UNIVERSITÀ DEGLI STUDI DELL'INSUBRIA

Scuola di Dottorato in Scienze economiche e statistiche

DOCTORAL PROGRAM IN METHODS AND MODELS FOR ECONOMIC DECISIONS

UNIVERSITÉ DE BORDEAUX

Ecole Doctorale Entreprise, économie, société

DOCTORAL PROGRAM IN ECONOMICS

Candidate:

Julien SEAUX

Migration and Innovation : An analysis based on patent data

Thesis supervisors: **Andrea Vezzulli** and **Francesco Lissoni**

Co-supervisor: **Stefano Breschi**

Date of Thesis defense: 16 December 2019

Committee members:

M. Brambilla Carlo Santo
Professor, Università degli studi dell'Insubria, **Chair of PhD defense**

Mme Plunket Anne
Professor, Université de Lorraine, **Reviewer**

Mme Lenzi Camilla
Associate professor, Politecnico Milano, **Reviewer**

Mme Roux Pascale
Associate professor, Université of Bordeaux, **Examiner**

Mme Elena Maggi
Associate professor, Università degli studi dell'Insubria, **Guest member**

M. Vezzulli Andrea
Associate professor, Università degli studi dell'Insubria, **Supervisor**

M. Lissoni Francesco
Professor, Université de Bordeaux, **Supervisor**

M. Breschi Stefano
Professor, Universià « Luigi » Bocconi, **Co-supervisor**

Title: Migration and innovation: An analysis based on patent data

Abstract: This thesis investigates the migration of inventors by studying their contribution to the innovation at both destination and in their home country, after controlling for individual characteristics such as gender, education, experience, company mobility and others. In addition, we decompose the flows of migrants by entry channel in the destination country, such as the education channel, the multinational channel or when the inventors change of company and investigate the selection of migrants and the productivity gap between natives and migrants. Also, in the analysis, we decompose the cohort of entry in the destination country to compare the productivity differences among migrants themselves. Finally, we study whether return migrants are more productive than their non-migrants' colleagues in origin countries, as a function of their experiences abroad.

Keywords: Migration, Inventor, Return Migration, Productivity, Self-selection

Titre : Migration et innovation : Une analyse basée sur des données de brevet

Résumé : Cette thèse a pour but d'analyser la migration des inventeurs en étudiant leur contribution à l'innovation de leur pays de destination ainsi que leur pays d'origine, en contrôlant par leurs caractéristiques individuelles telles que leur genre, éducation, expérience, mobilité interentreprise et d'autres. De plus, nous décomposons les flux de migrants par canaux d'entrer dans le pays de destination, tel que le canal de l'éducation, multinational ou si l'inventeur change d'entreprise, et analysons la sélection des migrants et le gap de productivité entre les natifs et les migrants. Aussi, dans cette analyse, nous décomposons les cohortes d'entrées dans le pays de destination en comparant la productivité entre les migrants eux-mêmes. Finalement, nous étudions si les migrants de retour sont plus productifs que leurs collègues non migrants dans leur pays d'origine en fonction de leur expérience acquise à l'étranger.

Mots clés: Migration, Inventeur, Migration de retour, Productivité, Autosélection

Titolo : Migrazione e innovazione : Un’analisi basata su dati brevettuali

Abstract : L'obiettivo principale di questa tesi è lo studio della migrazione degli inventori e il loro contributo all’innovazione nei Paesi di destinazione ed origine, controllando per una serie di caratteristiche individuali, come ad esempio genere, livello di educazione, esperienza e mobilità. Inoltre, differenziamo il flusso migratorio secondo lo specifico canale di entrata nel Paese di destinazione, come ad esempio educazione, riallocazione di sede all’interno di imprese multinazionali, o mobilità tra imprese, per analizzare la selezione e il gap di produttività tra migranti e non-migranti. Per comparare la produttività tra migranti, la nostra analisi distingue i differenti periodi di entrata nel Paese di destinazione. Infine, analizziamo se gli inventori che tornano nel loro Paese di origine siano più produttivi dei loro colleghi non-migranti in funzione delle esperienze acquisite nel Paese di destinazione.

Parole chiave : Migrazione, Inventori, Migrazione di ritorno, Produttività, Autoselezione

"Alone we can do so little. Together we can do so much."

Helen Keller

Acknowledgements

I would like to thank all people and institutions that helped to develop this work and made this experience successful.

I want to thank the Department of Economics of the University of Insubria to have accepted me in their PHD program, and the GREThA at the University of Bordeaux with which the co-tutelle took place. In that matter, I would like to thank the program Vinci initiated by the Università Italo-Francese to have provided me funding for the multiple travels between these institutions. Finally, I would like to thank the Università Luigi Bocconi and the CESPRI that guested me on multiple occasions during these three years of visiting.

I would like to thank the members of my thesis committee Camilla Lenzi, Anne Plunket, Pascale Roux, Carlo Brambilla and Elena Maggi for their support and for the time spent evaluating this thesis.

I am greatly in dept to my three supervisors; Andrea Vezzulli, Francesco Lissoni and Stefano Breschi. Their advises go beyond the reduce frame of a thesis, they became examples along the way.

Stefano, I thank you for this period in Bocconi, I learnt so much about programming thanks to you, your personality helped me to be comfortable since the first day and I hope that you enjoyed our collaboration.

Andrea, I thank you to have been always here for me, I hope I did not annoy you too much or/and too often. I am grateful for all the advises in econometrics, writing, presenting... I hope you enjoyed Bordeaux and Saint-Emilion, it was a very good moment there.

Francesco, it is very difficult to express something so simple, so I will tell it straight: you are like father to me. During these years, you became a true example of who I would like to be. You taught me so much and helped me to push my limits.

I address my sincere gratitude to Valerio Sterzi, my advisor and friend since my third year of bachelor's degree, for his trust, support and precious comments. I did my first internship with you, you always believed in me, and supervised me on multiple occasions during this thesis.

I would like to thank Pascale Roux, Valerio Sterzi, Olivier Baron and Tanguy Bernard for the teaching responsibilities they gave me.

I would like to emphasis a special thanks to Ernest Miguelez, Gianluca Orsatti, Massimiliano Coda Zabetta, Eduardo Ferrucci and Christian Chacua.

I would like to thank my colleagues Dan, Louis, Jeanne, Martin, Marine, Damien, Leo, Lucie, Lucile, Erwan, Laura, Sebastien, Diane, Suneha, Nico, Viola, Nicolas Bedu, Pierre, Tomi, Marianne and Jeremy.

To the other member of the GREThA, DIECO and CESPRI; professors and fellow doctoral students thank you for creating a friendly atmosphere that made working in these different departments a pleasant experience.

I give my most loving appreciation to my mother and sister for their endless support, and in particular my father, without whom I would have never continued so far in my studies and who gave me the spirit of never give up.

Finally, thank you Valentina, for all your love and support in my good and bad times. We have been through a lot during these last years and I can’t wait for the next adventures with you.

Table of contents

List of Figures

List of Tables

Introduction

Since the earliest times, humanity has been on the move. Some people move in search of economic opportunities, to join family, or to study; others move to escape conflicts, persecution, terrorism, or human rights violations. Increasingly, nowadays, others move in response to the adverse effects of climate change, natural disasters, or other environmental factors. Between the years 2000 and 2017, the number of international migrants has grown from 173 to 258 million, of individuals, and the share of international migrants in the world population has grown from 2.8% to 3.4%[1]. Hence, today, more people than ever before live in a country different from the one in which they were born.

Within migration flows, the one of high skilled migrants has particularly increased. The share of high skilled individuals, defined in official statistics as individuals with a completed tertiary education, on the total migrant population has grown from 27% to 50% between 1990 and 2010 (IOM, 2018). This may be due to the increasing education levels worldwide and to a growing demand for skilled labour, particularly from developed economies. Also, better wages and employment conditions, better information, recruitment, and lower transportation costs are encouraging skilled migrants to seek jobs in developed economies. Graduates in Science, Technology, Engineering, and Mathematics (STEM) significantly contribute to this flow, especially flows connecting China, India, and Eastern Europe to the United States and other English-speaking countries (Docquier & Rapoport, 2012; Freeman, 2013).

This issue raises several issues on the migrants' role in the innovation process in their countries of destination and origin. This thesis consists of three related chapters that investigate as many issues. Chapter 1 relates to the lack of data on STEM workers' migration and describes the methodology used to create the main data source for the following chapters, namely the "Linked Inventor" dataset. The Linked Inventor database matches information on inventors, retrieved from LinkedIn (a professional-oriented social media), with patent data collected from the United States Patent & Trademark Office (USPTO), the European Patent Office (EPO) and the World Intellectual Property Organization (WIPO). The result is a highly detailed set of information on inventors, thanks to which we are able to investigate the different characteristics of migrant and native inventors working in several countries.

In chapter 2 we investigate the difference in productivity between Indian and native inventors in the US Among migrants, we differentiate by two possible channels of entry at destination, namely education and work. Among work-entry migrants, we furtherly distinguish if they change of company or relocated within the same one, most typically, between different country locations in the same multinational enterprise. Also, we investigate the productivity difference of migrants entering the US during different periods of the H1-B visa policy, that we use as a proxy for the migrants' degree of selection. Finally, we

[1] United-Nation Migration Data Portal: https://migrationdataportal.org/?i=stock_abs_&t=2017&m=1 (last visited on August 2019)

study the company's mobility of migrants and natives at destination, finding a different effect of mobility during the patenting activity and before (after) the first (last) patent filed. We show that migrants are more company mobile than natives at destination.

In chapter 3, we focus on Europe and compare the productivity, in the country of origin, of return migrants and stayers. Among return migrants, we distinguish between inventors having worked abroad and those with just an educational experience. We only find a productivity premium for return migrants with a working experience abroad.

This dissertation confirms that some brain gain potential exists for both the migrants' host and home countries but finds that it varies conditionally to the migrants' or return migrants' experiences or entry channels. Thanks to our new data, we strengthen the previous results by controlling for crucial individual characteristics and can consider country policies such as the H1-B visa in the United-States. To the best of our knowledge, ours is the first attempt to cross migration and company mobility literature for both natives and migrants. Furthermore, ours is also the first attempt to investigate and to account for both positive and negative selection of European return migrants in STEM fields. Finally, we are able to control for a set of "early stage" characteristics of the inventors that go usually unobserved in other studies, such as: the country where the highest level of education was obtained, the mobility patterns and work experience before the first patent filed.

Chapter 1

The Linked Inventor database: Methodology and Contents

This first chapter discusses the methodology used to build an original database on a worldwide population of inventors, to which we will refer as the Linked Inventor database. Its contents result from matching 424.497 public LinkedIn profiles associated to employees of large Semiconductor and ICT companies active in the US with inventor data as found on patent documents filed by the same companies from 1950 to 2016. The goal is to enrich the information one can retrieve from patent data with information on the migrant vs. native status of the inventor, as well as information on education and labour market experience. To assign a country of origin, we exploit both the information from LinkedIn as well as further information such as the results of name analysis, language proximity or inventor's nationality as reported on a subset of USPTO patent applications filed according to the PCT procedure before 2011. The final database contains 98.853 inventors with a LinkedIn profile and 138 different countries of origin.

1.1 Introduction

The LinkedIn Inventor database contains detailed information gathered from LinkedIn and patent data. We focus on inventors who work in the ICT sector for 178 US public companies that filed at least 200 patents. The patent applications filed by our population are registered in the United States Patent and Trademark Office (USPTO) between 1960 and 2016. By exploiting the information collected from the LinkedIn profiles, and combining name analysis and language proximity, we are able to develop an algorithm, the Homeland algorithm, that assigns a country of origin, among 137 possibilities, to each inventor.

Differently from other data sources on inventors, the Linked Inventor database includes not only information on patenting activity, but also information on the inventors' education and work experience. This allows us to have accurate information on inventors' individual characteristics and mobility determinants. Also, the diversity in terms of inventors' country of origin, and the representativeness of inventors working in ICT in the US, makes it a unique tool to explore the main topic of this dissertation, namely the relationship between migration and innovation. Throughout all the steps of the database construction, that are data analysis, scientific computing, approximation and phonetic matching of strings, regular expression operations, we use Python as coding language. Figure 1.12 in the Appendix provides a short description of the various step.

Users of the Linked Inventor database need to achieve some understanding of the methodology used to identify inventors, gather the primary data, and on the subsequent matching procedures and data management performed. This is what this chapter aims to present, alongside with some summary information on the database contents. The rest of this chapter is organized as follows: section 2 presents the main databases that deal with the migration of highly skilled workers in general or inventors in particular, and an overview of the most relevant studies using them; section 3 presents a summary of the main steps of the Linked Inventor database construction; section 4 shows some summary statistics and stylized facts we can get from the data; section 5 concludes.

1.2 Literature review

The international mobility of skilled workers, associated with both brain gain and brain drain, has gained prominence in public policy decisions on economic growth driven by innovation. The importance of high skilled migration is a well-recognized phenomenon in the literature. Earliest contributions stress the adverse consequences of the loss of nationally trained human capital from developing countries working and living abroad (Bhagwati & Hamada, 1974). Advances in the better understanding of the impact of highly skilled workers migration is notably due to the new data becoming available over the last 20 years and, started by the pioneers that adopted a more positive view of the migration of highly-skilled workers in the 90s' (Borjas, et al., 1992; Topel & Ward, 1992; Borjas & Bratsberg, 1996).

One of the first attempts to build in a systematic way a dataset on emigration rates by educational attainment is Carrington & Detragiache (1998). They provide 1990 emigration rates for 61 sending countries to OECD destinations, they estimate skill levels by extrapolating the schooling levels of US immigrants. It was followed by Docquier & Marfouk's (2006) and Defoort's (2008) works on migration stocks, and the breaking down of each stock by the level of education. Later Beine et al. (2007) provided the entry age of immigrants and Docquier et al. (2009) a breakdown by gender.

Nevertheless, the existing datasets provide a skills breakdown according to three schooling levels, such as primary, secondary, and tertiary education, which only offers a rough differentiation of skills.

When we focus on the migration of inventors, using patent applications can overcome several limitations associated with migrant stock data. Inventors represent a specific type of highly skilled workers more homogenous than the entire group of educated tertiary workers.

Patents are one of the significant sources of information for the analysis of innovation and technological change (Griliches, 1998). Most of the empirical studies on inventors' mobility make use of patent data. One of the forerunners in this literature is the work of Almeida and Kogut (1999). They focus on inventors in the US semiconductor industry and observe that regions differ in the degree of localization of knowledge, interpreting this result as the effect of workers' mobility, i.e. inventors' mobility. The authors rely on patent data for three different aspects:

1) Inventors are considered as a proxy for skilled workers as well as researchers due to their direct involvement in knowledge development in the varied organizations they are employed.
2) The data on the applicants (Company, University, or the inventor himself) are used to track the inventor's mobility across organizations and regions. An inventor is considered mobile when he/she applies at least two patents held by two different applicants.
3) Finally, data on citations are used as a proxy for knowledge flows.

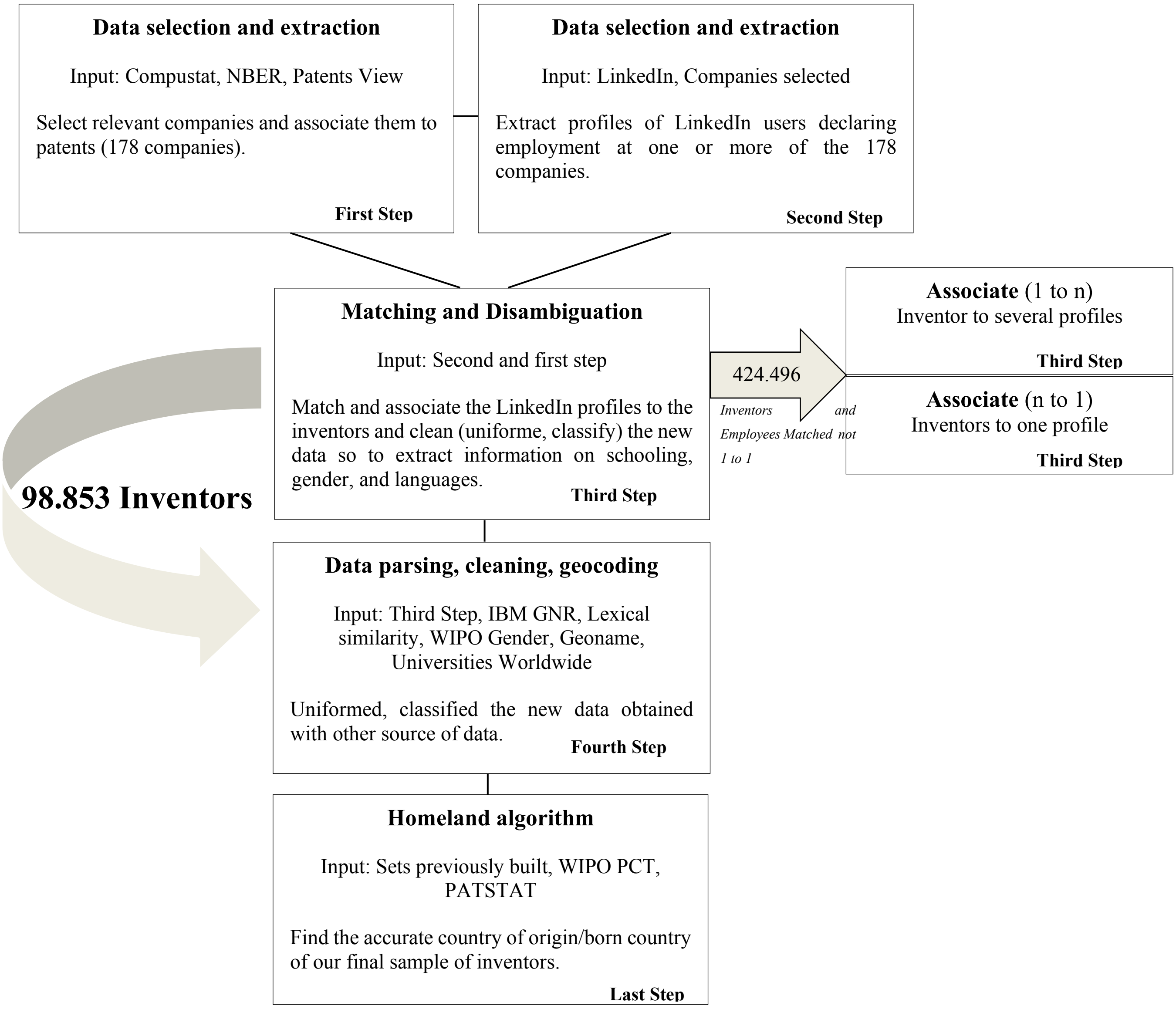
The three previous cited aspects have been widely applied to several topics within the broad field of economic studies on innovation (Trajtenberg, 2005; Hoisl, 2007; Breschi & Lissoni, 2009; Breschi, et al., 2010; Lenzi, 2010). The availability and quantity of patent databases, such as those maintained by the United States Patent and Trademark Office (USPTO) or the European Patent Office (EPO), make the use of patent data even more appealing. Moreover, patent datasets cover several countries, years, and type of organizations.

