US. Instead, for inventors whose first event outside India was an employment in the US, migration year was set equal to the starting year of the corresponding employment spell. Out of 1,589 migrant inventors to the US for work reasons, the first event in the US was a patent for 1,280 (i.e. 81%) of them.

It is important to emphasize the asymmetry in estimating the migration date for inventors whose migration motive was education as compared to inventors who migrated for work reasons. Whereas this estimate is likely to be fairly accurate for inventors who moved for education reasons, this is less likely to the case for inventors who moved for work reasons. As noted above, for the majority of the latter the first event in the US that we could detect on the basis of available information was a patent application. Yet, it might be that these inventors moved to the US before this application and we are simply unable to spot the time of the move because inventors' resume does not report sufficiently detailed and accurate information to date the migration event more precisely.

For descriptive purposes, Figure A8 reports the percentage distribution of migrant inventors to the US, who migrated for education motives, by migration year. Most of the migration for these reasons was concentrated in the two decades 1990-1999 and 2000-2009. Of all Indian inventors that migrated to the US for education reasons, 44% of them did it in the period 1990-1999, and 33% did in the period 2000-2009.

Figure A9 illustrates the percentage distribution of migrant inventors to the US whose first event in the US was a patent. The distribution appears quite different from the one observed for education migrants. Only 16% of inventors of the inventor who migrated in this way did so in the decade 1990-1999, whereas 68% of them did it in the decade 2000-2009. Finally, Figure A10 shows the percentage distribution of migrant inventors to the US whose first observable event was an employment. Keeping in mind that the number of inventors in this subset is lower than in the other two cases, one can notice a sort of cyclical pattern. A first peak is observed in the years from 1997 to 2001 (corresponding to the development of the dot com economy), whereas a second peak is observed in the years 2011 and 2012.

Figure A11 reports the percentage distribution of the 3,943 migrants for education motives by age at migration. Around 84% of all migrant inventors for education motives had an age at migration comprised between 23 and 27, suggesting that the vast majority of those who moved to the US for this reason went there to attain either a MSc or a PhD.

**Figure A8: Percentage distribution of «education migrants to the US» by migration year (3,943 obs.)**

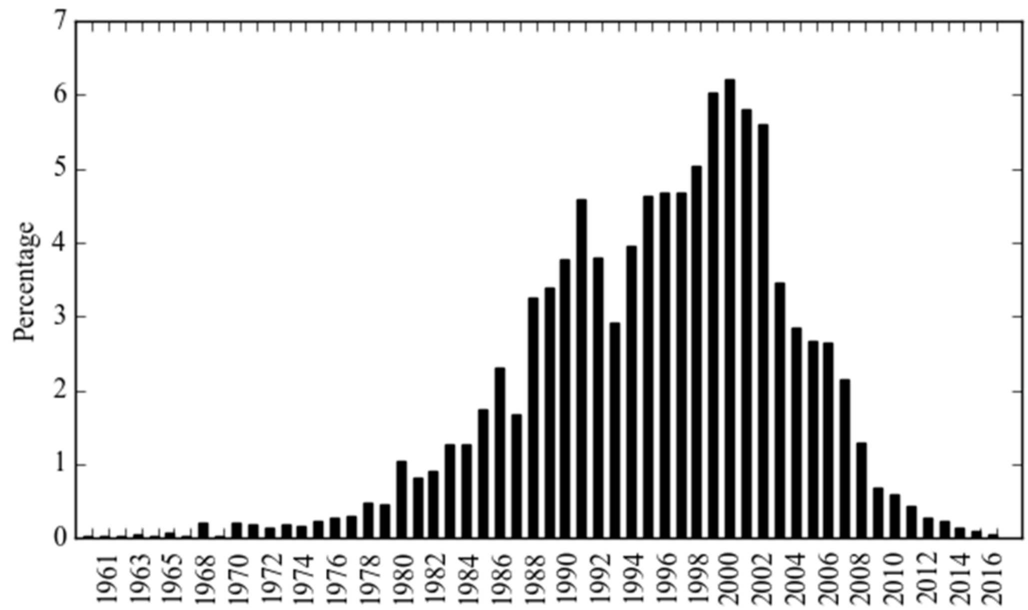


**Figure A9: Percentage distribution of «work migrants to the US» whose first event in the US was a patent by migration year (1,280 obs.)**