Nevertheless, patent data are limited when investigating migration. Even though they record the inventor's localization in space, information on the inventor's nationality and country of origin is most often missing. The first mean to detect individual's migration pattern experimented by researchers has consisted in identifying the likely origin of inventors' names and surnames (Kerr, 2008; Agrawal, et al., 2011; Foley & Kerr, 2013). However, guessing the country of origin using names may not always capture recent migratory backgrounds or may overlook immigrant inventors with names sharing the same cultural roots of the host country. In that matter, the contribution of Miguelez and Fink (2013) was to build a dataset on the international mobility of inventors using the information contained in the patents' applications filed under the Patent Cooperation Treaty (PCT), such as the residence and the nationality of the inventors. Doing so they were able to enlarge the number of countries and period of study and to rely on migratory background information revealed by inventors, rather than indirectly inferring a possible migration history through the cultural origin of names. Nevertheless, other scholars used and improved methods in assessing the origin's country of inventors based on the individual's name and surname, using, for instance, IBM Global Name Recognition (Breschi, et al., 2017) as well as improving disambiguation technics (Pezzoni, et al., 2014).

1.3 Database construction

In this section we offer a detailed description of the methodology used to build the LinkedIn database. Figure 1.1 provides an overview and a brief description of the various steps.

Figure 1.1: Overview of the main steps



1.3.1 Data selection and extraction

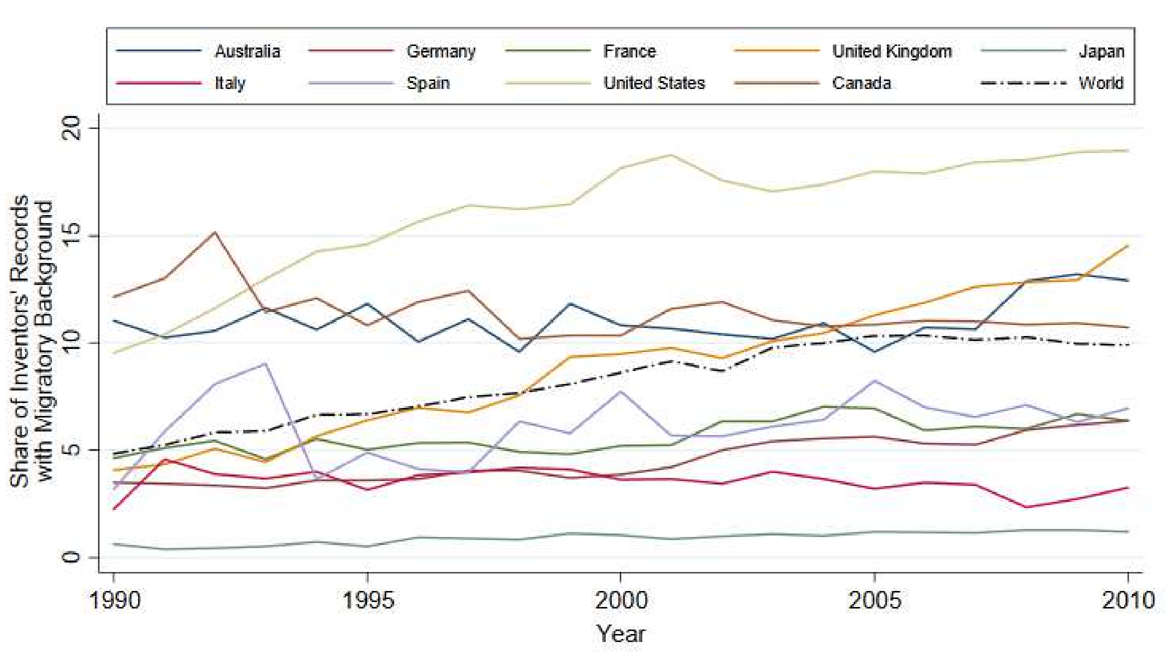
The first step in our methodology consists in selecting a representative population of inventors. Due to computational constraints we cannot gather all inventor profiles on LinkedIn and then perform a disambiguation procedure. Hence, we focus on one destination country (the Untied States) and one sector (ICT).

The choice to focus on the United States is driven by the fact that the country attracts many high-skilled migrants, and inventors in particular. As Figure 1.2 shows, in 2010 the US are the country that hosted 20% of the total migrant inventors, being the country with the highest share. Also, in the United States, migrants hold a disproportionate share of jobs in Science, Technology, Engineering and Math (STEM) occupations. As shown by the Figure 1.3, in 2013 the foreign-born workers accounted for 43 percent of STEM workers.

The second choice, conditional to the previous one, is to focus on the Information and Communications Technology (ICT) sector, that is known for its high patenting activity. We only consider the company that filed at least 200 patents, ending up with 178 ones[2]. To select the ICTs' US companies, we use the database of the National Bureau of Economic Research (NBER) which use USPTO patents from the North American Compustat data at Wharton Research Data Services. Then, we use Patentsview that is a patent database provided by the United States Patent Trade Office (USPTO), that allows us to obtain the patent's assignee name alongside the inventors' names. Matching Patentsview to the NBER dataset using the assignee name and company code, we obtain each company's patents. At the end of this crucial step, we end up with 424.496 LinkedIn profiles of employees working for one of the 178 selected companies and 262.849 inventors that have patented for one of the 178 selected companies.

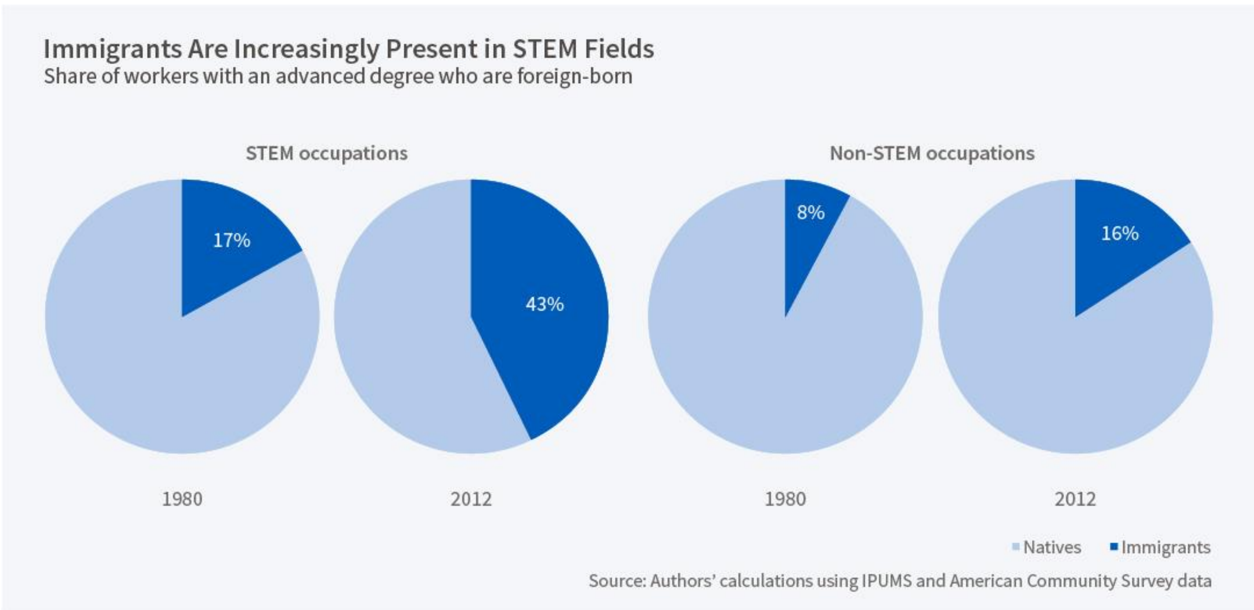
[2] Find in Appendix 1.6, Table 1.8, the list of the 178 companies.

Figure 1.2: Share of migrant inventors over total inventors within each country



Source: Miguelez and Fink (2013)

Figure 1.3: Share of migrant in STEM in the US



Source: (NBER, 2016)

1.3.2 Matching and Disambiguation

Using the selected LinkedIn profiles, we perform a matching followed by a disambiguation procedure. We match two lists, one of inventors' names reported on the patents, and one of names reported in the LinkedIn profiles by company such as:

- List A: names of inventors who made patents for company i
- List B: names of people reporting company i as one employer

We perform the matching in various ways, among others exact name match, inverted names, variants of first name, string similarity. The same inventor can be matched to multiple LinkedIn profiles, for example: Inventor 6278442-5 JULIEN SEAUX is matched to:

- Profile 1: CA5a0a87f3447f9c69 JULIEN R SEAUX
- Profile 2: CA4d24cd7724d7a086 JULIEN SEAUX

This requires a second stage where:

1) Profiles are disambiguated: this step is necessary as some people open more than one LinkedIn account. Typically one of the two is not updated, but they both correspond to the same person. These profiles need to be consolidated.
2) Assign a unique LinkedIn profile to multiple inventors: This is necessary, because the same profile can refer to apparently different inventors who are the same person. The latter problem is most likely due to lack of recall (presence of many false negatives) in the disambiguation of inventors.

First, we compute the number of profiles per person, and we separate the inventors with only one profile matched. For the inventors with more than one profile, we disambiguate and consolidate them. For this purpose, we proceed as follows:

1) For each inventor, we create all possible pairs of profiles that were matched to him/her.
2) We select the pairs where at least one employer (exactly named) is in common. The idea is that if the same individual has more profiles, at least one employer should match.
3) We merge information on the start and end date of employment for a given employer (if an individual has created a new profile, for the employer that the two profiles have in common, the starting and/or end date should match), the job title, the skills reported in the profiles, and the education institution (again the idea is that if an individual has created a new profile, the two profiles should have some of those dimensions, i.e. skills, education, job title, in common)
4) We consolidate two LinkedIn profiles (i.e. we assume that the two profiles belong to the same person who probably opened two accounts) if:

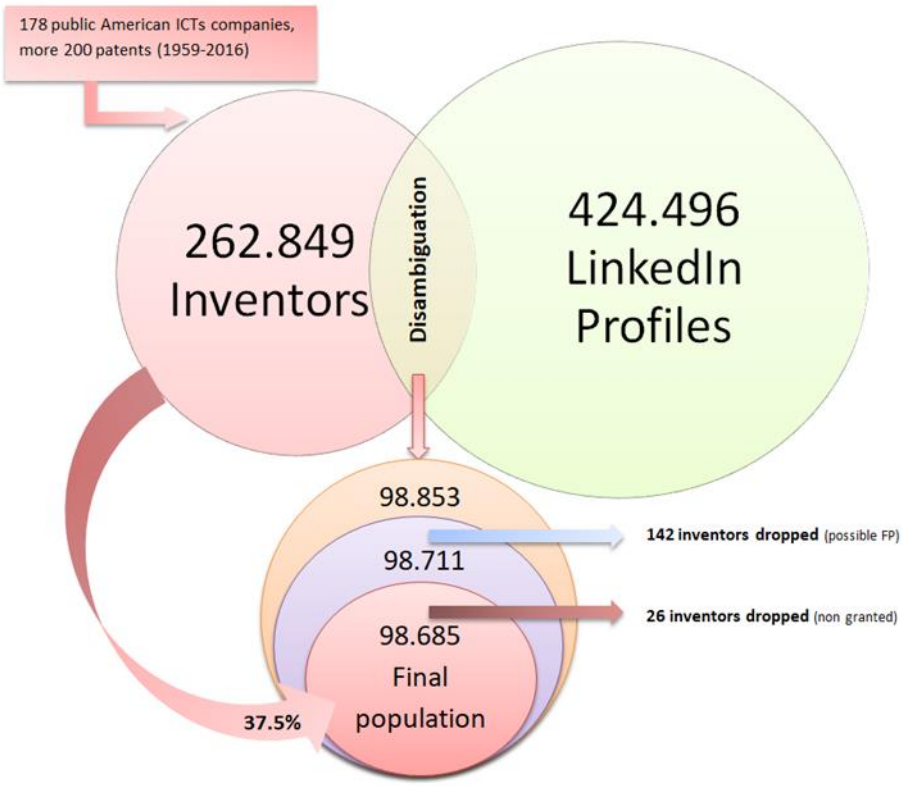
a. The starting date and the end date of employment for (at least one) of the common employers match.
b. The starting date of (at least one) of the common employers matches, the end date does not match, but the two profiles have at least one common job title (string similarity across job titles computed with Jaro Winkler and two job titles matched if >0.85). When a person opens a new account, the old one still exists but it is not updated by the person. However, LinkedIn updates the old profile automatically, therefore reporting that the end date of the last employer is "Present", whereas in the "new" profile the end date for that employer is not necessarily the same. Hence, the two end dates might not coincide.
c. (At least one) name of the employers exactly match in the two profiles, and they have (at least one) exactly named job title in common. Here, we do not check for the starting and end dates, but the fact that employers and job titles exactly match should be enough guarantee that the two profiles belong to the same person.
d. At least two names of the employers exactly match in the two profiles.
e. The two profiles share at least two common schools.
f. The two profiles share at least two common groups.
g. The two profiles share at least two common skills.

Notice that the criteria above are not mutually exclusive. In other words, two profiles are consolidated and assumed to belong to the same person if either one of the above criteria hold.

After consolidating the profiles, we take the inventors that were associated to multiple profiles (see above) and re-compute the total number of profiles after the consolidation. If an inventor has only two profiles and the two profiles have been consolidated, we take these two profiles as "a match" for the inventor. It is *as if* this inventor has been matched to only one profile, i.e. the one resulting from the consolidation.

For all other inventors, we assume that the association to multiple profiles is not due to problems of duplication of profiles, but it is a genuine problem of associating the true profile to the inventor among different potential candidates (e.g. near homonyms, etc.). This disambiguation is based on an algorithm that takes into account the names, companies, locations and patent's technologies of the inventors, and attributes a score based on the comparison of these variables. The threshold defined aims at decreasing the number of false positives rather than a higher rate of matching. At the end of the disambiguation process, we end up with 98.853 inventors linked with their LinkedIn profile out of 262.849 inventors and 424.496 LinkedIn profiles, hence we are able to correctly identify the 37.5% of the inventor's population. Figure 1.4 provides a summarization of the various steps.

Figure 1.4: LinkedIn-Inventor matching procedure



1.3.3 Data parsing, mining and geocoding

In this section, we report the most important steps of the Homeland algorithm that use the raw data extracted from LinkedIn to attribute a country of origin (or birth) to the matched inventors.

First, based on their names, we assign a gender to each inventor. We do so by using two Python packages, *gender guesser* and *chichsexer*, that associate a probability to be a male or a female to a list of several names. The two different packages are complementary to each other since the first one, *gender guesser*, is not accurate with Asian names. We complete this procedure by performing a robustness check based on WIPO gender[3], that is a database from the World Intellectual Property Organization (WIPO) that has information on inventors' gender recorded in PCT patents (Martínez, et al., 2016). To avoid confusion, we also exploit the information on the inventors' presumed country of origin, as several first names can be associated to different genders depending on the country of origin[4].

Using the WIPO database we find a match for around 93.000 inventors, obtaining the same gender for 96.94% inventors correctly matched. The match is based on the first name and country of origin. For the remaining 5.000 inventors for who we do not find any possible match with WIPO database, we assign the gender based on our algorithm.

[3] For more information about the database see Appendix 1.6 sub section 1.6.1.g

[4] For instance Andrea, that in Italy is a masculine name while feminine in Spain.

We then treat the data on education. LinkedIn education data consist of a list of institutions and degrees obtained by the profile holder. However, the institution's country is often not indicated. Hence, to get the country location of the universities, we geocode this information extracting the relevant information from Google Maps. To correct potential errors, we cross data from Google with the dataset "Worldwide universities"[5], that is publicly available. The latter contains geographical information on 9362 universities in the world.

We then classify the degrees delivered using the following categories: High School Diploma, Bachelor, Master, PhD, MBA, Juris Doctorate and Post-Doc. Table 1.1 reports typical final output for a sample inventor:

Table 1.1: Education treated output

Inventor	University	Degree	Country	Start	End
58745	Simon Fraser	BSC	Canada	1967	1971
58745	Waterloo	MSC	Canada	1971	1973

Besides providing valuable information *per se*, data on education help us to assign a country of origin to each inventor. To do this, we combine this information with a linguistic analysis of the inventors' names and surnames, based on IBM GNR as in Breschi and al. (2017). The IBM Global Name Recognition (GNR) contains tools to manage, search, analyze and compare name data sets by leveraging culture specific name data and linguistic rules that are associated with the name's culture.