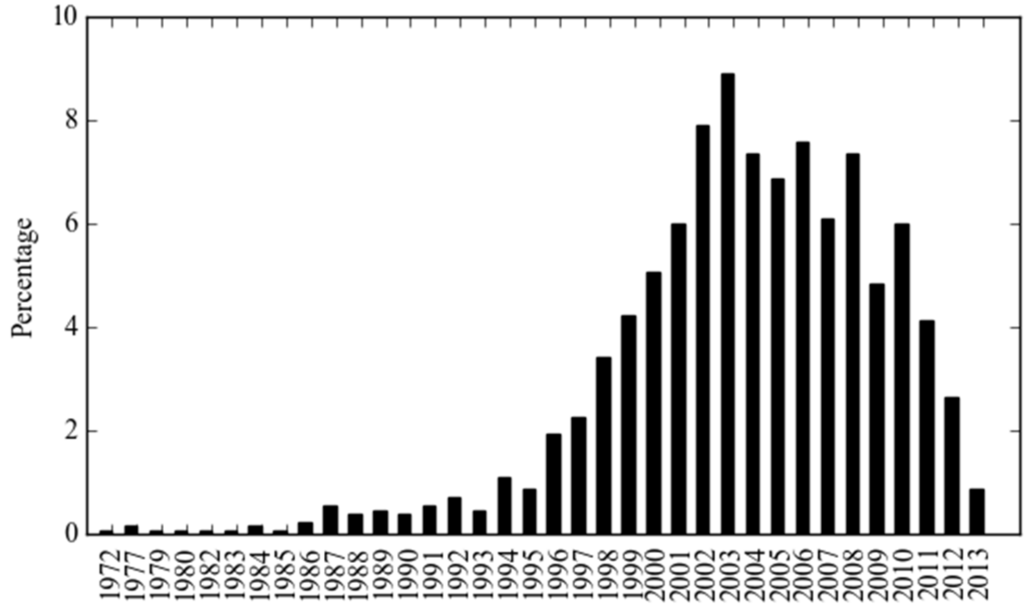



**Figure A10: Percentage distribution of «work migrants to the US» whose first event in the US was an employment by migration year (309 obs.)**

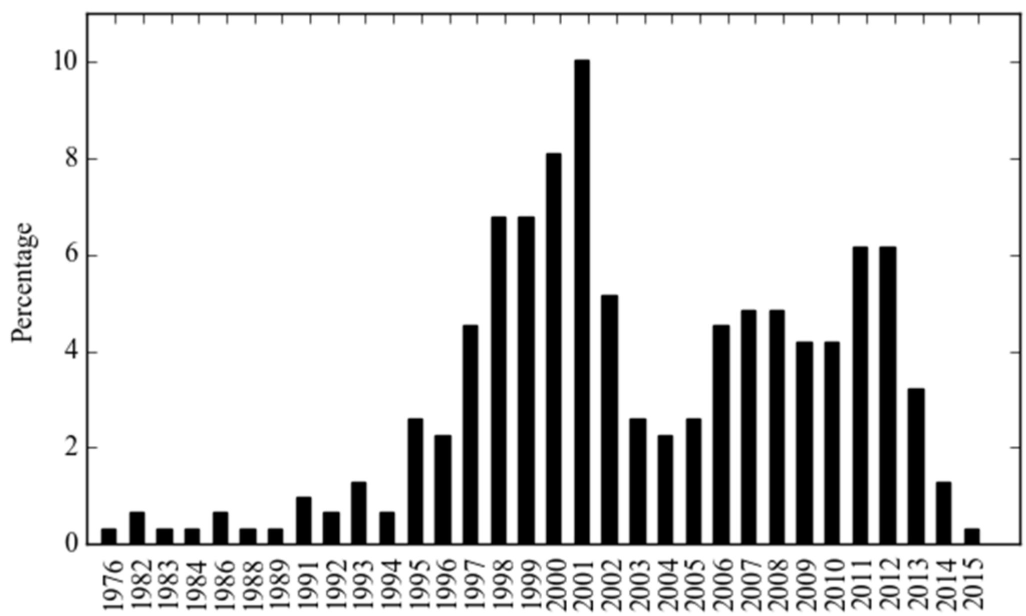
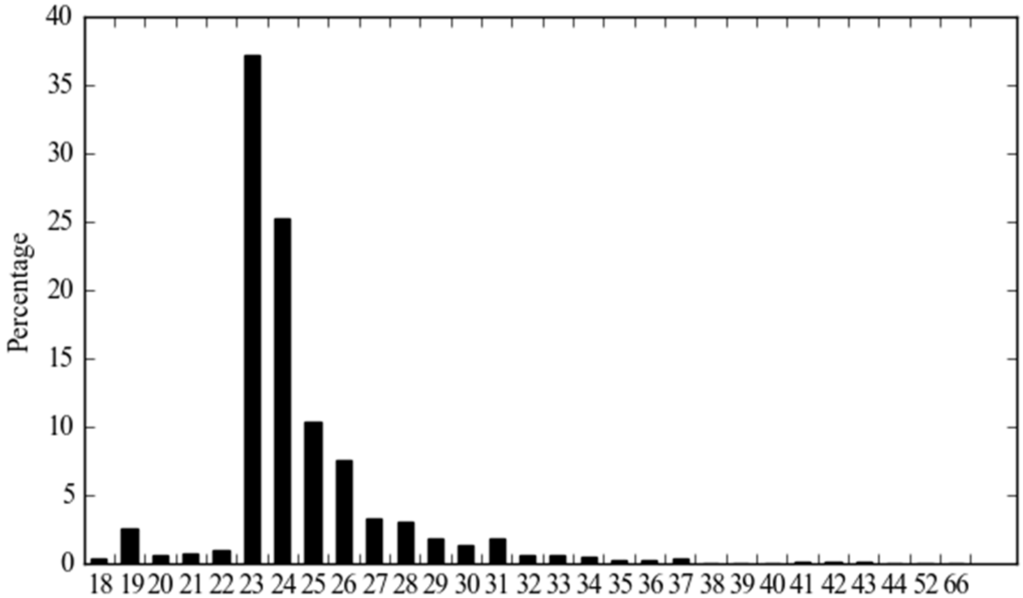


**Figure A11: Percentage distribution of «education migrants to the US» by age at migration (3,943 obs.)**




Note that, given the way in which we coded the year of birth, the age at migration should in principle display only three values, i.e. 19, 23, 25 and 27, depending on the type of education level used for the estimation. Yet, due to misreporting errors in the data, the age at migration is not necessarily the same for all inventors who moved to the US to attain a certain education level. To illustrate this issue, consider the case of the inventor shown in Table A8. The resume reported three educational attainments. For each of them, the resume reported the start but not the end year. As the starting date of the BSc in India was in 1988, we accordingly estimated that the inventor was born in 1969 (i.e. =1988-19). Yet, the resume also reports that the inventor started a MSc degree programme in the US in 1990. Hence, the inventor's migration year was set equal to 1990 and her age at migration was equal to 21 (i.e =1990-1969), whereas the age at the start of the MSc is in general equal to 23.

**Table A8: Inventor D, migrant to the US for education motives**

| University | Start year | End year | Degree |
|---|---|---|---|
| Indian Institute of Technology, Kharagpur | 1988 | . | B.Tech. (Honors) in Computer Science and Engineering |
| The University of Texas at Austin | 1990 | . | M.S. in Computer Science |
| The University of Texas at Austin | 1995 | . | Ph.D. in Computer Science |

Despite all our efforts to carefully clean and check raw data, a few errors, inconsistencies and more generally noise are still present in the data. Notwithstanding this, we believe that the general pattern reported in Figure A11 is reassuring about the quality of the data used in our analysis. Moreover, some deviations from the general pattern might be due to genuine devitations from the typical educational pattern. This is particularly the case of inventors whose age at migration is greater than 27. Consider for example the inventor reported in Table A9. The inventor started a BSc in India in

1949 and her birth year was accordingly estimated as 1930 (i.e. =1949-19). In 1963, when she was 33 years old, she started (i.e. migrated to) a PhD in the US.