Before we assign the inventors' country of origin, we develop a solution to compensate the over representativeness of some populations that historically migrated into the United States, making the previous algorithms less likely to predict the exact inventors' country of origin, due to the existing procedures (Kerr, 2008), as noticed by Breschi and al. (2017). Another limitation of IBM GNR is that it does not predict the United States as possible origin's country. Table 1.2 we show a concrete example of the limitation of IBM GNR crossed with education information.

Table 1.2: IBM GNR prediction

First name	Last name	IBM GNR (with the highest prediction)	Education Country	Education level (latest one)
ANNA	POVZNER	Russia (81%)	Ukraine	Bachelor
JIANHUI	GU	China (92%)	Singapore	High School

In particular, we use language similarity index developed by Tyshchenko (1999) and exploit it in two ways. First, we solve cases in which the linguistic analysis of inventors' names and LinkedIn information on education suggest different countries of origin. For example, many inventors whose name is Russian, according to linguistic analysis, declare to have received their education in Ukraine. Knowing that Russian and Ukrainian are close languages, we decide that the inventor's name was in fact Ukrainian and that the country of origin is Ukraine, therefore giving more weight to the information on the education).

[5] For more information about the database please see Appendix 1.6, sub section 1.6.1.h

Also, based on the Language proximity analysis and Geoname[6], a database with information on the languages spoken in each country, we build the following associations, described in Table 1.3, linking over-represented and under-represented countries based on the languages spoken.

Table 1.3: Main country's extension, example

Main country	Languages associated	Countries associated
China	Cantonese, Mandarin, Vietnamese, Thai	Thailand, Vietnam, China, Hong Kong, Macao, Singapore, Taiwan
Russia	Russian, Ukrainian, Polish	Belorussia, Kirghizistan, Kazakhstan, Moldavia, Russia, Ukraine, Poland
Germany	Germanic, Dutch	Austria, Belgic, Switzerland, Germany, Italy, Liechtenstein, Luxembourg, Curaçao, Netherland, Suriname, Saint-Marin (Netherland part)
India	Hindi, Malayalam, Gujarati, Punjabi, Urdu, Tamil, Nepali, Bangla	India, Pakistan, Sri Lanka, Malaisie, Singapore, Nepal
Japan	Japanese	Japan

Finally, LinkedIn data report the languages spoken by the inventors. Even if only 20% of the gathered LinkedIn profile report at least one language spoken, is important to consider this information, crossed with the Geoname database, when predicting the country of origin. For simplification, we only keep the information on the native language or languages in which the profile holder declares bilingual proficiency.

[6] For more information about the database see Appendix 1.6.1.i

1.3.4 Homeland algorithm

From the previous steps, we build several sets of countries associated to each inventor, following the recorded information on the LinkedIn profile or the name and surname. As earlier mentioned, IBM GNR doesn't allow to identify US as a possible country of origin. For this purpose, we cross IMB GNR with WIPO Gender[7].

Below, two examples of the Homeland algorithm input: Table 1.4 reports an example of a set of countries associated to a given name and surname, while Table 1.5 reports the same example with a possible set of countries of origin, extended using the LinkedIn information on education location, and the Language Proximity.

Table 1.4: Countries associated with Name and Surname

	IBM GNR			
Inventor	Union of all possibilities between the inventor name and surname	Intersection of all possibilities between the inventor name and surname	Selection using the highest likely country of origin	WIPO Gender
ANNA POVZNER	(16 possibilities)	Russia, Israel	Russia	(46 possibilities)

Table 1.5: Countries associated to LinkedIn information and Language Proximity

Inventor	Language Spoken (native one only)	Education localisation	Language Proximity extension
ANNA POVZNER	nd	Ukraine, United States	Bielorussia, Kyrgyzstan, Kazakhstan, Moldavia, Russia, Ukraine, Poland

The Homeland algorithm is based on intersection of countries' sets. The main steps are the following:

1) We take the intersection between the IBM GNR and WIPO sets of possible countries
    a. If WIPO set contains US, we take the result of 1) + US.
    b. If the result of 1) is an empty set, we take the IBM GNR 'Union' set instead of the IBM GNR 'intersection' and recompute.
        i. If b) is still empty, we take the Union between IBM GNR 'Union' and WIPO Gender sets and we go in step 2).
        ii. If b) found a unique solution, we perform a robustness check based on the 'Education localisation' and 'Language proximity extension' sets.
        iii. If b) found several solutions, we go in step 2).
    c. Otherwise, continue in step 2)

[7] For more information about the database see Appendix 1.6.1.e

2) We take the result of 1) and take the intersection with the 'Language Spoken', 'Education localisation' and 'Language proximity extension' sets, as well as their three pairwise intersection.
    a. If 2) gives an empty set or more than one solution, we focus on the inventor's education and language proximity sets, note that the following decision rules are sorted by robustness:
        i. If the education localisation corresponds to the inventor's High School degree, we take this country as country of origin.
        ii. If 1) gives only one possibility, we use the language proximity analysis. If this new set match with the Bachelor localisation, we take this as country of origin.
        iii. If only IBM GNR 'intersection' set gives only one possibility, we do like 2.a.ii)
        iv. If the list of countries associated to 'Language Spoken' match with the Bachelor localisation, we take it as country of origin.
        v. We take the bachelor's localisation as country of origin.
        vi. If at least one localisation about education match with the list of countries of WIPO Gender, we take it as country of origin.
        vii. At last, the shutdown condition, we take the LinkedIn localisation corresponding where the LinkedIn profile has been created as country of origin.
    b. If 2) gives one solution, we take it as country of origin.

The education set is mostly composed by US degrees, as 59% of inventors completed their education, and 65% at least a part of it, in the United-States.
The Homeland algorithm predicts 137 different countries of origin. We find that 63% of inventors are from United-States, this result is in line with the selected population and the existing literature. In fact, we focus on Americans' ICTs companies where 30% of the working age population in Science, Technology, Engineering and Mathematics (STEM) were international students in United-States (Ruggles, et al., 2010). In addition, Hanson & Liu (2017) found that the share of US workers who are from India rises from near zero in 1960 to 9.3%, accounting for 33% of all foreign-born workers in 2010-2012. One of the challenges that this algorithm solves is to differentiate individuals with Anglo-Saxons names in Canadians, Australians, Americans, Irish, and British.

The table below, Table 1.6, summarizes the top 10 countries of origin, for the extended list see Appendix, Table 1.10.

Table 1.6: Top 10 countries of origin

Country of origin	Frequency / Percentage
United States	58671 / 61%
India	8272 / 9%
China	3539 / 4.9%
Great Britain	3256 / 4.7%
Canada	3193 / 4.6%
Israel	2009 / 2.1%
Germany	1809 / 1.9%
France	1364 / 1.8%
Japan	990 / 1.4%
Taiwan	678 / 1%

We check the robustness of our algorithm by comparing the country of origin with the nationality given by WIPO PCT application's data[8]. We are aware of the issues associated with approximating the country of origin with nationality, as individuals may be naturalized in the host country, and this is even more likely at the time of PCTs applications that usually occur quite late in the inventors' career. Also, not all the countries are members of WIPO, meaning that it will not be possible to find a correspondence for all the countries, and this may underestimate the prediction accuracy. This is a particular issue for countries with limited recognition, such as Taiwan. Yet, at the best of our knowledge, this is the best solution to carry out a robustness check on the Homeland algorithm. In the end, we obtain a 81% precision score.

1.4 Results and summary statistics

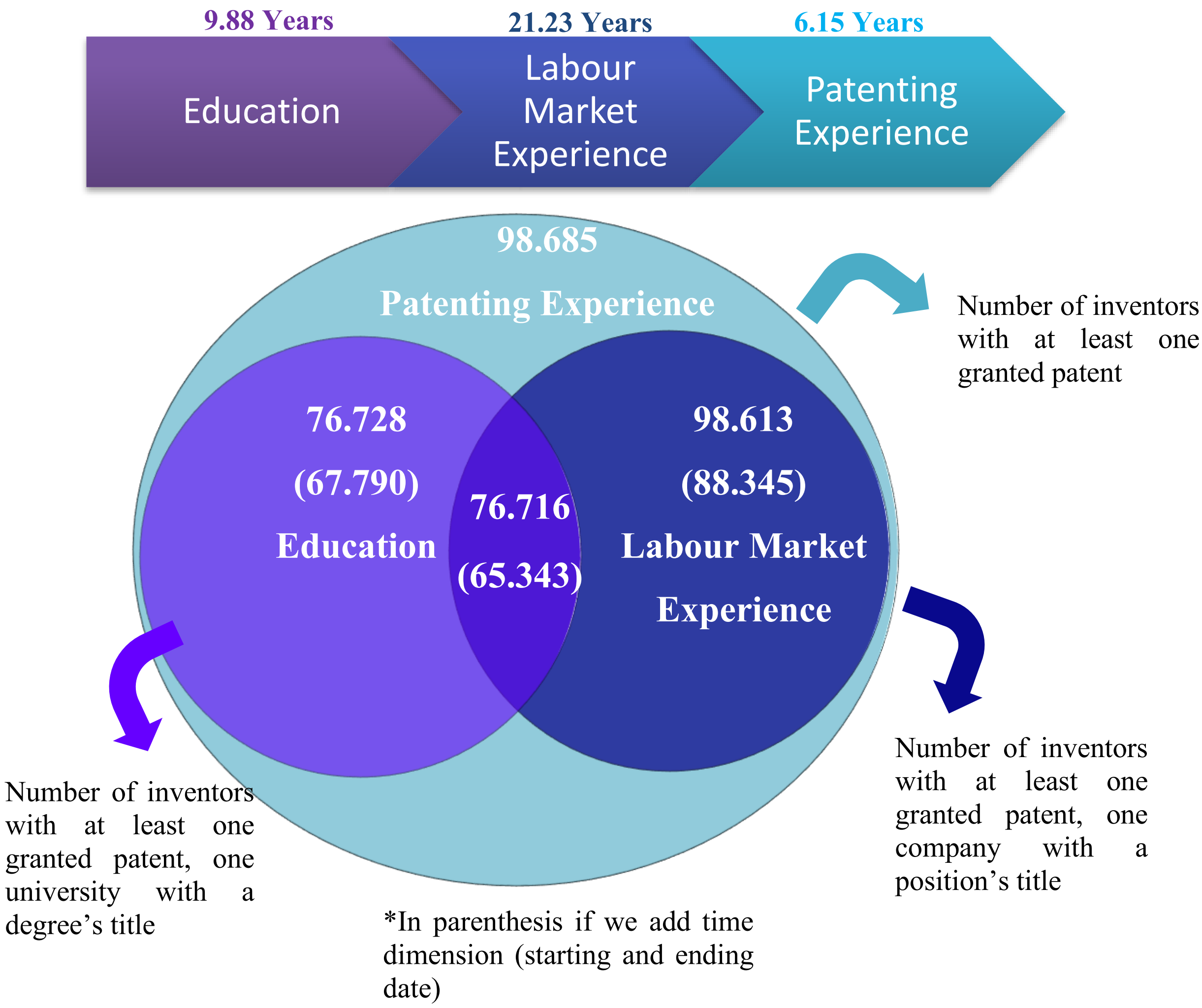
1.4.1 General database and population description

In this sub section we present some basic statistics of our sample of inventors. Figure 1.5 summarizes the information we gathered on inventors' education, participation to labour market and patenting experience. On average, we have information on inventors over 10 years of education[9] and 21 years of labour market activity, including 6 years of patenting activity. This is particularly important, as our data enable us to identify migration or mobility status before the first patent filed, overcoming one of the main limitations of patent data.

[8] We thank Carsten Fink and Ernest Miguelez for providing us the dataset.

[9] Can include high school in some cases as well as some period where the inventor stopped his education and continue it later, hence, explaining the important span.

Figure 1.5: Database general overview



However, not all LinkedIn profiles are complete. In fact, on the 98.685 profiles, only 76.716 have at least one information about education, the company and the position. If we furtherly restrict the sample to profiles that provide information on the starting or ending time of their education of work experience, we obtain 65.343 inventors.

Basic inventor's characteristics are summarized Table 1.7. The database is composed by 76% of stayers and 24% of inventors with at least one year of experience abroad. Within the migrant population we find 12% of returnees. The data accuracy allows us identify the entry channel in the destination country, we find that 46% of migrants entered the destination country during education, and 54% during labour experiences. Within the latter, only 11% entered in the host country after the first patent, underlying the limitation of studying inventors' migration only with patent data, that underestimate the real flow of migrants.

Table 1.7: Main inventor's characteristics

Main characteristics	Most frequent category	%
Migration status	Native	76%
	Migrant	24%
	Return migrant	12%
Education level	Bachelor	32%
	Master	36%
	Ph.D./MBA	31%
Migration motives	Education	46%
	Labour	54%
	Patent	11%
Gender	Male	88%
Degree	Computer science	6%
Company	IBM	7.6%
Position	Software engineer	3.5%
Last school	Stanford	2.2%
Last school (no US)	Waterloo	2.6%

Besides general statistics about migration status, entry channel, and level of education, we also have specific information on the inventor's degree and the current position in the company. As we focus on ICTs companies in the United States, we have an important representation of IBM degrees and positions related to computer science and software. When we consider the whole database, we find that the most represented university is Stanford. When we do not consider the US, the most represented university is the University of Waterloo in Canada. A possible explanation for this is that since 2002 the founder of Blackberry, Mike Lazaridis, has financed the computer science department in Waterloo and developed an important network of scientists and engineers between the Silicon Valley and the Institute of Quantum computing, also located in Waterloo. Also, the university of Waterloo has been ranked as the Canadian most innovative university for the past 25 years and is ranked 22th[10] best University in Computer science despite the overall University being ranked 170th.

We then investigate the collaborations and the production of patents (Figure 1.6 summarizes the main findings). Until 2016, the inventors in our sample filed 449.107 patents, 374.356 of these patents where co-invented, and 159.110 where co-invented with other inventors in our sample. One third of inventors filed only one patent.

[10] We use the QS World University Rankings : https://www.topuniversities.com/university-rankings

Figure 1.6 Patent and collaboration general overview

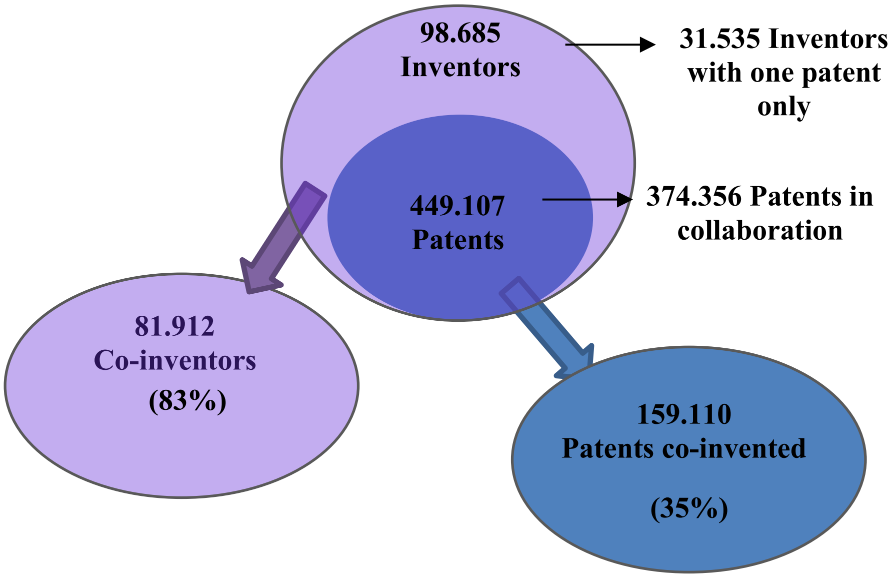
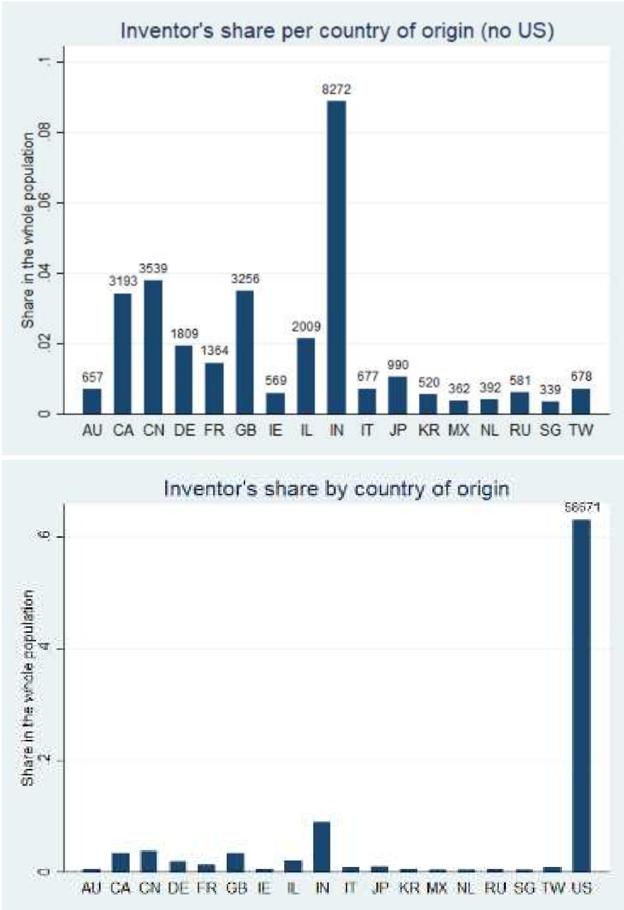


Figure 1.7 shows the sample decomposition by country of origin. Due to the selection of American companies, inventors from the US are over-represented in our data. Inventors from India account for almost 28% of all the other ethnicities, again, due to the specific field, the ICTs, this finding comes without surprise.