**Table A9: Inventor E, migrant to the US for education motives**

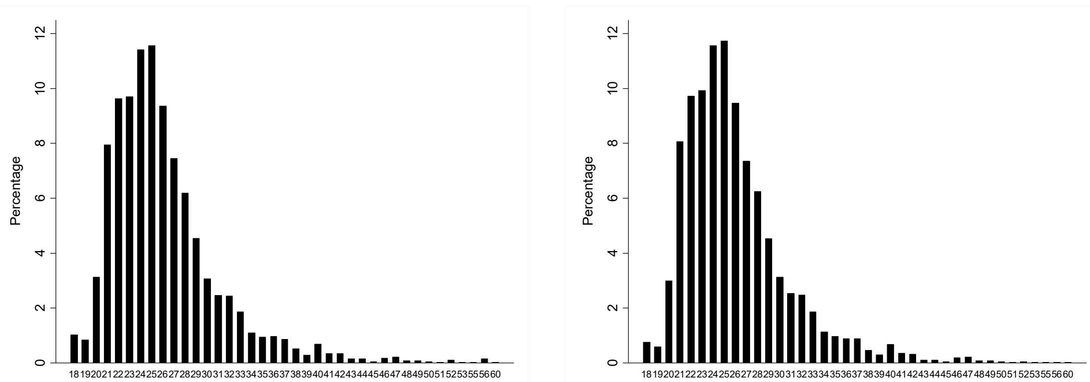
| University | Start year | End year | Degree |
|---|---|---|---|
| University of Lucknow | 1949 | 1952 | B.Sc. Physics |
| Brooklyn Polytechnic | 1963 | 1968 | Ph.D. Electrical Engineering |

Figures A12a,b look at the distribution of the age at migration of, respectively, all Indian immigrants and Indian immigrants with college, master or PhD education, both employed in STEM occupations, for the migration cohorts 1990s and 2000s. The raw data are the pooled samples of the American Community Survey 2000 and 2010 – extracted from IPUMS USA. The modal values are 24 and 25. 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 A13). The most numerous group is the one at ages 25-34 years old, followed by the group 35-44.

**Figure A12: Percentage distribution of Indian immigrants from ACS 2000 & 2010, employed in STEM occupations**

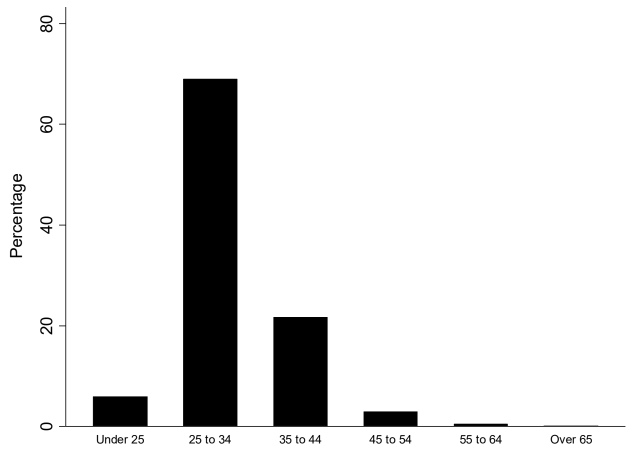
**A12a : All Indian immigrants**

**A12b : Indian immigrants with college, master or PhD education**




Note: STEM occupations include Computer and Mathematical Occupations, and Architecture and Engineering Occupations The raw data to build these figures are the pooled samples of the American Community Survey 2000 and 2010 – extracted from IPUMS USA.

**Figure A13: Percentage distribution of H-1B petition filings by age, 2007-2017, all origins**




**Source:** US Citizenship and Immigration Services

Figure A14 reports the age at migration for the work migrants to the US. Not surprisingly, we observe a substantial difference with the distribution of age at migration of inventors who migrated for education reasons. Work migrants tend to be significantly older than education migrants at the time of migration. In comparing the two distributions, however, one should keep in mind that our ability to estimate the year of birth based on educational attainment was lower for work migrants than for education migrants. For a substantial fraction of the latter, we had to estimate the year of birth as the average age at the time of the first patent (see discussion above and Figure A2), which is necessarily a rather crude estimate. Out of 3,943 education migrants to the US we estimated age on the basis of the average age at the time of the first patent for 35 inventors (i.e. less than 1%). On the other hand, out of 1,589 work migrants to the US we had to estimate age on the basis of the average age at the time of the first patent for 194 inventors (i.e. about 12% of them).

As a robustness check, Figure A15 plots the percentage distribution of work migrants by age at migration, excluding the 194 inventors for whom age was estimated as the average age at the time of the first patent. Once again, we observe that the modal value is at an age of 32 and that the distribution appears concentrated on older ages than the distribution of education migrants.

**Figure A14: Percentage distribution of «work migrants to the US» by age at migration (1,589 obs.)**

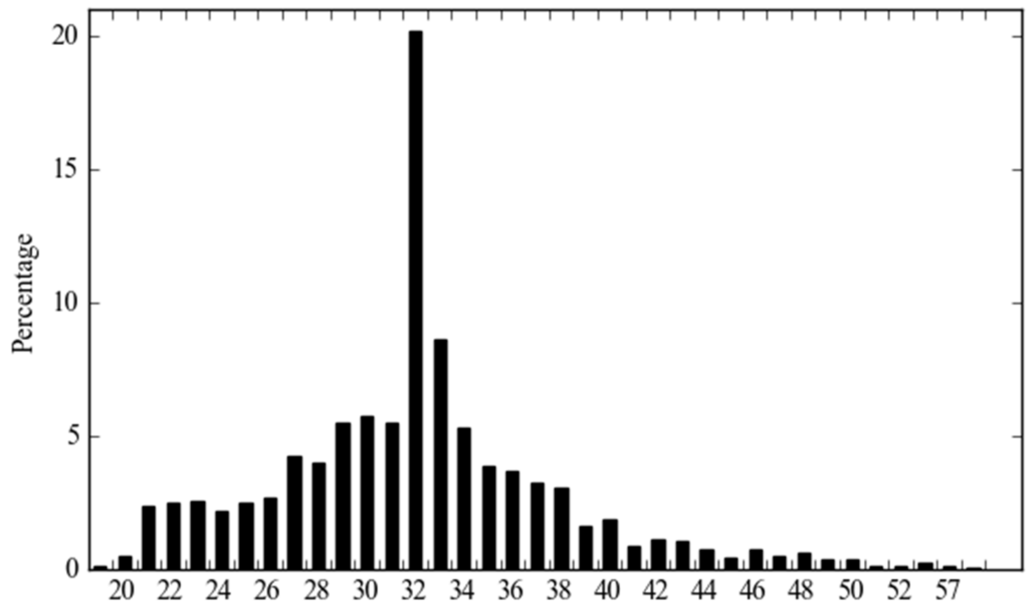
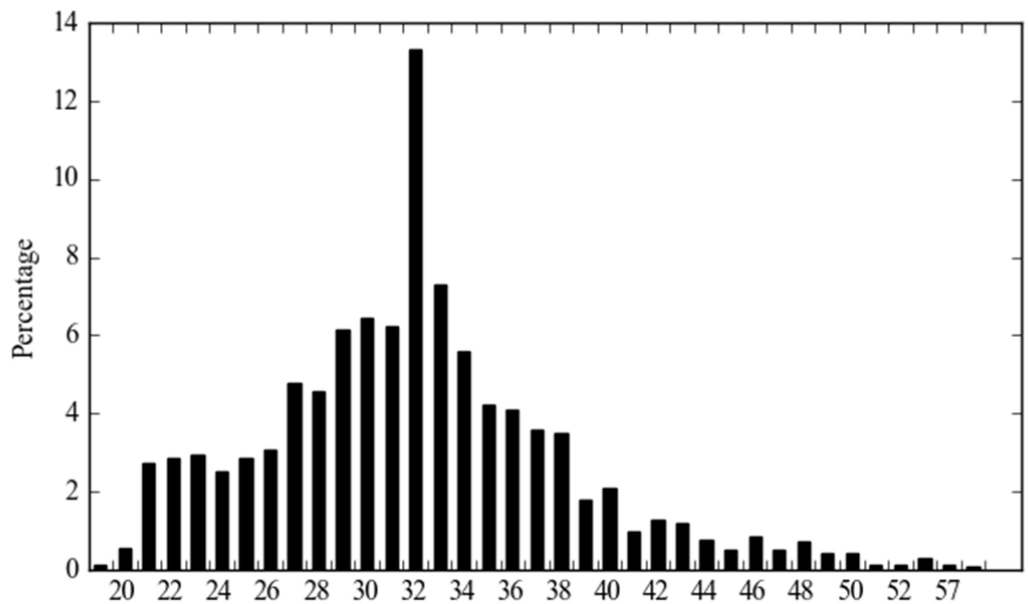