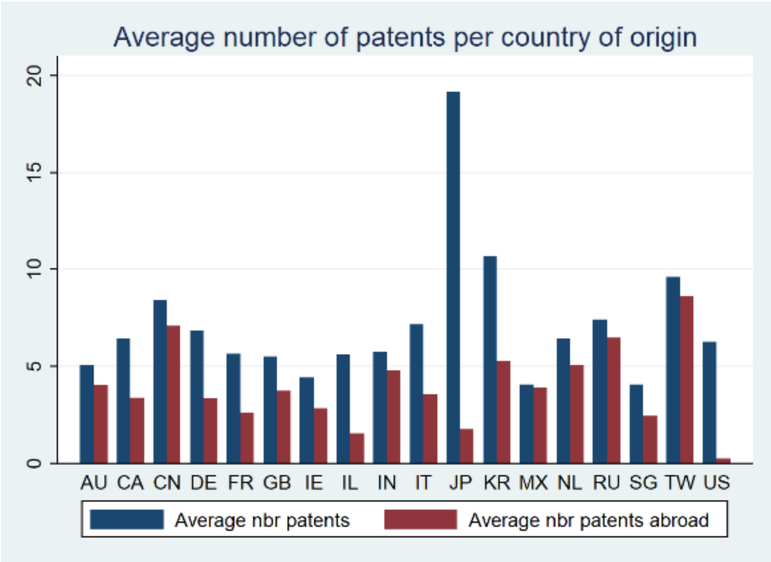
Figure 1.7 Sample decomposition by country of origin



(Source: Author's calculation from the Linked Inventor database)

Figure 1.8 gives general hints on the patent productivity across country of origin. We observe that Japanese' inventors filed twice more patents than the others. This result can be explained by differences across patent offices. In fact, the Japanese patent office (JPO) is known as granting patents with smaller claims, hence, the same innovative product required more patents than the one done in Europe or the United States. Despite this disparity in the number of patents granted by country of origin we observe heterogeneity in the location. In fact, inventors from the US or Japan patent more in their home country rather than abroad for the rest of the sample.

Figure 1.8 Number of patents granted by inventors decomposed by country of origin

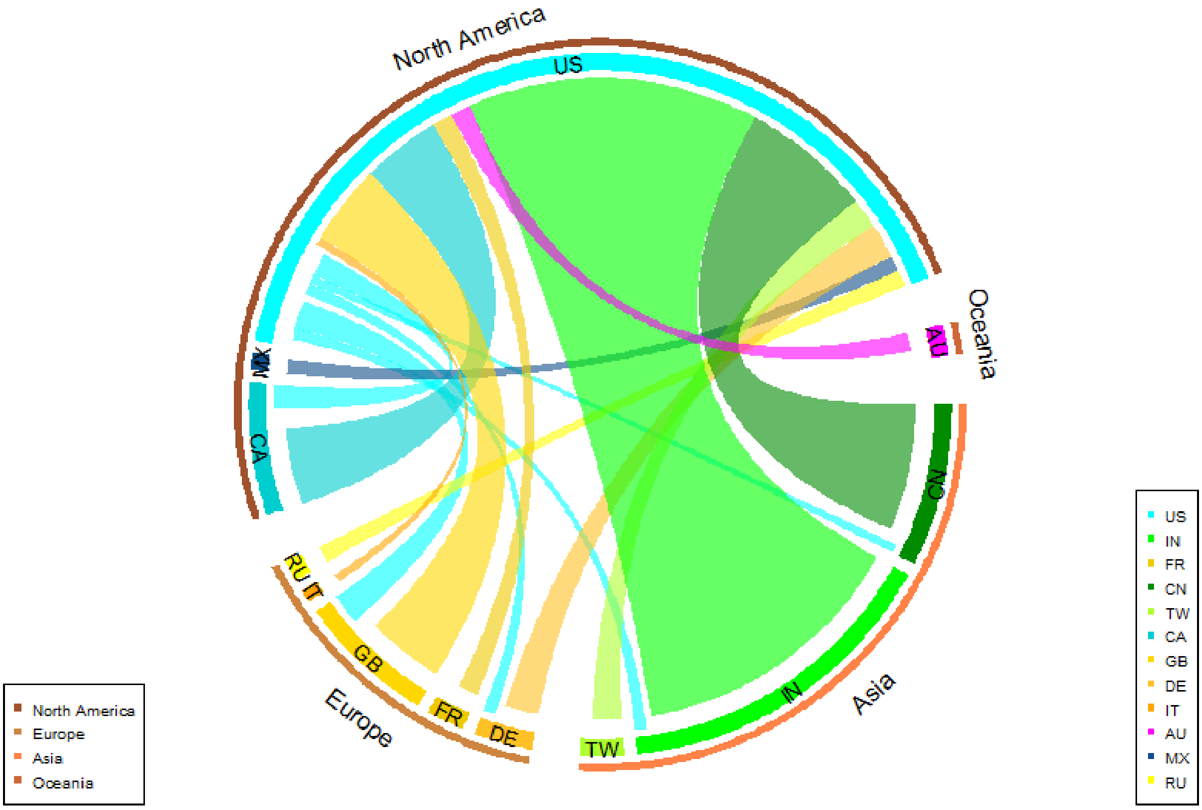


(Source: Author's calculation from the Linked Inventor database)

1.4.2 Migration Outflow and Inflow

In what follows we present some statistics on the main migration flows over the whole period covered by the Linked Inventor database (1950-2016). Figure 1.9 shows the most important inflows and outflows of inventors that are directed toward the US and coming from mostly India and China.

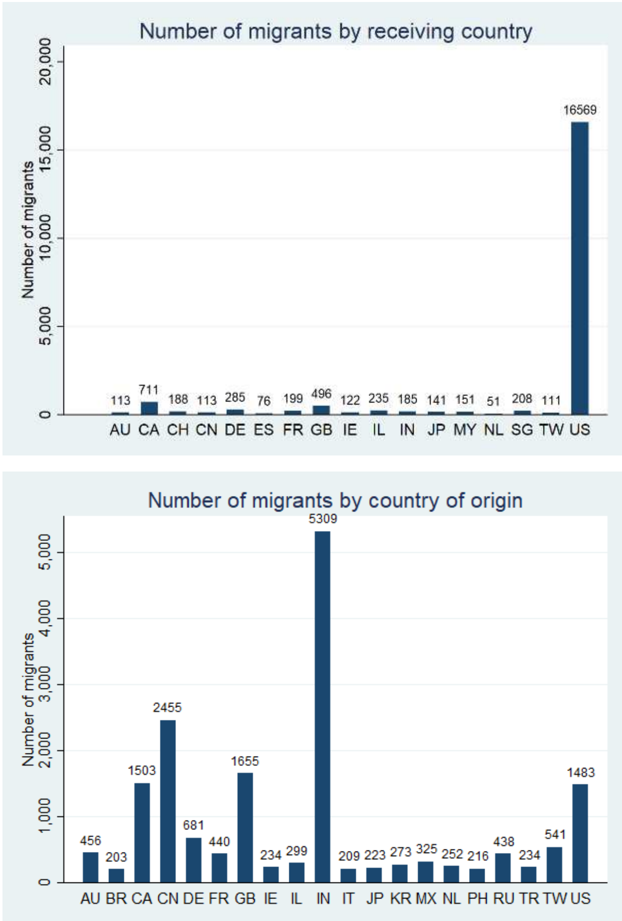
Figure 1.9: Most important migrations' outflows of inventors



(Source: Author's calculation from the Linked Inventor database)

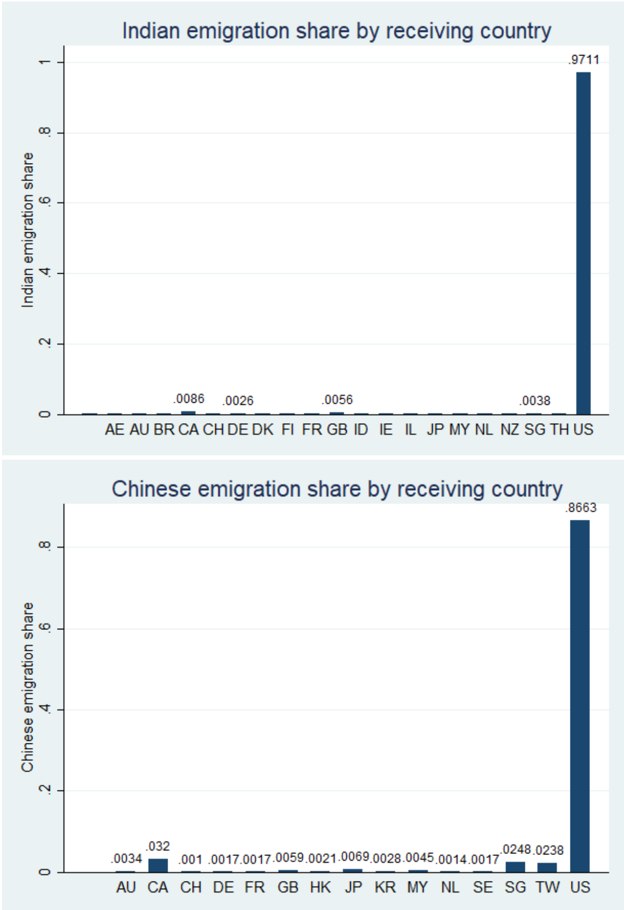
We can also notice some important flows between the US and the Canada, possibly due to regional proximity. As for Europe, the most important countries of origin are Great Britain, Germany, France, Italy and Russia. Figure 1.10 below shows the decomposition by receiving and sending countries and Figure 1.11 focus on the main sending countries, India and China.

Figure 1.10: Most important migrations' inflows and outflows of inventors



(Source: Author's calculation from the Linked Inventor database)

Figure 1.11: Most important migrations' outflows of inventors



(Source: Author's calculation from the Linked Inventor database)

1.5 Conclusion

The Linked Inventor database contains rich information on inventors in the ITC sector, with patenting experience in at least one US company, and with a LinkedIn profile. Besides biographical information (such as gender,country of origin, and country of residence at different points in time), and information on patenting activity (such as the number of granted patents, citations or claims) the database provides information on inventors' education (such as institutions and degrees) and career (such as the companies the inventor worked for and the positions held).

The database is particularly helpful to investigate the main topic of the present dissertation, that is the relationship between migration and innovation. First, it allows to consider a specific population, that is inventors working in the ICT sector in the US, making the comparison between migrants and natives more comparable, as the individuals are more homogeneous. Also, the data are accurate enough to track the inventors' mobility across companies, allowing us not only to focus on migration, but also on inter-company mobility.

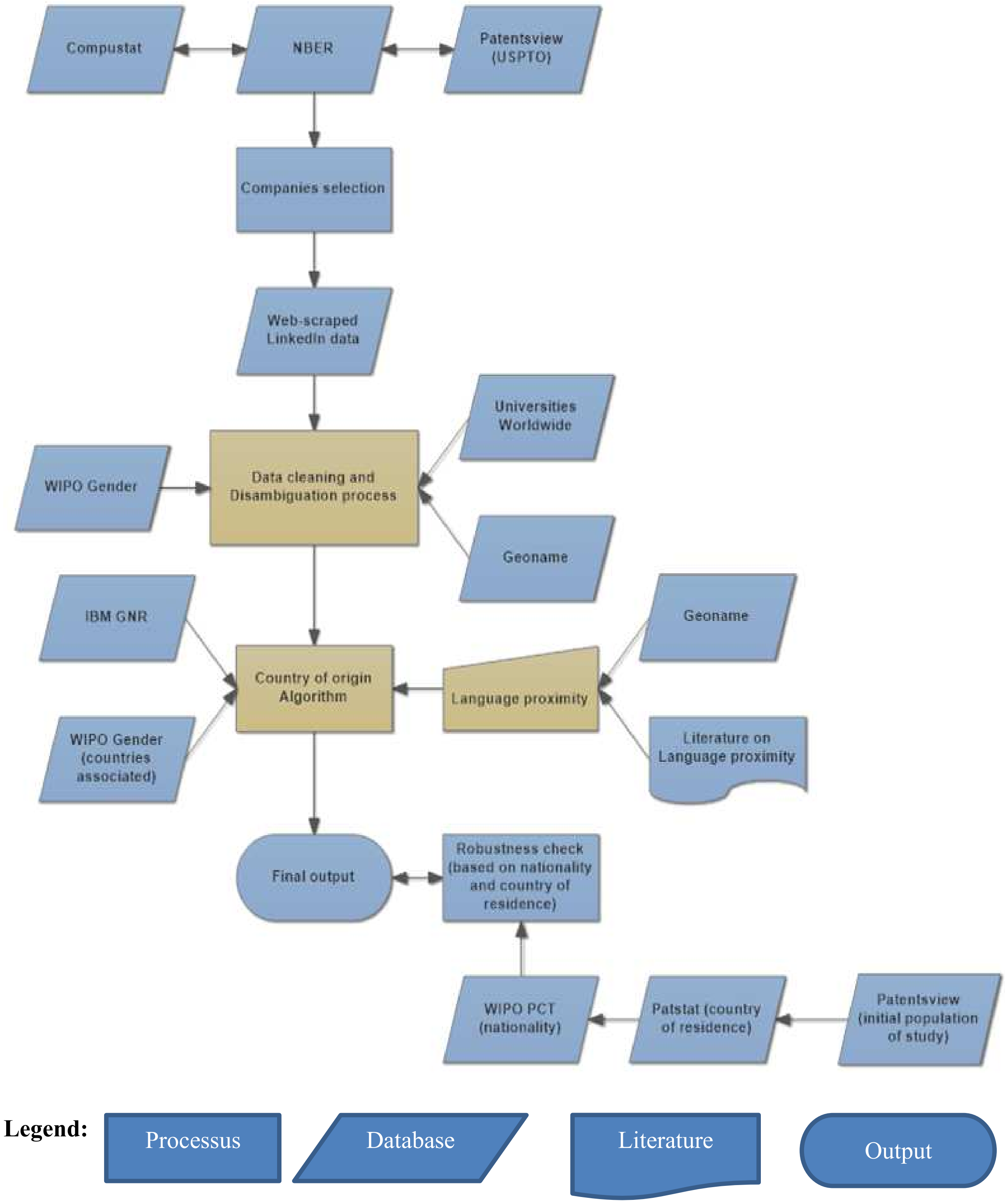
Besides the collected data, this chapter presented the Homeland algorithm, that is a new and original algorithm aiming at assigning a country of origin to each inventor. We refer to the literature on language proximity to refine the main ethnicities and find more under-represented ones. By doing so, one of the challenges that this algorithm solves is to differentiate individuals with Anglo-Saxons names in Canadians, Australians, Americans, Irish, and British.

The following empirical chapters use the Linked Inventor database as main data source. In Chapter 2, we compare the productivity of Indian inventors and natives in the United States, by controlling for company mobility. Thanks to the Linked Inventor database, we are able to break down the migrants' flows by entry channel (work or education) in the US, and to control for selection using the H1-B visa policy implementation in 1990 as an exogenous shock. In Chapter 3, we compare in the country of origin the productivity of European return migrants and stayers. We do so by taking into account the issue of return migrants double selection, that is positive when migrating and negative when returning.

1.6 Appendix

1.6.1 Appendix A: Description of the databases used

Figure 1.12: Databases used and main steps



a. Patentsview (USPTO)

PatentsView is a patent data visualization and analysis platform provided by the United States Patent and Trademark Office. Since 2012, the USPTO has built a newly developed database that links inventors, their organizations, locations and patenting activity. We use Patentsview (crossed with Compustat) to attribute the patents, hence the inventors of the 178 ICTs' companies and in a second time as a bridge with Patstat. Patentsview can be download in free access US patent data here: http://www.patentsview.org.

b. Patstat (EPO)

The European Patent Office (EPO) Worldwide Patent Statistical Database contains patent data about 90 national and international patent offices with different degree of coverage. Data include bibliographic data, citations and family links. Adding the coverage of Patstat with the data from Patentsview allow us to extend the production of patents for inventors not located in the US, the two databases can easily be linked using the patent application id here[11]: http://rawpatentdata.blogspot.com/search/label/patentsview and Patstat can be found here: https://www.epo.org/searching-for-patents/business/patstat.html#tab-1.

c. WIPO PCT

The World Intellectual Property Organization (WIPO) under the Patent Cooperation Treaty (PCT) assists applicants in seeking patent protection internationally for their inventions. However, what makes this database important for our concern is the inventor's nationality[12] that is recorded for each patents. Even though only few patents are filed under the PCT, we use this precious information as robustness check for the country of origin algorithm.

d. Compustat

The Compustat database created by Standard & Poors, presents financial, statistical and market information on active and inactive companies worldwide. We make the use of this database only during the selection of the US companies, in fact we selected the firms in ICT using the US SIC (Standard Industrial Classification) code provided. After selection we match them with the assignees in Patentsview. Compustat can find here: https://wrds-www.wharton.upenn.edu/pages/support/data-overview/wrds-overview-compustat-north-america-global-and-bank/.

[11] We would like to personally thank Gianluca Tarasconi for his help in that matter.
[12] We would like to personally thank Ernest Miguelez and Carsten Fink to have provided us the data.

e. NBER (Patent data project)

The National Bureau of Economic Research (NBER) patent data project uses US patent data for 1976-2006 associating the assignee with Compustat. Hence, we use this database to link the companies selected from Compustat with their patent on Patentsview. Here NBER Patent can be found in free access: https://sites.google.com/site/patentdataproject/Home/downloads.

f. Universities worldwide

We are using the Universities Worldwide database to enhance the accuracy of the geolocalisation procedure, by retrieving from this database the university country location. After matching with the universities of each profiles, we are adding to the request done via the API Google Map, this new information. Universities worldwide can be download in free access here: https://univ.cc/.

g. WIPO Gender

The WIPO Gender database gives for 174.849 unique names their gender and the countries associated to. We use this database for two reasons. First to use it for the gender algorithm, to give a gender to the inventors and, to retrieve the countries in order to feed our country of origin algorithm.

h. IBM GNR

The IBM Global Name Recognition (GNR) contains technologies to manage, search, analyze and compare name data sets by leveraging culture specific name data and linguistic rules that are associated with the name's culture. We make the use of IBM GNR for the country of origin algorithm, giving to the API the name and surname of our inventor's population, as output we gather a list of countries associated to each name and surname with statistics of occurrence. This association originates from a database produced by US immigration authorities in the first half of the 1990s, which registered all names and surnames of all foreign citizens entering the US, along their nationality. IBM GNR can be found here: https://www.ibm.com/support/knowledgecenter/en/SSEV5M_4.2.0/KC_ditamaps/pv_welcome_gnm42.html.

i. Geoname

The Geoname geographical database covers all countries and contains over eleven million placenames. The geographical information retrieved from this database are used for geolocalisation procedures as well as for the linguistic proximity algorithm by using the languages and their localization worldwide. Geoname can be download in free access here: https://www.geonames.org/about.html.

j. Zephyr

The Zephyr database created by Moody's records company deal information. We make the use of Zephyr by collecting merge, acquisitions from the 178 selected companies. We do so to only assign patents developed by the company and not acquired later. Zephyr can be download here: https://www.bvdinfo.com/en-gb/our-products/data/specialist/zephyr.

Table 1.8 : List of US public ICT companies

IDX	Company name	IDX	Company name	IDX	Company name
0	3COM CORP	**60**	EMULEX CORP	**120**	NETWORK APPLIANCE INC
1	ACTEL CORP	**61**	EXTREME NETWORKS INC	**121**	NETWORKS ASSOCIATES
2	ADC TELECOMMUNICATIONS INC	**62**	F5 NETWORKS INC	**122**	NOVELL INC
3	ADOBE SYSTEMS INC	**63**	FAIRCHILD SEMICONDUCTOR INTL	**123**	NUANCE COMMUNICATIONS INC
4	ADTRAN INC	**64**	FEI CO	**124**	NVIDIA CORP
5	ADVANCED MICRO DEVICES	**65**	FINISAR CORP	**125**	OMNIVISION TECHNOLOGIES INC
6	AFFYMETRIX INC	**66**	FIRST DATA CORP	**126**	ORACLE CORP
7	AGERE SYSTEMS INC	**67**	FORMFACTOR INC	**127**	PITNEY BOWES INC
8	AGILENT TECHNOLOGIES INC	**68**	FOUNDRY NETWORKS INC	**128**	PLANTRONICS INC
9	AKAMAI TECHNOLOGIES INC	**69**	FREESCALE SEMICONDUCTOR INC	**129**	PMC-SIERRA INC
10	ALTERA CORP	**70**	GATEWAY INC	**130**	POLYCOM INC
11	AMETEK INC	**71**	GENESYS TELECOMM LABS INC	**131**	POWER INTEGRATIONS INC
12	AMKOR TECHNOLOGY INC	**72**	GOOGLE INC	**132**	QLOGIC CORP
13	AMPHENOL CORP	**73**	HARMAN INTL INDUSTRIES INC	**133**	QUALCOMM INC
14	ANALOG DEVICES	**74**	HARRIS CORP	**134**	QUANTUM CORP
15	APPLE INC	**75**	HEWLETT-PACKARD CO	**135**	QWEST COMMUNICATION INTL INC
16	APPLIED MICRO CIRCUITS CORP	**76**	HUTCHINSON TECHNOLOGY INC	**136**	READ-RITE CORP
17	ARRIS GROUP INC	**77**	I2 TECHNOLOGIES INC	**137**	RED HAT INC
18	AT&T CORP	**78**	IMMERSION CORP	**138**	RESEARCH IN MOTION LTD
19	AT&T INC	**79**	INFINERA CORP	**139**	ROGERS CORP
20	ATHEROS COMMUNICATIONS INC	**80**	INTEGRATED DEVICE TECH INC	**140**	SANDISK CORP
21	ATI TECHNOLOGIES INC	**81**	INTEL CORP	**141**	SCIENTIFIC-ATLANTA INC
22	ATMEL CORP	**82**	INTERMEC INC	**142**	SEAGATE TECHNOLOGY
23	AUTODESK INC	**83**	INTERSIL CORP	**143**	SENSORMATIC ELECTRONICS
24	AVANEX CORP	**84**	INTL BUSINESS MACHINES CORP	**144**	SIGMATEL INC
25	AVAYA INC	**85**	INTL RECTIFIER CORP	**145**	SILICON GRAPHICS INC
26	BEA SYSTEMS INC	**86**	INTUIT INC	**146**	SILICON IMAGE INC
27	BECKMAN COULTER INC	**87**	IOMEGA CORP	**147**	SILICON LABORATORIES INC
28	BELL & HOWELL OPERATING CO	**88**	IXYS CORP	**148**	SILICON STORAGE TECHNOLOGY
29	BELLSOUTH CORP	**89**	JUNIPER NETWORKS INC	**149**	SILICONIX INC
30	BIO-RAD LABORATORIES INC	**90**	L-3 COMMUNICATI ONS HLDGS INC	**150**	SKYWORKS SOLUTIONS INC
31	BMC SOFTWARE INC	**91**	LATTICE SEMICONDUCTOR CORP	**151**	SPANSION INC
32	BROADCOM CORP -CL A	**92**	LEVEL 3 COMMUNICATIONS INC	**152**	STANDARD MICROSYSTEMS CORP
33	BROCADE COMMUNICATIONS SYS	**93**	LEXMARK INTL INC - CL A	**153**	STORAGE TECHNOLOGYCP
34	CA INC	**94**	LINEAR TECHNOLOGY CORP	**154**	SUN MICROSYSTEMS INC
35	CADENCE DESIGN SYSTE MS INC	**95**	LORAL SPACE & COM MUNICATIONS	**155**	SYBASE INC
36	CASCADE MICROTECH INC	**96**	LSI CORP	**156**	SYMANTEC CORP
37	CERTICOM CORP	**97**	LUCENT TECHNOLOGIES INC	**157**	SYMBOL TECHNOLOGIES
38	CIENA CORP	**98**	MAXIM INTEGRATED PRODUCTS	**158**	SYMYX TECHNOLOGIES INC
39	CIRRUS LOGIC INC	**99**	MAXTOR CORP	**159**	SYNAPTICS INC
40	CISCO SYSTEMS INC	**100**	MCI INC	**160**	SYNOPSYS INC
41	CITRIX SYSTEMS INC	**101**	MENTOR GRAPHI CS CORP	**161**	TEKTRONIX INC
42	COGNEX CORP	**102**	METHODE ELECTRONICS - CL A	**162**	TELECOMMUNICATION SYS INC

Continues at page 43

43	COHERENT INC	**103**	METROLOGIC INSTRUMENTS INC	**163**	TELLABS INC
44	COMMVAULT SYSTEMS INC	**104**	MICREL INC	**164**	TERADYNE INC
45	CONEXANT SYSTEMS INC	**105**	MICROCHIP TECHNOLOGY INC	**165**	TEXAS INSTRUMENTS INC
46	CORNING INC	**106**	MICRON TECHNOLOGY INC	**166**	TRIQUINT SEMICONDUCTOR INC
47	CREDENCE SYSTEMS CORP	**107**	MICROSEMI CORP	**167**	UNISYS CORP
48	CREE INC	**108**	MICROSOFT CORP	**168**	UNIVERSAL DISPLAY CORP
49	CYPRESS SEMICONDUCTOR CORP	**109**	MICROVISION INC	**169**	UNIVERSAL ELECTRONICS INC
50	DALLAS SEMICONDUCTOR CORP	**110**	MINDSPEED TECHNOLOGIES INC	**170**	VARIAN INC
51	DELL INC	**111**	MITEL NETWORKS CORP	**171**	VIASAT INC
52	DIEBOLD INC	**112**	MKS INSTRUMENTS INC	**172**	WESTERN DIGITAL CORP
53	DIGIMARC CORP	**113**	MOLEX INC	**173**	WORLDCOM INC - CONSOLIDATED
54	DIRECTV GROUP INC	**114**	MONOLITHIC POWER SYSTEMS INC	**174**	XEROX CORP
55	EBAY INC	**115**	MOTOROLA INC	**175**	XILINX INC
56	ECHOSTAR CORP	**116**	NATIONAL INSTRUMENTS CORP	**176**	YAHOO INC
57	ELECTRONIC DATA SYSTEMS CORP	**117**	NATIONAL SEMICONDUCTOR CORP	**177**	ZILOG INC
58	ELECTRONICS FOR IMAGING INC	**118**	NCR CORP	**178**	ZORAN CORP
59	EMC COR P/MA	**119**	NETLOGIC MICROSYSTEMS INC		

1.6.2 Appendix B: country of origin whole list

Table 1.9: Predicted origin's country

ID COO	Country of origin	Frequency / Percentage	ID COO	Country of origin	Frequency / Percentage
1	US	58671	70	IS	11
2	IN	8272	71	SI	11
3	CN	3539	72	TN	11
4	GB	3256	73	IQ	10
5	CA	3193	74	OO	10
6	IL	2009	75	BA	9
7	DE	1809	76	EE	9
8	FR	1364	77	NP	9
9	JP	990	78	ET	8
10	TW	678	79	AM	7
11	IT	677	80	GT	7
12	AU	657	81	JM	6
13	RU	581	82	LU	6
14	IE	569	83	PA	6
15	KR	520	84	TT	6
16	NL	427	85	LT	5
17	MX	362	86	MT	5
18	SG	339	87	QA	5
19	BR	318	88	AE	4
20	PH	302	89	CY	4
21	MY	298	90	KW	4
22	CH	262	91	MK	4
23	TR	261	92	SV	4
24	ES	243	93	TZ	4
25	DK	239	94	AL	3
26	RO	219	95	CU	3

Continues at page 44

27	SE	212	96	KE	3
28	EG	199	97	ME	3
29	IR	194	98	MM	3
30	BE	160	99	UY	3
31	HK	157	100	XX	3
32	PK	144	101	ZM	3
33	ZA	141	102	AF	2
34	PL	134	103	DO	2
35	NZ	123	104	EC	2
36	GR	118	105	GH	2
37	AT	113	106	LA	2
38	FI	113	107	LV	2
39	VE	101	108	MD	2
40	AR	92	109	SN	2
41	NO	86	110	SO	2
42	CZ	83	111	SR	2
43	BD	82	112	ZW	2
44	PR	68	113	AO	1
45	VN	67	114	AZ	1
46	UA	65	115	BH	1
47	BG	63	116	BM	1
48	HU	61	117	BO	1
49	TH	54	118	CD	1
50	ID	48	119	CM	1
51	LB	43	120	DM	1
52	CO	40	121	GE	1
53	RS	40	122	GU	1
54	JO	39	123	HN	1
55	BS	32	124	HT	1
56	PT	28	125	KP	1
57	HR	22	126	KZ	1
58	CR	19	127	MO	1
59	LK	18	128	MR	1
60	BY	17	129	MU	1
61	CL	16	130	MZ	1
62	NG	14	131	NI	1
63	SK	14	132	OM	1
64	SY	13	133	SC	1
65	YU	13	134	SD	1
66	DZ	12	135	UZ	1
67	MA	12	136	VC	1
68	PE	12	137	VI	1
69	SA	12			

Chapter 2

Are migrant inventors more productive than native ones? A study on Indian inventors in the US

We contribute to the literature on migration and innovation by comparing the productivity of foreign (Indian) and native ICT inventors in the United States, as measured by the number of patents filed and the number of citations received. As a baseline result we find migrants to be more productive than natives, but observe that, in the destination country, migrant inventors tend to change employer more often than natives, which may be a relevant concurrent factor. This requires that we control not just for migration motives such as education, access to the labour market, and cohort effects, but also for mobility across companies and within them (mobility across different locations of the same company). We do so by exploiting a rich database obtained by merging information on inventors coming from both patents and social media. In particular, we compare migrant inventors to both within-company mobile and non-mobile natives and find them to be more productive than both control groups. At the same time, migrant inventors who entered in the United States via an internal re-assignment within the same company perform better than migrants who changed company or entered for education reasons.

2.1 Introduction

Highly skilled international migration raised in recent decades (Docquier & Rapoport, 2012). In the United States, 19% of the tertiary-educated population was foreign-born in 2013, and specific fields such as science, technology, engineering, and mathematics (STEM) more than 30% of graduates were foreign-born (Ruggles, et al., 2010). The direct contributions of migrants in the host country are nowadays well-known, mainly thanks to a quite rich literature on the United States. Compared to a foreign-born population of 12% in 2000, 26% of US-based Nobel Prize recipients from 1990-2000 are migrants (Peri, 2007) and 25% of new high-tech companies with more than one million dollars in sales in 2006 were founded by foreign-born entrepreneurs (Wadhwa, et al., 2007). Stephan and Levin (2001) show that migrants are over-represented among members of the National Academy of Sciences and the National Academy of Engineering as well as among highly-cited authors. Migrants also contribute exceptionally to the host country's patenting activity. Kerr (2007) shows that the share of US patents awarded to US-based inventors with Chinese and Indian names account for 12% of the total in 2004.

Over the past 20 years, the percentage of scientists and engineers on the total flows of global migration has increased substantially (Freeman, 2010). These flows are fed by an increasing number of countries of origin, most notably China, India, and former soviet-bloc countries. Most of the studies have focused on the United States as a destination country. Kerr (2008) finds that there is a growing contribution by ethnic minorities to US domestic patent filings, primarily by migrants of Chinese and Indian ethnicity who have become a relevant part of US inventors in ICTs sectors. Hunt and Gauthier-Loiselle (2010) show that migrant inventors in the US generally perform better, in terms of patenting rates, than the US-born ones because they are more highly educated and with degrees in Science, Technology, Engineering and Mathematics (STEM). Ruggles and al. (2010) show that in the US about 19% of the working-age population with a bachelor's degree or higher were foreign-born in 2013, and in STEM more than 30% were migrants. In STEM-related jobs, the share of US immigrant workers from India accounts for one-third of all foreign-born workers in 2010-2012 (Hanson & Liu, 2017). Most of the empirical investigations have shown that high-skilled migrants increase their destination country's innovation potential. Although, with two exceptions: Kerr and Lincoln (2010) find that, at the city level, high-skilled migrants from India and China increase the overall patenting rate, but do not increase the patenting rate of natives; and Borjas and Doran (2012) find that Soviets migrants in the field of mathematics have a negative impact on the productivity of American natives in the same field. Positive results appear to be in line with a more general evidence on cultural diversity and various economic performance measures at the country, region and city levels (Ozgen, et al., 2012; Ottaviano & Peri, 2006; Alesina, et al., 2016) and at the state level (Hunt & Gauthier-Loiselle, 2010).

We contribute to this literature by assessing whether migrant inventors are more productive than natives. Even though the determinants of inventors' productivity are nowadays well known (Hoisl, 2007; Hoisl, 2009; Latham, et al., 2011; Zwick, et al., 2017), research comparing migrants and natives' inventors' productivity are still scant and usually rely on a limited set control variables to account for differences among migrants and natives' characteristics. We add new evidence to this literature, by controlling for an important and often neglected source of heterogeneity: the inventors' intra-country mobility experience (changing of company) for both the natives in their home country and the migrants at destination, hence putting together the literature on mobility and productivity with the one on migrants' productivity at destination. Finally, since the productivity of migrants may be affected by the type of entry in the destination country, we are breaking down migrants according to their entry channel in the destination country such as: education, inter-firm and intra-firm (multinational) channels. Finally, we correct for endogeneity issue by controlling for self-selection using the H1-B visa program.

Besides the previously cited literature, one of the main reasons for the scarcity of individual-level studies comparing the productivity between migrants and natives is the unavailability of appropriate data on inventors' characteristics especially in the early stages of their career. Many papers have exploited patent data and focused on inventors as a representative category of STEM workers; but they have not obviated to the major limitation of this type of data, namely the sparse information they provide on the inventors' biographical data. When it comes to their use for studying migration and innovation, they do not provide information on their date and place of birth and/or nationality and education. This makes it difficult to ascertain their migrant vs. native status and, in the case of migrants, their countries of origin. While name analysis, as in Kerr (2007) and Breschi et al. (2014) can obviate to this problem, it is a solution that works well with some countries of origin, whose linguistic distribution is very different from that of destination (as with Indian migration to the United States, to name a very important bilateral flow). But not for the others (as with other important flows such as those from Great Britain to the United States or from Germany to Switzerland). An additional problem with patent and inventor data is that they do not record the inventors' career moves to employers for which they do not file any patent and in particular no career move at all for inventors with just one patent; and yet they may be relevant for both migration and innovation studies to the extent that their contribution to innovation and knowledge diffusion may go beyond the patent(s) they sign. This solves one of the major limitations of previous studies on mobility and productivity of inventors from patent data, which were able to track the mobility only for those inventors with at least two patents, based on differences in the addresses reported in one and the other documents (Hoisl & de Rassenfosse, 2014; Hoisl, 2007; Hunt & Gauthier-Loiselle, 2010).

Finally, in the absence of information on the inventors' education, it is difficult to distinguish between the different entry channels in the destination country, whether through higher education institutions. In all these respects, the database we have presented in chapter 1 fills many gaps, as it provides information on education, labour market activity and patenting activity of inventors who we can classify as native or migrant (with details on the potential country of origin). Although a large pool of available options, we decide to focus on Indian migrants in the US for two reasons. First, for the database building we chose to focus on a highly innovative sector as ICTs, that is also the one where the Indians migrants are the most represented, alongside the Chinese, in the US. Second, we prefer to focus on Indian rather than Chinese for a statistical reason. The Indians waves of high-skilled migration are older than the Chinese ones that start in the 2000s. Instead of the 90s for the Indian ones, hence, we can follow, on average, an Indian longer than a Chinese in the US.