**Figure A15: Percentage distribution of «work migrants to the US» by age at migration excluding inventors for whom age was estimated as average age at the first patent (1,395 obs.)**




*N. Coding return migration and return year*

The final methdological step consisted of coding the events of return migration. To this purpose, we exploited again information on the location of the three types of activities included in our data, i.e. education, patenting, and employment. For each of the 5,532 «migrants to the US», we identified «returnees to India» by looking at their career path. An inventor was defined as a «returnee to India» when she either made a patent, attained education or reported a job located in India in a year *following the one of migration to the US*. Return year was set equal to the date of the first event (if any) taking place in India after migration. For example, for an inventor who migrated to the US in 1990 and subsequently made a patent in India in 1995, the return year was set equal to 1995.

*O. Indian-born and second-generation Indians*

As explained above, we defined Indian-born inventors or «potential migrants» as those inventors who met the following conditions:

1) given and family names were classified as having an Indian-origin, and
2) *at any time* in their career, the inventors either made a patent, were educated or «self reported» a job location in India.

The condition that the inventors had to show some experience in India *at any time* during their career might introduce some false positives in the sample of Indian-born «potential migrants». For instance, consider the case of a second-generation Indian inventor born and educated in the US, who at some point starts working or patenting in India. This individual will be considered as a potential migrant, whereas in fact she is not. Although this is arguably a relatively uncommon case, we cannot completely rule out it. A potential solution to this issue would be to include in our sample only inventors who attained a BSc in India. The problem with this solution is that many Indian-born inventors *do not* report information in the resume on the BSc (and thus where this has been attained), including only information on the MSc or higher degree often attained in the US. According to the logic explained above, we should treat them as second-generation migrants and exclude them from the pool of potential migrants, whereas in fact they are. In other words, we would generate many false negatives reducing our sample size.

Rather than trying to solve the issue, we keep our definition and in this section we show that the concern illustrated above is likely to be rather limited. To this purpose, we focus attention on the 5,532 migrants to the US. For each of them, we attempted to reconstruct their career path *before* migrating to the US. In particular, we assessed to what extent the sample of migrants to the US includes inventors who were *active* in some way in India *before* the year of migration to the US. Specifically, for each inventor we recorded the year of the *first event* (i.e. education, patent or a self-reported employment) in India and we compared it with the year of migration to the US. To the extent that the first event in India was preceding the year of migration, we can exclude that the inventor is a second-generation Indian who at some point returned to India. Out of 5,532 migrants to the US 5,230 (i.e. 94.5% of all migrants) were active in India in the sense specified above (i.e. they either made a patent, attained education or had a job) *before* the migration event.

Note that one should not consider the other 302 inventors, for which we have no trace of educational or professional activity in India *before* the migration, as false positives, namely second-generation Indian inventors who went to India for professional reasons. Rather, the majority of them are likely to be genuine migrants, who simply did not record in their resume any experience in India made *before* the choice of migration. The case discussed above of the inventor attaining a BSc in India, without reporting it in the resume, and recording only the MSc or the PhD attained in the US fits this picture.

To dig more into this problem, we further split the 302 potential migrants without any trace of experience in India *before* migration: of them, 122 are returnees to India, while 180 are inventors who attained in India education for which we could not define the start and end dates, which provide the crucial information to define *before* migration events. Of the 180 inventors who attained education in India at some unknown date, 4 attained a BSc, 17 a MSc, 4 a PhD, and 158 other unclassified education titles[18]. Moreover, a casual inspection reveals that most of the unclassified education titles relate to secondary school or college level education. Of the 122 returning inventors, 7 attained a BSc, 13 a MSc, 2 a PhD, and 33 other unclassified education titles in India at some unknown date. Overall, of the 120 returning inventors, 49 got some education in India at some unknown date. As it is quite reasonable to assume that second-generation US born Indian inventors are unlikely to go and get any education in India, it is likely that 229 (=180+49) out of 302 potential false positives are actually genuine migrants. Following this logic, the potential problem of having false positives in the sample of potential migrants is restricted to only 73 individuals, i.e. around 1.3% of all migrants to the US.