From the database, we extract a large sample (40.806 individuals) of inventors with at least one patent filed in an ITC –related technological field from an address in the United States. We have 36.010 are natives, and 4.796 are Indian's migrants, observed during the period between 1969 and 2016. Find the Worldwide (Appendix E, Figure 2.E.1) as well as the US localization (Appendix E, Figure 2.E.2) of the Indians inventors from whom we have the LinkedIn profile.

Our baseline results, where we only distinguish between migrants and natives, suggest that migrants are more productive than natives. However, an individual can decide to move into the destination country through three different channels. The first one is education that the migrant can choose to partially or entirely acquire at destination. If the individual completed his or her education in the country of origin, he could move in the host country by changing of company or by a re-allocation within the same company (if, for instance, he is working for a multinational).

When we break down the migrant population by entry channel, we find that migrants entering the US via re-allocation are more productive than natives and migrants that entered through the other channels discussed earlier (education or company change). To account for selection issues, we exploit the H1-B visa program, a policy established in the 90s that aimed at regulating the number of migrants entering in the US. In a first time, we do not find any difference in productivity between migrants and natives. We then exploit the yearly ratio of the number of available H1-B visa and the number of visa seekers to proxy the intensity of selection. When we take into account this measure, we find that the cohorts of migrants entering in the US when the intensity of selection was higher performed better than natives and the migrants that entered in a period where the selection was less intense.

This chapter is organized as follows. We first provide an overview of the background literature in section 2, in section 3, we present the data alongside some descriptive statistics, followed by the model specification, estimation, and the results. In section 4, we perform some robustness checks. Finally, section 5 summarizes the findings and draws some conclusions.

2.2 Literature review

2.2.1 Migration of high-skilled

Research on the economics of labour migration, despite his long and controversial history, has undergone in the past years significant and crucial transformation. Emigration today is a contentious issue in most of the countries. Public debate and policies have become characterized by the contrast between restrictive attitudes towards low-skilled migration and more welcome views for migration of high-skilled[13] workers. An increasing number of developed and developing countries are currently in the process of redesigning their immigration systems and implementing various skill-selecting and attracting policy instruments. Since the 1990s, countries have intensified their efforts to attract and retain international talent, driven by the general perception that highly skilled workers are contributing to receiving countries, by fostering innovation and competitiveness, hence contributing to the host country economic growth. This has resulted in a global diffusion of high-skilled migration policies. In 2015, almost half of the 172 UN member states declared an explicit interest in increasing the level of high-skilled migration, by attracting foreign or retaining native talent (Czaika, 2018). Mostly Western countries are at the vanguard of this global trend, with two-thirds of OECD countries have implemented or implemented policies targeting, mostly, high-skilled migrants (Czaika & Parsons, 2017). The rapid aging of the population as well as the increasing skill shortages in highly developed countries, but also in newly emerging economies such as Brazil, Russia, India, and China have led such states to implement policies targeting high-skilled workers. The emergence of a global labour market for occupations in high demand has led to a process of mutual selection between skilled migrants and targeting states (Chiswick, 2011).

International migration of high-skilled has risen in recent decades (Docquier & Rapoport, 2012). Between 2000 and 2006, the United States attracted 1.9 million of tertiary-educated migrants, and, European OECD countries attracted 2.2 million (Widmaier & Dumont, 2011). In percentage of the tertiary-educated population, high-skilled migrants represented about 11% in OECD countries (Bertoli, et al., 2009) in 2000. Nevertheless, the distribution of skilled migrants in OECD countries is very uneven, with the US economy hosting almost half of all the skilled migrants. This clearly shows that the US represents by far the preferred destination among skilled migrants (Kerr, et al., 2017).

Nevertheless, studies comparing the foreign-born inventors' productivity with the native's ones are still scant.

[13] The term "skilled" is often referred in terms of education attainment, a skilled migrant is considered as an individual with any tertiary education.

The analysis of highly skilled workers' determinants of productivity (often measured with the wage) undertake by other streams of literature, more focus on the assimilation and positive selection of mobile workers.

2.2.2 Self-selection and assimilation

Emigrants are not randomly selected from the population of the country of origin, and insofar they may either self-select based on both observable and unobservable characteristics (such as education or non-quantifiable skills, respectively) or affected by skill-selective immigration policies.

On self-selection, the so-called Roy model (Borjas, 1987) proposes that the relative returns to human capital in the country of origin and destination country determine whether high or low skilled individuals tend to migrate. Borjas finds that the higher are the returns to human capital in a country, the more high-skilled individuals tend to migrate in this country. This hypothesis is confirmed empirically by (Moraga, 2011; Dustmann & Glitz, 2011; Parey, et al., 2017).

The economics of assimilation, that starts with Chiswick (1978), finds that earnings of newly arrived migrants are significantly lower than those of natives with the same observed characteristics because migrants' skills are not always transferable or immediately valued in the host country. The literature attributes assimilation to the positive selection of migrants for their innate ability, high motivation for labour market success, and their higher incentives to invest in the host country's specific human capital. Indeed Carliner (1980) Borjas (1982) find that migrant earnings reach parity with native ones within ten years of residence, exceeding it afterward. Questioning the earliest empirical results, Borjas (1985) finds that the assimilation effects confound with cohort effects; in fact, the assimilation effect on earnings of migrants with different ages can overstate if the quality of more recent migrant cohorts is lower than older ones. Consequently, Borjas (1985) finds that migrants may not assimilate as rapidly as the former view hypothesized, and earnings of more recent cohorts may never reach parity with the earnings of natives. In recent years, other researchers explore the contribution of highly skilled migrants to invention and innovation in their host country. The broad literature on the relationship between age, cohort and period, and inventive performance shows that heterogeneity between highly skilled individuals is persistent over the life span, and that performance may decline with higher age and is lower for older cohorts (Turner & Mairesse, 2005; Hoisl, 2007; Lissoni, et al., 2013).

Nevertheless, Walsh and Nagaoka (2009) find a positive relationship between age and inventive performance, due to a selection process, according to which only the most successful workers remain in research and development activities. Previous findings of the H-1B visa impact on the US patenting activity seems to corroborate the thesis previously cited. Hunt (2011) shows that college-educated migrants outperform college-educated natives in wages, patenting, and publishing, especially for migrants who initially arrived on student or work visas. Kerr and Lincoln (2010), using a reduced-form framework, find that an increase in the national H-1B visa population also increases patenting in cities with a high dependence on H-1B visa workers, especially by inventors of Chinese and Indian ethnicity.

2.2.3 Brain drain and Brain gain

The emigration of highly skilled workers from their country of origin to another country may result in a depletion of knowledge and abilities in the former. Traditionally, the low- and middle-income countries, the less developed economies, are more affected by this phenomenon, especially when they are already suffering from a severe scarcity of human capital endowments. The term brain drain implies that the net flows of talented people are unbalanced in one direction (Koser & Salt, 1997).

The brain drain can be seen as a negative externality on the population left in the source country (Bhagwati & Hamada, 1974), due to imperfect substitution between skilled and unskilled labour. The negative impact of the brain drain has, for long been stressed in the growth literature. Most studies underline the positive effects of migration on human capital formation but turning to a negative effect on growth (Miyagiwa, 1991; Haque & Kim, 1995).

In the 1990s, however, a more nuanced view of skilled migration has emerged, which emphasize several channels through which skilled emigration can be advantageous for sending countries. For instance, the contribution of migrants' remittances to their home countries' GDP and the reduction of inequalities that may go with it are now widely recognized (Barham & Boucher, 1998; Adams, et al., 2005). There has also been a growing interest in the role of migrating knowledge workers as carriers of international knowledge flows. The importance of knowledge diffusion for innovation and productivity growth is nowadays well known. According to Saxenian (Saxenian & Quan, 2002; Saxenian, 2006), migrant entrepreneurs and their communities provide a significant mechanism for the international diffusion of knowledge and upgrading of local capabilities. In some parts of the world, the high skilled emigration has been effectively converted into brain circulation as talented engineers and scientists return home to pursue promising opportunities.

2.2.4 Migrant's contribution to innovation at destination

Skilled migrants bring with them specific competences earned during their education, and their direct impact on innovation activities in the destination country is expected to be positive. Using individual data on workers in Science and Engineering mostly in the United States, a large set of studies shows that the total number of inventions increase through the contribution of migrant inventors (Stephan & Levin, 2001; Chellaraj, et al., 2005; Hunt & Gauthier-Loiselle, 2010; Kerr & Lincoln, 2010). In the US, migration policies and self-selection drivers make, on average, migrants more educated than natives (Batalova & Fix, 2017), and, more likely to work in Science and Engineering occupations (Hunt & Gauthier-Loiselle, 2010). Self-selection plays a positive role: if the migrant workforce is more motivated or has better competences than the native one, the migrant's contribution to innovation is enhanced.

Another part of the literature on migrants' contribution at destination explain the positive effect of migration focusing on the impact of ethnic diversity on innovation. Evidence suggests that firms with ethnically diverse workforce tend to be more innovative (Parrotta, et al., 2014). At the individual level, Ferrucci & Lissoni (2019) found a positive relationship between team diversity and patent quality as well.

2.2.5 Mobility and productivity

Employees' mobility affects the performance of both the source and recipient organizations, as well as that of the employees themselves (Rosenkopf & Almeida, 2003; Somaya, et al., 2008; Campbell, et al., 2012).

Employee mobility is facilitated by collaborative work, which can, in turn, enhance innovation outcomes (Feldman, et al., 2012). Analyzing a sample of US inventors patenting at the European Patent Office (EPO), (Breschi, et al., 2010) suggests that the mobility of inventors is an important knowledge diffusion channel when they generate new social networks in the recipient areas. Breschi and Lenzi (2010) offer evidence that job moves lead to an effective transfer of knowledge as measured by patent citations. Parallel to these studies, another line of research has studied mobility by focusing on inventors' performance. For instance, mobility and productivity relationships have been considered by Lenzi (2010) for Italian, by Hoisl (2007) for German and by Shalem and Trajtenberg (2009) for Israeli inventors.

Nevertheless, the literature on the impact of individual mobility on individual performance is conflicted. On the one hand, it is not clear empirically whether individual productivity has a positive or negative effect on the likelihood to move (Hoisl, 2007; Campbell, et al., 2012; Di Lorenzo & Almeida, 2017). On the other hand, it is not evident whether changing of employer increases or decreases the individual productivity of the moving employee (Hoisl, 2007; Fernandez-Zubieta, et al., 2015). The lack of consistency in the results provided in the previous studies indicates that mobility is a complex phenomenon with uncertain effects on individual performance.

2.3 Data source and sample

2.3.1 Data description

In this section, we discuss the methodological approach to test our hypothesis, namely that migrant inventors are more productive than native ones. First, we decided to focus on just one destination and one country of origin of migrants, respectively the United States and India. That is, our analysis concerns only US-resident inventors from both the United States itself and India, although it includes information on both classes of inventors' mobility patterns within both the United States and abroad.

We assign inventors the status of migrants when their country of origin (variously defined as the country of birth and/or the country of primary/secondary education, depending on availability of information) is different from the country where the inventive activity takes place[14]. Furthermore, the migration status of individuals does not change over time, since we observe their productivity only in the United States[15]. The database we use is subset of the Linked Inventors database described in chapter 1.

The ultimate goal of the Linked Inventor database is to enrich the inventor information one can retrieve from patent data (address at the time of the patent, name of applicant, identity of co-inventors, and other patent contents) with information on the migrant vs native status of the inventors (plus, for migrants, their country of origin and year of entry in the US), as well as information on education and labour market experience extracted from online social network (in particular, LinkedIn). Notice, in this chapter, however that we restrict our attention to inventors who have been active at least once in the US (that is, we do not consider inventors appearing on patents that do not report at least one inventor's address in the United States, even when their LinkedIn profile mentions some work experience there).

To assign a country of origin to inventors, the Linked Inventor datasets exploits both the information from LinkedIn (such as the country where the earliest education degree has been attained, the individual's native language, and any useful biographical detail) as well as the following patent information:

- The inventor's nationality, as reported on a subset of USPTO patent applications filed according to the PCT procedure before 2011 (Miguelez & Fink, 2013).
- The results of name analysis, based on the combination of statistics on the ethnolinguistic origin of names and surnames from the IBM-GNR dataset (Breschi, et al., 2014) as well as additional linguistic analysis (Tyshchenko, 1999).

These data also allow us to track a substantial part of the inventors' careers, most notably their mobility before the first and after the last patent filed (the two dates coincide for the vast number of inventors with just one patent over their lifetime).

Another advantage of the Linked Inventor dataset is that it allows us to identify and classify different types of mobility, besides migration, such as:

1) Education mobility through changing of university (within or not the same country): when the inventors move during their education path, including the different universities in which they have studied.

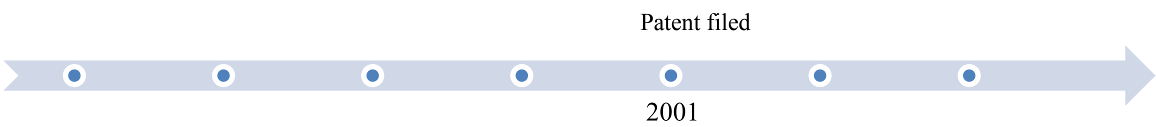
[14] We assess the country of origin of the inventor using a newly developed algorithm (Chapter 1; Section 1.4) that exploits the variety of information gathered from LinkedIn like the education localisation or the spoken language, as well as the language proximity literature and the existing literature that is analysing the name and surname to guess the country of origin.
[15] Nevertheless, thanks to our database we observe and control for other characteristics when the inventors were in their country of origin.

2) Labour market mobility through changing of company (within or not the same country): when inventors move during their professional career and the number of different companies where they have worked.
3) Using 1) and 2) we can also distinguish whether the inventors who migrate (mobile across countries) do so for education or work reasons and whether the move abroad coincides with patenting there, or instead precede/follow it.

Furthermore, the rich information about the inventor's lifespan allows us to progressively enrich our information on the inventors in our sample, thus obtaining three samples, all of them with the same observations (inventors) but an increasingly precise information on each inventor move across countries, companies, and educational institutions, including information on left and right censoring/truncation. We will then run the same econometric test on each different sample, so to evaluate how much the results change thanks to the information we can get from LinkedIn, on top of the most basic information provided by patent data.

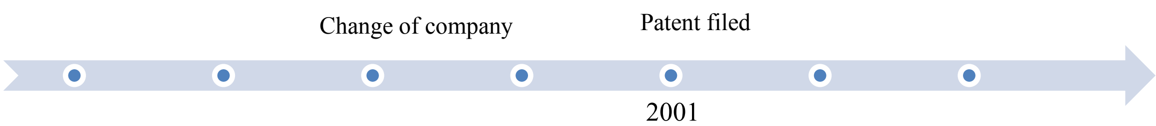
When using patent data only, one can detect inventors' moves across companies only by comparing the patent applicants. Besides, one cannot investigate occasional inventors, those who patented only once, such as inventor A in figure 2.1.

Figure 2.1: Inventor A, using patent data only



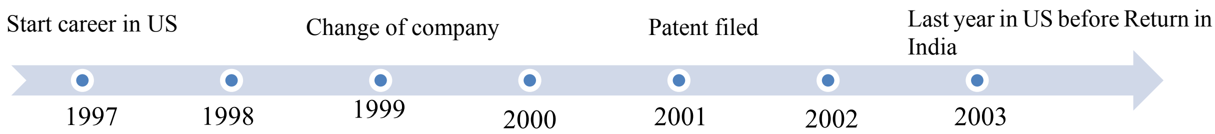
By, adding information from LinkedIn we can detect the possible moves of occasional inventors either before the patent, as in figure 2.2, afterwards. Even in cases in which we do not know, or cannot use, the exact year of the move, this suffices to establish whether an inventor is either a migrant or not and/or whether he/she has experience some work mobility.

Figure 2.2: First sample: Inventor A, using patent + LinkedIn data



Finally, when we can get and use the whole set of LinkedIn information we can track exactly all of the inventor changes of company as well the entry-to/exit-from the United States.

Figure 2.3: Second sample: Inventor A, using patent + LinkedIn data and extending the observations beyond the patent date



It is important to stress that the longitudinal nature of the data allows us to deal with some endogeneity issues and helps to establish a causal link between our independent variables of interest and productivity. The final panel consists of 40.806 inventors for which we have full information on education, labour market, and patenting. In the sample described Figure 2.2, this corresponds to 253,239 inventor-time observations. In the sample described in Figure 2.3, where we retain ore data points for each inventor, of the inventor-time observations are 912,526.

For a robustness check, in section 4, we will truncate our observations to the right, at the time of filing of the last patent, to account for some non-observable events that could occur after then, such as retirement. In this case, we reduce the time data points and the inventor-time observations drop to 570,019.

2.3.2 Descriptive statistics

Table 2.1 presents some general descriptive statistics. The sample consists of 253.239 observations for 40.806 inventors, 88.15% of which are natives and 11.85% are Indians migrants (dataset in Figure 2.3). When we break down Indians migrants by entry channel, we see that 65% of them enter the US for education motives, while 35% enter directly on the US labour market. Among the latter, 19% enter through internal mobility within a multinational compnay and 81% by changing employer. Besides migration status, information about mobility after education indicate that 73% have changed company at least once. Finally, each inventor has filed, on average, 13 patents[16] all among his/her career (or until 2016), with an average of 290 citations[17] received. Hoisl (2007), whose sample includes only inventors with at least two patents, has, on average, 14.7 patents by inventor.

Our sample is composed mainly of male inventors (88%) aged between 25 and 77 (with an average age of 38 in 2002). This first statistic shows that our sample has a population of younger inventors with respect to other studies (Trajtenberg, 2005; Hoisl, 2007; Hunt, 2004; Kerr, 2008), which had a sample of older inventors (on average) for instance Hoisl (2007) found an average age in 2002 of 54.04 years. Concerning inventors' behavior after having completed their education, 65% of the migrants stayed in the country of destination after having completed their education. The level of education is almost equally distributed, 32% have a bachelor, 36% a master and 31% have a Ph.D. or an MBA, this is similar

[16] We consider all granted patent family from the EPO Patstat database.
[17] We consider all the forward citations received by the patent family up to 2016.

to the findings by Hoisl (2007) where 36% of her population was holding a Ph.D., and 52% was holding a Bachelor or a Master degree. The sample is composed mostly of engineers (63%), managers (17%) and almost 6% of University scholars. The remaining 12% are split between inventors working as head or founder of a company (CEO, President, Director…) and others such as consultant or support functions (Accounting, Finance or Human resources).

Table 2.1: Sample general characteristic

Descriptive statistics (N=40.806)

Variable	Mean	S.D.	Min	Max
Migration status				
% Native	0.88		0	1
% Migrant	0.12		0	1
- Education motives	0.65		0	1
- Labour motives	0.35		0	1
Within company	0.19		0	1
Across company	0.81		0	1
Mobility				
Total Nbr. Interfirm mobility made	1.33	1.56	0	38
% Interfirm mobility	73.09		0	1
Other characteristics				
Total Nbr. of patents made	13.36	20.4	1	443
Total Nbr. of citations received	290.56	553.2	0	14746
Age of the inventors in 2002	38.19	7.51	25	77
% Gender (1 = male)	0.88		0	1
Level of education				
- Bachelor	0.32		0	1
- Master	0.36		0	1
- Ph.D./MBA	0.31		0	1
Title				
- Engineer	0.63		0	1
- Manager	0.17		0	1
- Company's head	0.07		0	1
- Scholar	0.06		0	1
- Founder	0.02		0	1
- Others	0.03		0	1

Table 2.2 provides the main variable of interest separated for Natives (column 3) and Migrants (column 6) representing respectively 88.15% and 11.85% of the whole sample. Then, we decompose the status of migration for two different kinds of mobility: stayers and movers. Columns (1) and (4) represent the native/migrant subsample that never changed company while columns (2) and (5) show the native/migrant subsample that at least have changed company once.

As Table 2.2 shows, changing company is a much more common phenomenon than changing country, with both natives and migrants having changed company at least one (65% and 71%, respectively). Similar findings by Lenzi (2010) confirm our results: in her study using survey data about Italians inventors, she found that 64% of them changed of company at least once while using patent data only 23% of them were company mobile.

Comparing the education level, we notice that migrants (column 6) are, on average, more educated than natives (column 3): 39% of the migrants have a Master degree and 41% a Ph.D., while 36% of the natives have a Master and only 26% a Ph.D. This preliminary result suggests a positive selection of migrants based on education. We are using two different measures for inventor's productivity: the number of patents he/she signed, and the number of citations received by those patents (forward citations).

First, we can observe that natives are, on average, less productive than migrants: a native file on average 13.28 patents along with his/her career while a migrant produces 14.16. This gap appears especially significant when we consider that, on average, we observe migrants for a shorter time than natives. In fact, a migrant's average patenting career length in the US is shorter than that of a native (3.8 versus 5.4 years, respectively; see Appendix A, Table 2.A.1)[18]. As for the overall work career in the US, this is about 22 years for a native and less than 15 years for a migrant (While Appendix B, Table 2.B.1).

Second, company movers are, on average, significantly more productive than non-movers for both the native (12.14 vs. 13.71 patents) and the migrant population (13.31 vs. 14.39 patents). We also observe a significant difference between native movers (13.71) and migrants that changed of company (14.39). Instead, when it comes to patent quality, which we measure with the number of patents' forward citations, it is natives who outperform migrants, we observe a significant difference between natives (291.5) and migrants (281.5).

Table 2.2: Natives and Migrants broke down by Inter-mobility status

Descriptive statistics (N=40.806)

	Natives			Indian Migrants		
	Stayers (1)	Movers (2)	All (3)	Stayers (4)	Movers (5)	All (6)
No. of inventors	12442	23568	36010	1386	3410	4796
Total Nbr. of patents	12.14	13.71	13.28	13.31	14.39	14.16
Total Nbr. of citations	240.4	310.9	291.5	232.6	294.9	281.5
Gender (1 = male)	0.87	0.90	0.89	0.80	0.85	0.84
Level of education						
- Bachelor	0.36	0.33	0.34	0.09	0.11	0.11
- Master	0.37	0.36	0.36	0.44	0.43	0.43
- Ph.D./MBA	0.26	0.30	0.28	0.47	0.45	0.46
Title						
- Engineer	0.67	0.61	0.63	0.59	0.59	0.59
- Manager	0.18	0.16	0.17	0.22	0.19	0.19
- CEO	0.05	0.09	0.07	0.05	0.07	0.06
- Scientist	0.06	0.06	0.06	0.12	0.09	0.10
- Founder	0.004	0.02	0.02	0.003	0.02	0.02
- Others	0.03	0.03	0.03	0.01	0.02	0.02

[18] In fact, besides the short period of observation due to inventors that patented only once, we observe Indian migrants for a shorter period of time because some of them simply started their career while at home and migrated after, a native, however has been always observed because he started his career in the country where we are investigating them.

2.3.3 Model specification

We estimate the relationship between migrant status and productivity with a random effect GLS[19] model for the two dependent variables of interest: the number of patents and the number of forward citations, as for equation (1).

$$Productivity_{it} = \beta_0 + \beta_1 MIG_i + \beta_2 CMOV_{it} + \beta_3 X_{it} + \beta_4 X_i + \epsilon_{it} \quad (1)$$

where MIG is a dummy that takes value one for migrants and zero for natives, and CMOV stands for the number of firms for which the inventor has worked for, X_{it} and X_i stand respectively for time-varying and time-invariant controls, the full list of which is provided in Table 2.3.

In a different specification (equation 2) we decompose the migration status into three entry channels in the destination country: educational (the inventor enters during his/her studies; EDUC_MIG); company change (the inventors leave a company in his/her country of origin and moves to a different one at destination; COMP_MIG) or within-company (the inventor moves to the destination country through internal mobility within the same company he/she worked for at home; WITHIN_MIG).

$$Productivity_{it} = \beta_0 + \beta_1 EDUC\_MIG_i + \beta_2 COMP\_MIG_i + \beta_3 WITHIN\_MIG_i + \beta_4 CMOV_{it} + \beta_5 X_{it} + \beta_6 X_i + \epsilon_{it} \quad (2)$$

2.3.3.1 Measuring productivity

We consider patent production as a direct measure of inventor productivity. In particular we refer to NPAT as the yearly number of patents that the inventors' files alone or with co-inventors between 1969 – 2015[20]The patent count was used in different studies investigating the inventor's performance and mobility (Hoisl, 2007; Breschi & Lissoni, 2005). We focus only on patents filed when the inventor resides in the US. That is, we do not consider any patent filed by migrants in their origin's country in the core chapter, except in the robustness checks reported in Appendix F.

When it comes to measuring patent quality (importance and value of the invention) we rely on forward citations per year of life of the patent NCIT). Gambardella et al. (2006) have shown that the number of citations received by a patent is a good proxy for the value of a patent. However, measures based only on the number of forward citations have some limitations (Hall, et al., 2001). For example, large firms might have larger portfolios of citing patents compared to smaller companies and universities, hence affecting the number of citations that their patents receive by self-citations. Furthermore, citations cannot be made to or by inventions that have not patented and so, underestimating the importance of some of them.

[19] Please note that we are using by default for OLS random effect estimation the generalized method of moments (GMM).
[20] This time window corresponds to the observed years across this chapter, hence using only patent data as observations.

2.3.3.2 Accounting for inventor characteristics

In our dataset, we observe a high heterogeneity in the observed duration of inventors' careers. Some span for many years and others were active for a short period. For this purpose, we control for the duration of an inventor' labour career (LMEX), reducing the bias due to left truncation and right truncation for those who retired in our dataset. In addition, we control for the inventor position in the company as an explanatory factor for such productivity or inactivity during an inventor's career (TITLE).

To consider, the increasing importance of highly skilled migration toward the United States in the last decades, we control, as suggested by Borjas (1985) and Hall and al. (2007), for cohort effects (COH) and age (AGE). We also control for the inventor gender (GEND). To consider the widely developed literature on the positive selection of mobile highly skilled workers, we control for individual skills, approximated by the level of education (EDLEV). Finally, we control for the patent production in the country of origin for the Indians' migrants (PAT_IND). Hence PAT_IND takes the value 0 if the inventor is a native or a migrant with no patent at home.

Table 2.3: List of variables

Variable	Definition or formula to calculate
Productivity measures	
$NPAT_{it}$	The number of patents granted by each inventor i the year t. Application date 1969-2016 (Source. EPO)
$NCIT_{it}$	The number of forwards citations received by the patent of the inventor i the year t up to 2016 (Source. EPO).
$Log(1 + NPAT_{it})$	The logarithm of one plus the number of patents granted by each inventor i the year t. Variable used with GLS specification.
$Log(1 + NCIT_{it})$	The logarithm of one plus the number of forwards citations received by the patent of the inventor i the year t. Variable used with GLS specification.
Migration status	
MIG_i	Dummy variable indicating whether inventor i is a migrant or a native.
$COMP\_MIG_i$	Dummy variable indicating whether inventor i migrated in changing of company or not.
$EDUC\_MIG_i$	Dummy variable indicating whether inventor i migrated for education reasons or not.
$WITHIN\_MIG_i$	Dummy variable indicating whether inventor i migrated within the same company or not.
Interfirm Mobility	
$CMOV_{it}$	Count variable indicating how many time the inventor i has changed of company up to year t.
Career Measures	
$TITLE_{it}$	Categorical variable (can be break into six dummy variable) giving the position in the current company of inventor i the year t. We observe six positions, such as Engineer; Manager; Founder or Co-Founder; Company's head (CEO); Scientist and Others.
Education Measures	
$EDLEV_i$	Count variables giving the education level of inventor i among three; Bachelor, Master, or Ph.D./MBA.
Other Demographical information	
COH_i	Categorical variable giving the inventor's cohort among five; starting with the inventors born before 1950 and ending for those born between 1980 – 1990.

2.3.4 Estimation and endogeneity issues

Due to the excess of zeroes and overdispersion in our dependent variables, we first apply a logarithmic transformation of the depended variables and estimate the coefficients with ordinary least squares. As a robustness check, we then estimate the coefficients using the non-transformed dependent variables with a negative binomial regression model, finding overdispersion (see LR test, Appendix A, Table 2.A.2 for the number of patents and 2.A.3 for the number of forward citations) for both the number of patents and the number of forward citations. Due to the invariant characteristic of our main variable of interest, migrant, or native (MIG), we are using random effect models. Furthermore, we consider heteroscedasticity issues by using robust standard errors.

Both our models may suffer from endogeneity issues. Migrants are positively selected, meaning that it exists a correlation between the migration's explanatory variables and the error term, driven by inventor's unobservable characteristics such as ability and intrinsic motivations, that can bias our coefficients upward. We are taking into account of this issue by decomposing our migration variable by cohort of entry in the US. In this way we aim to proxy for inventor's intrinsic motivations by considering cohorts characterized by a different degree of difficulty and waiting times for entering in the US, measuring these hampering factors with the exogenous variation in the H-1B visa program (see subsection 3.5.2 and Appendix D, Table 2.D.1 for more details on the variable building). The H-1B is a visa in the United States, under the Immigration and Nationality Act, that allows and regulates the hiring of foreign workers by US employers. This law limits the number of granted H-1B visas each year. The H-1B system attempts to protect American workers by requiring that employers pay an H-1B worker the usual wage for his position, experience, and qualifications. Furthermore, this visa system enhances the migrant selection based on information that is difficult to measure, but observable by employers that can make decisions based on a case-by-case basis (Kerr, 2018). The introduction of this program in 1990 may even have enhanced the positive selection of migrants entering the United States. Furthermore, before the H-1B visa, there was an unlimited of H-1 visas available. It is only since the H-1B that quotas for qualified workers have been installed. It is why we assume that before 1990 there was no selection between highly skilled migrants.

2.3.5 Results

2.3.5.1 Migration and entry channel: Using patent data as time frame

In this subsection, we analyze how the migrant-native difference in productivity changes when we control for several inventor's characteristics such as its previous mobility across companies, level of education, cohort, and position in the current company.

In Table 2.4, we present the estimation results on the partial correlation between the migrant status and the log number of patents (Panel A) or log number of forward citations (Panel B). In the first column, we show a specification whose only covariates are the foreign-born dummy (MIG) and the year fixed effect dummies. We can notice that migrants are more productive than natives of 8.5% in terms of the number of patent and 29.5% in terms of number of citations. In the following columns, we proceed by adding controls.

Column 2 shows that adding the position in the company has a negligible impact on the migrant advantage on productivity, which is now 8% for patents production and 28.1% for the number of citations. Moreover, the inventor's position in the company (TITLE) contributes to decreasing the productivity gap between migrants and natives.

Table 2.4: Migrants vs. Natives productivity

Panel A: Log Annual Number of Patents Log(1 + $NPAT_{it}$)

	(1) GLS	(2) GLS	(3) GLS	(4) GLS	(5) GLS
MIG	0.085*** (0.008)	0.080*** (0.008)	0.054*** (0.008)	0.055*** (0.008)	0.045*** (0.008)
CMOV				-0.015*** (0.002)	-0.017*** (0.003)
CMOV*CMOV				0.001** (0.001)	0.001*** (0.001)
EDLEV					0.034*** (0.003)
Observations	253,239	253,239	253,239	253,239	253,239
Nbr Inventors	40,806	40,806	40,806	40,806	40,806
R–squared	0.021	0.024	0.029	0.030	0.032
Year	Yes	Yes	Yes	Yes	Yes
TITLE	No	Yes	Yes	Yes	Yes
COH	No	No	Yes	Yes	Yes

Panel B: Log Annual Number of Citations Log(1 + $NCIT_{it}$)

	(1) GLS	(2) GLS	(3) GLS	(4) GLS	(5) GLS
MIG	0.295*** (0.021)	0.281*** (0.021)	0.192*** (0.021)	0.192*** (0.021)	0.163*** (0.021)
CMOV				-0.016** (0.005)	-0.020** (0.008)
CMOV*CMOV				0.003** (0.0005)	0.004*** (0.001)
EDLEV					0.083*** (0.008)
Observations	253,239	253,239	253,239	253,239	253,239
Nbr Inventors	40,806	40,806	40,806	40,806	40,806
R-squared	0.034	0.036	0.044	0.044	0.045
Year dummies	Yes	Yes	Yes	Yes	Yes
Title	No	Yes	Yes	Yes	Yes
Cohort	No	No	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1

This table estimates the effect of being a migrant on inventors' productivity with different combinations of characteristics. Standard errors appear in parenthesis and are clustered at the inventor level. Panel A shows the effect on the number of patents produced yearly per inventor. Panel B shows the effect on the total number of citations received yearly per inventor. Both Panel A and B are estimated with a random effect GLS model.

In column 3, we control for cohort effects to consider variations in the characteristics the individuals that change over time with the length of their work experience. This drastically reduces the migrant-native productivity gap from 8% to 5.4% concerning the patenting productivity and from 28.1% to 19.2% for quality.

In column 4, we add the inter-firm mobility covariate (CMOV) and its quadratic form to investigate the effect of mobility on productivity for inventors moving more than once. Differently from the previous literature (Hoisl, 2007; Hoisl, 2009; Hoisl & de Rassenfosse, 2014) we find a negative impact of mobility on the productivity proxies: -1.5% on patents and -1.6% on citations with a positive sign of quadratic effect, suggesting that inventors moving more than once recover the loss in productivity associated with the first move. This difference from the existing literature can be explained by the different measures used for mobility. Thanks to our newly developed data, we can measure inter-firm mobility before the first patent. Hence, unlike the previously mentioned empirical evidence, our data enable us to keep in the sample inventors that patented only once. Doing so, we are considering not only professional inventors but also inventors whose inventive career was occasional. We, therefore, consider individuals who may move across firms or national boundaries after one occasional patent, and not with the intent of pursuing an inventive career.

Finally, in column 5, we add the education covariate (EDLEV) to account for the difference in skills between natives and migrants. Our findings are in line with the existing literature on positive selection of migrants (Hunt, 2004), since, we control for education that reduces the migrant-native gap, keeping a migrant advantage of 4.5% for the patent production and 16.2% for the number of citations received. This result seems to be robust and even stronger when we use a Negative Binomial model (Appendix A Table 2.A.2).

Next, we decompose the migration variable (MIG) by entry channel in the destination country. Doing so, we are providing evidence on how different migration channels are related to the migrants' productivity. Table 2.5 presents the coefficients of the three migration channels. Panel A shows the results for the Log number of patents and, Panel B for the log number of forward citations. In the first column of Panel A, we only use the entry channels and year fixed effects as covariates. We observe that all three coefficients are positive and statistically significant. We observe that migrants entering in the US for education motives[21] perform better in terms of patent production with respect of the natives, while migrants with a host country education have an increase in productivity of 9.7%, migrants entering in the US when changing of company have an increase of 6.3% and when entering in US within the same company of 6.4%. These results confirm the previous ones, namely that migrants are more productive than natives, especially if they acquired the education in the host country.