[18] Please note that the sum is greater than 180 since some inventors reported multiple educational attainments.

## List of US public ICT companies used in the paper

| IDX | Company name | IDX | Company name |
|---|---|---|---|
| 0 | 3COM CORP | 89 | JUNIPER NETWORKS INC |
| 1 | ACTEL CORP | 90 | L-3 COMMUNICATIONS HLDGS INC |
| 2 | ADC TELECOMMUNICATIONS INC | 91 | LATTICE SEMICONDUCTOR CORP |
| 3 | ADOBE SYSTEMS INC | 92 | LEVEL 3 COMMUNICATIONS INC |
| 4 | ADTRAN INC | 93 | LEXMARK INTL INC -CL A |
| 5 | ADVANCED MICRO DEVICES | 94 | LINEAR TECHNOLOGY CORP |
| 6 | AFFYMETRIX INC | 95 | LORAL SPACE & COMMUNICATIONS |
| 7 | AGERE SYSTEMS INC | 96 | LSI CORP |
| 8 | AGILENT TECHNOLOGIES INC | 97 | LUCENT TECHNOLOGIES INC |
| 9 | AKAMAI TECHNOLOGIES INC | 98 | MAXIM INTEGRATED PRODUCTS |
| 10 | ALTERA CORP | 99 | MAXTOR CORP |
| 11 | AMETEK INC | 100 | MCI INC |
| 12 | AMKOR TECHNOLOGY INC | 101 | MENTOR GRAPHICS CORP |
| 13 | AMPHENOL CORP | 102 | METHODE ELECTRONICS -CL A |
| 14 | ANALOG DEVICES | 103 | METROLOGIC INSTRUMENTS INC |
| 15 | APPLE INC | 104 | MICREL INC |
| 16 | APPLIED MICRO CIRCUITS CORP | 105 | MICROCHIP TECHNOLOGY INC |
| 17 | ARRIS GROUP INC | 106 | MICRON TECHNOLOGY INC |
| 18 | AT&T CORP | 107 | MICROSEMI CORP |
| 19 | AT&T INC | 108 | MICROSOFT CORP |
| 20 | ATHEROS COMMUNICATIONS INC | 109 | MICROVISION INC |
| 21 | ATI TECHNOLOGIES INC | 110 | MINDSPEED TECHNOLOGIES INC |
| 22 | ATMEL CORP | 111 | MITEL NETWORKS CORP |
| 23 | AUTODESK INC | 112 | MKS INSTRUMENTS INC |
| 24 | AVANEX CORP | 113 | MOLEX INC |
| 25 | AVAYA INC | 114 | MONOLITHIC POWER SYSTEMS INC |
| 26 | BEA SYSTEMS INC | 115 | MOTOROLA INC |
| 27 | BECKMAN COULTER INC | 116 | NATIONAL INSTRUMENTS CORP |
| 28 | BELL & HOWELL OPERATING CO | 117 | NATIONAL SEMICONDUCTOR CORP |
| 29 | BELLSOUTH CORP | 118 | NCR CORP |
| 30 | BIO-RAD LABORATORIES INC | 119 | NETLOGIC MICROSYSTEMS INC |
| 31 | BMC SOFTWARE INC | 120 | NETWORK APPLIANCE INC |
| 32 | BROADCOM CORP -CL A | 121 | NETWORKS ASSOCIATES |
| 33 | BROCADE COMMUNICATIONS SYS | 122 | NOVELL INC |
| 34 | CA INC | 123 | NUANCE COMMUNICATIONS INC |
| 35 | CADENCE DESIGN SYSTEMS INC | 124 | NVIDIA CORP |
| 36 | CASCADE MICROTECH INC | 125 | OMNIVISION TECHNOLOGIES INC |
| 37 | CERTICOM CORP | 126 | ORACLE CORP |
| 38 | CIENA CORP | 127 | PITNEY BOWES INC |
| 39 | CIRRUS LOGIC INC | 128 | PLANTRONICS INC |
| 40 | CISCO SYSTEMS INC | 129 | PMC-SIERRA INC |
| 41 | CITRIX SYSTEMS INC | 130 | POLYCOM INC |
| 42 | COGNEX CORP | 131 | POWER INTEGRATIONS INC |
| 43 | COHERENT INC | 132 | QLOGIC CORP |
| 44 | COMMVAULT SYSTEMS INC | 133 | QUALCOMM INC |
| 45 | CONEXANT SYSTEMS INC | 134 | QUANTUM CORP |
| 46 | CORNING INC | 135 | QWEST COMMUNICATION INTL INC |
| 47 | CREDENCE SYSTEMS CORP | 136 | READ-RITE CORP |
| 48 | CREE INC | 137 | RED HAT INC |
| 49 | CYPRESS SEMICONDUCTOR CORP | 138 | RESEARCH IN MOTION LTD |
| 50 | DALLAS SEMICONDUCTOR CORP | 139 | ROGERS CORP |
| 51 | DELL INC | 140 | SANDISK CORP |
| 52 | DIEBOLD INC | 141 | SCIENTIFIC-ATLANTA INC |
| 53 | DIGIMARC CORP | 142 | SEAGATE TECHNOLOGY |
| 54 | DIRECTV GROUP INC | 143 | SENSORMATIC ELECTRONICS |

| | | | |
|---|---|---|---|
| 55 | EBAY INC | 144 | SIGMATEL INC |
| 56 | ECHOSTAR CORP | 145 | SILICON GRAPHICS INC |
| 57 | ELECTRONIC DATA SYSTEMS CORP | 146 | SILICON IMAGE INC |
| 58 | ELECTRONICS FOR IMAGING INC | 147 | SILICON LABORATORIES INC |
| 59 | EMC CORP/MA | 148 | SILICON STORAGE TECHNOLOGY |
| 60 | EMULEX CORP | 149 | SILICONIX INC |
| 61 | EXTREME NETWORKS INC | 150 | SKYWORKS SOLUTIONS INC |
| 62 | F5 NETWORKS INC | 151 | SPANSION INC |
| 63 | FAIRCHILD SEMICONDUCTOR INTL | 152 | STANDARD MICROSYSTEMS CORP |
| 64 | FEI CO | 153 | STORAGE TECHNOLOGY CP |
| 65 | FINISAR CORP | 154 | SUN MICROSYSTEMS INC |
| 66 | FIRST DATA CORP | 155 | SYBASE INC |
| 67 | FORMFACTOR INC | 156 | SYMANTEC CORP |
| 68 | FOUNDRY NETWORKS INC | 157 | SYMBOL TECHNOLOGIES |
| 69 | FREESCALE SEMICONDUCTOR INC | 158 | SYMYX TECHNOLOGIES INC |
| 70 | GATEWAY INC | 159 | SYNAPTICS INC |
| 71 | GENESYS TELECOMM LABS INC | 160 | SYNOPSYS INC |
| 72 | GOOGLE INC | 161 | TEKTRONIX INC |
| 73 | HARMAN INTL INDUSTRIES INC | 162 | TELECOMMUNICATION SYS INC |
| 74 | HARRIS CORP | 163 | TELLABS INC |
| 75 | HEWLETT-PACKARD CO | 164 | TERADYNE INC |
| 76 | HUTCHINSON TECHNOLOGY INC | 165 | TEXAS INSTRUMENTS INC |
| 77 | I2 TECHNOLOGIES INC | 166 | TRIQUINT SEMICONDUCTOR INC |
| 78 | IMMERSION CORP | 167 | UNISYS CORP |
| 79 | INFINERA CORP | 168 | UNIVERSAL DISPLAY CORP |
| 80 | INTEGRATED DEVICE TECH INC | 169 | UNIVERSAL ELECTRONICS INC |
| 81 | INTEL CORP | 170 | VARIAN INC |
| 82 | INTERMEC INC | 171 | VIASAT INC |
| 83 | INTERSIL CORP | 172 | WESTERN DIGITAL CORP |
| 84 | INTL BUSINESS MACHINES CORP | 173 | WORLDCOM INC-CONSOLIDATED |
| 85 | INTL RECTIFIER CORP | 174 | XEROX CORP |
| 86 | INTUIT INC | 175 | XILINX INC |
| 87 | IOMEGA CORP | 176 | YAHOO INC |
| 88 | IXYS CORP | 177 | ZILOG INC |
| | | 178 | ZORAN CORP |