We find similar results in Panel B, for patent quality. Migrants having entered in the US for education reasons have a productivity advantage over natives of 32.6%, while for migrants entering by changing company the same figure is 22.7%, and for those migrating within the same company is o26.6%. In column 2, when we include the inventor's position in the company as control variable, we observe a decrease in the impact of migration on the inventor's productivity for the Panel A and B. Adding the

[21] Hence getting one degree and entering in the host labour market

cohort effect in column 3, we find a loss in significance for both the inter and within company migration, with still migration for education reasons having the highest impact on the inventor productivity (6.3%). In Panel B, the coefficients of migration for education motives and within the same company increase and their partial correlation effect on the inventor's productivity quality is respectively of 21.2% and 21.4%. In Column 5 we add the inter-firm mobility covariate (CMOV) that strengthens the positive impact of all three migration dummies for both the Panels, with a loss in significance for the impact of migration within the same company on the number of patents filed. Finally, in column 5, when adding the education covariate (EDLEV), we find an intriguing result: we observe that migrants entering in the destination country via re-allocation within the same company seem to outperform the other migrants. This result is robust to a Negative Binomial specification, and the coefficient has an increase in the magnitude (Appendix A Table 2.A.3). A possible explanation for this result is that migrants may be positively selected when re-allocated within the same multinational firm. Nevertheless, this result must be taken with a pinch of salt, as we do not find any difference across the channels coefficients for both the GLS and Negative Binomial specifications, based on a Wald test.

Similar to the H-1B visa program, the selection is performed on a case to case basis, and consequently, firms are able to observe the characteristics of an individual during his working experience in his home country. Then, after selection at home, the best employees will be more likely to be re-located at destination in the US, where the research in ICTs is at the forefront, due to the massive expenditures in R&D, the presence of the best collaborators-colleagues and the best equipment at their disposal.

Table 2.5: Migrant channels vs. Natives productivity

Panel A: Log Annual Number of Patents Log(1 + $NPAT_{it}$)

	(1) GLS	(2) GLS	(3) GLS	(4) GLS	(5) GLS
EDUC_MIG	0.097*** (0.097)	0.091*** (0.010)	0.063*** (0.010)	0.066*** (0.010)	0.048*** (0.011)
COMP_MIG	0.063*** (0.015)	0.060*** (0.015)	0.033** (0.015)	0.036** (0.015)	0.037** (0.015)
WITHIN_MIG	0.064*** (0.021)	0.061*** (0.021)	0.049** (0.021)	0.041* (0.021)	0.032 (0.021)
CMOV				-0.015*** (0.003)	-0.017*** (0.003)
CMOV*CMOV				0.001** (0.0001)	0.001*** (0.001)
EDLEV					0.033*** (0.003)
Observations	253,239	253,239	253,239	253,239	253,239
Nbr Inventors	40,806	40,806	40,806	40,806	40,806
R-squared	0.021	0.024	0.029	0.030	0.032
Year	Yes	Yes	Yes	Yes	Yes
Title	No	Yes	Yes	Yes	Yes
Cohort	No	No	Yes	Yes	Yes

Panel B: Log Annual Number of Citations Log(1 + $NCIT_{it}$)

	(1) GLS	(2) GLS	(3) GLS	(4) GLS	(5) GLS
EDUC_MIG	0.326*** (0.026)	0.308*** (0.026)	0.212*** (0.026)	0.214*** (0.026)	0.168*** (0.027)
COMP_MIG	0.227*** (0.038)	0.222*** (0.038)	0.133*** (0.038)	0.135*** (0.038)	0.137*** (0.039)
WITHIN_MIG	0.266*** (0.057)	0.261*** (0.056)	0.214*** (0.058)	0.212*** (0.058)	0.190*** (0.057)
CMOV				-0.015* (0.008)	-0.019** (0.009)
CMOV*CMOV				0.003** (0.001)	0.004*** (0.001)
EDLEV					0.083*** (0.008)
Observations	253,239	253,239	253,239	253,239	253,239
Nbr Inventors	40,806	40,806	40,806	40,806	40,806
R-squared	0.034	0.037	0.044	0.044	0.045
Year	Yes	Yes	Yes	Yes	Yes
Title	No	Yes	Yes	Yes	Yes
Cohort	No	No	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1

This table estimates the effect of being a migrant decomposed by entry channel in the United States on inventors' productivity with different combinations of characteristics. Standard errors appear in parenthesis and are clustered at the inventor level. Panel A shows the effect on the number of patents produced yearly per inventor. Panel B shows the effect on the total number of citations received yearly per inventor. Both Panel A and B are estimated with a random effect GLS model.

2.3.5.2 Migration and entry channel: Using LinkedIn data as time frame

In this subsection, we analyze again the migrant-native difference in productivity changes. But differently from the previous section, we now observe the inventors during all their US labour market activity up to 2015, for those still active in the US at this time.

First, in table 2.6 we observe a consistency of the results for the migrant variable (MIG), which still suggests that migrants are more productive than natives at destination. The results are confirmed when we use a negative binomial specification (Appendix B, Table 2.B.2). However, we now find a lower In fact, in Table 2.6, Panel A, Column 5, we find that being a migrant increases the number of patents produced by 2.1%, instead of 4.5%. When we consider the number of citations received, Table 2.6, Panel B, Column 5, we find that being a migrant increase the number of citations received by 6.1%, instead of 16.3%. This last result shows that short time frame tends to bias upward coefficients. When we consider only the patent data time frame, we analyze the inventor productivity during his active patenting period, hence, when he is active only. While, when we consider the whole labour market experience in the US, we artificially extend the inactive patenting period of an inventor, the patenting experience being most of the time a short period in the individual's labour career.

When we consider inter-firm mobility (CMOV), we observe a different result. While, in the previous subsection we found that during the patenting period inter-firm mobility has a negative impact on the inventor's productivity; we find a different result when we consider the whole labour market experience in the US, see Table 6. Inter-Firm mobility has now a positive impact on inventors' productivity for both the number of patents filed and the number of citations received. This last finding strengthens the previous one if companies tend to retain the most prolific inventors during their patenting career, see Table 5. Hence those changing of company being less productive than the non-movers, overall changing of company has a positive impact on the inventor productivity, see Table 2.6 both Panel A and B; and Appendix B, Table 2.B.2 for the negative binomial specification. Thus, and as the previous literature on inventor's mobility has shown, inventors' movers can benefit from the new environment to learn different skills and acquire a better network, as well as improving the employee-employer match.

Table 2.6: Migrants vs. Natives productivity

Panel A: Log Annual Number of Patents Log(1 + $NPAT_{it}$)

	(1) GLS	(2) GLS	(3) GLS	(4) GLS	(5) GLS
MIG	0.035*** (0.003)	0.032*** (0.003)	0.031*** (0.003)	0.031*** (0.003)	0.021*** (0.003)
CMOV				0.003*** (0.001)	0.003*** (0.001)
CMOV*CMOV				-0.0006*** (0.001)	-0.0006*** (0.001)
EDLEV					0.027*** (0.001)
Observations	912,526	912,526	912,526	912,526	912,526
Nbr Inventors	40,806	40,806	40,806	40,806	40,806
R–squared	0.032	0.042	0.043	0.043	0.046
Year	Yes	Yes	Yes	Yes	Yes
TITLE	No	Yes	Yes	Yes	Yes
COH	No	No	Yes	Yes	Yes

Panel B: Log Annual Number of Citations Log(1 + $NCIT_{it}$)

	(1) GLS	(2) GLS	(3) GLS	(4) GLS	(5) GLS
MIG	0.086*** (0.008)	0.079*** (0.008)	0.089*** (0.008)	0.088*** (0.008)	0.061*** (0.008)
CMOV				0.020*** (0.002)	0.020*** (0.002)
CMOV*CMOV				-0.002*** (0.0003)	-0.002*** (0.0003)
EDLEV					0.072*** (0.003)
Observations	912,526	912,526	912,526	912,526	912,526
Nbr Inventors	40,806	40,806	40,806	40,806	40,806
R-squared	0.037	0.046	0.047	0.047	0.049
Year dummies	Yes	Yes	Yes	Yes	Yes
Title	No	Yes	Yes	Yes	Yes
Cohort	No	No	Yes	Yes	Yes

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

This table estimates the effect of being a migrant on inventors' productivity with different combinations of characteristics. Standard errors appear in parenthesis and are clustered at the inventor level. Panel A shows the effect on the number of patents produced yearly per inventor. Panel B shows the effect on the total number of citations received yearly per inventor. Both Panel A and B are estimated with a random effect GLS model.

Then Table 2.7, we investigate the difference in productivity by entry channel in the destination country. Using LinkedIn data as time frame strengthens the previous results. We still find migrants entering the US with a multinational the most productive, followed by those who entered for education reason. These results are confirmed when using a negative binomial specification as well Appendix B Table 2.B.3. Nevertheless, endogeneity issues driven by the positive selection of migrants remain.

Table 2.7: Migrant channels vs. Natives productivity

Panel A: Log Annual Number of Patents Log(1 + $NPAT_{it}$)

	(1) GLS	(2) GLS	(3) GLS	(4) GLS	(5) GLS
EDUC_MIG	0.043*** (0.004)	0.041*** (0.004)	0.039*** (0.004)	0.039*** (0.004)	0.024*** (0.004)
COMP_MIG	0.010** (0.005)	0.008* (0.005)	0.008 (0.005)	0.007 (0.005)	0.008* (0.005)
WITHIN_MIG	0.048*** (0.008)	0.040*** (0.008)	0.041*** (0.008)	0.040*** (0.008)	0.033*** (0.008)
CMOV				0.003*** (0.001)	0.003*** (0.001)
CMOV*CMOV				-0.0006*** (0.0001)	-0.0006*** (0.001)
EDLEV					0.026*** (0.001)
Observations	912,526	912,526	912,526	912,526	912,526
Nbr Inventors	40,806	40,806	40,806	40,806	40,806
R-squared	0.032	0.042	0.043	0.043	0.046
Year	Yes	Yes	Yes	Yes	Yes
Title	No	Yes	Yes	Yes	Yes
Cohort	No	No	Yes	Yes	Yes

Panel B: Log Annual Number of Citations Log(1 + $NCIT_{it}$)

	(1) GLS	(2) GLS	(3) GLS	(4) GLS	(5) GLS
EDUC_MIG	0.109*** (0.010)	0.103*** (0.010)	0.113*** (0.010)	0.112*** (0.010)	0.071*** (0.010)
COMP_MIG	0.017 (0.013)	0.013 (0.013)	0.024* (0.013)	0.023* (0.013)	0.026** (0.013)
WITHIN_MIG	0.121*** (0.021)	0.100*** (0.021)	0.110*** (0.021)	0.104*** (0.021)	0.087*** (0.021)
CMOV				0.020*** (0.003)	0.020*** (0.003)
CMOV*CMOV				-0.002*** (0.001)	-0.002*** (0.001)
EDLEV					0.071*** (0.003)
Observations	912,526	912,526	912,526	912,526	912,526
Nbr Inventors	40,806	40,806	40,806	40,806	40,806
R-squared	0.037	0.046	0.047	0.047	0.049
Year	Yes	Yes	Yes	Yes	Yes
Title	No	Yes	Yes	Yes	Yes
Cohort	No	No	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1

This table estimates the effect of being a migrant decomposed by entry channel in the United States on inventors' productivity with different combinations of characteristics. Standard errors appear in parenthesis and are clustered at the inventor level. Panel A shows the effect on the number of patents produced yearly per inventor. Panel B shows the effect on the total number of citations received yearly per inventor. Both Panel A and B are estimated with a random effect GLS model.

2.3.5.3 Selection and migration into the US

In this subsection, we control and further investigate the selection issue. Thanks to our data, we know the entry year of each one of our Indians migrants into the US labour market. Crossing this information and the those on the H1-B visa program, we know if, during a given year, the migrant could have been extremely selected or not. Using the number of H1-B visa available (cap) and the number of H1-B visas delivered (granted) we are building the variable (SELEC) as the number of granted visas on the available cap for each year (please refer to Appendix D, Table 2.D.1 for more details). The variable SELEC takes the value 0 if the individual is a native or a migrant that entered in the US labour market before 1990, date when the H-1B visa policy started.

From Table 2.8 we observe, when we control for selection, that the migrant variable is not positive and significant anymore and this, for both the number of patents (column (1,2,3)) and the number of citations (column (4,5,6)). Furthermore, the variable describing the selection (SELEC) is strongly significant and positive, indicating that more an inventor migrant entering the US is selected the more he is productive. Hence, confirming the previous results from the literature on the positive selection of migrant.

Table 2.8: Migrant vs. Natives productivity, Migration, and selection

	GLS (1)	GLS MLE (2)	NEG. BIN. (3)	GLS (4)	GLS MLE (5)	NEG. BIN. (6)
MIG	0.023 (0.022)	-0.012 (0.015)	0.923** (0.029)	0.069 (0.060)	0.025 (0.044)	0.933*** (0.002)
SELEC	0.019 (0.017)	0.039*** (0.012)	1.129*** (0.029)	0.089* (0.046)	0.092*** (0.035)	1.174*** (0.023)
CMOV	-0.012*** (0.002)	-0.010*** (0.001)	0.976*** (0.004)	-0.003 (0.006)	0.004 (0.005)	0.976*** (0.003)
CMOV*CMOV	0.001*** (0.001)	0.001*** (0.001)	1.002*** (0.001)	0.001 (0.001)	0.001*** (0.001)	1.001*** (0.001)
EDLEV	0.034*** (0.003)	0.022*** (0.002)	1.041*** (0.005)	0.083*** (0.008)	0.052*** (0.007)	1.071*** (0.004)
Observations	253,239	253,239	253,239	253,239	253,239	253,239
Nbr Inventors	40,806	40,806	40,806	40,806	40,806	40,806
R-squared	0.032			0.045		
Year, TITLE, COH	Yes	Yes	Yes	Yes	Yes	Yes
LR test sigma_u=0		3.0e^04***			2.1e^04***	
Wald test			4153***			7533***
LR test			3.8e^04***			1.0e^04***

*** p<0.01, ** p<0.05, * p<0.1

This table estimates the effect of being a migrant decomposed by cohort of entry in the United States, the baseline cohort being the natives, on inventors' productivity with different combinations of characteristics. Furthermore, we include the variable that captures the selection effect ($SELEC_{it}$) described table 9. Standard errors appear in parenthesis and are clustered at the inventor level. Column (1),(2) and (3) shows the effect on the number of patents produced yearly per inventor, measured as $Log(1 + NPAT_{it})$ for columns (1) and (2) and as $NPAT_{it}$ for (3). Column (4),(5) and (6) show the effect on the total number of citations received yearly per inventor, measured as $Log(1 + NCIT_{it})$ for columns (4) and (5) and as $NCIT_{it}$ for (6). All columns are estimated with a random effect model, with a maximum likelihood estimator (MLE) for column (2) and (5) as a robustness check.

To investigate further the selection issue, we are decomposing the migrant variable, based on the H1-B visa program, see Appendix D, Table 2.D.1 for more details and Table 2.9 for the results, by cohort of entry in the United States labour market. This new categorical variable COH_MIG takes the value of 0 if the inventor is a native and is issued as our baseline comparison. If the inventor migrant entered in the US before the H1-B visa implementation COH_MIG takes 1 (referred as COH_MIG_1 in Table 2.9), if she enters in the US between 1990 and 1998, so during the first cap period, the variable COH_MIG takes 2 (referred as COH_MIG_2 in Table 2.9). Thus, when the H1-B visa cap is raised twice between 1999 and 2003 the variable COH_MIG takes 3 (referred as COH_MIG_3 in Table 2.9), finally, when the H1-B visa policy is reduced and keep the same cap until the end of our observations time, between 2004 and 2015, the variable COH_MIG takes the value of 4 (referred as COH_MIG_4 in Table 2.9).

We observe that during the first period (COH_MIG_1) a migrant that entered the US before the H1-B visa policy is not more productive than a native and even less productive if we consider the negative binomial specification column 3 and 4, both for the number of patents (column 1 to 3) and the number of citations (column 4 to 6). When we consider the second period (COH_MIG_2), when the H1-B visa policy has been installed, we observe that the migrants are now positively and significatively more productive than natives. After that, when the cap of H1-B visa has been raised (COH_MIG_3), making more accessible for a migrant to obtain one, we do not observe any significant increase or decrease in the productivity for the migrants entering the US in that period. At last, when the H1-B visa cap has been lowered (COH_MIG_4), hence, increasing the difficulty to enter the US, we observe that it has a positive and significant impact on the migrant's productivity.

Table 2.9: Migrant vs. Natives productivity, Migration cohort based on cap

	GLS (1)	GLS MLE (2)	NEG. BIN. (3)	GLS (4)	GLS MLE (5)	NEG. BIN. (6)
COH_MIG_1	0.018 (0.039)	-0.023 (0.007)	0.895** (0.050)	-0.050 (0.11)	-0.167** (0.077)	0.883*** (0.034)
COH_MIG_2	0.040** (0.018)	0.013 (0.012)	0.970 (0.025)	0.166*** (0.046)	0.092*** (0.021)	1.023 (0.020)
COH_MIG_3	0.025* (0.014)	0.012 (0.010)	1.028 (0.022)	0.155*** (0.035)	0.136*** (0.029)	1.077*** (0.019)
COH_MIG_4	0.069*** (0.011)	0.062*** (0.010)	1.152*** (0.005)	0.220*** (0.025)	0.183*** (0.007)	1.255*** (0.019)
CMOV	-0.017*** (0.003)	-0.022*** (0.002)	0.951*** (0.005)	-0.019** (0.009)	-0.035*** (0.006)	0.957*** (0.004)
CMOV*CMOV	0.001*** (0.001)	0.003*** (0.001)	1.007*** (0.001)	0.004*** (0.001)	0.009*** (0.001)	1.006*** (0.001)
EDLEV	0.034*** (0.003)	0.023*** (0.002)	1.041*** (0.005)	0.084*** (0.008)	0.053*** (0.007)	1.062*** (0.004)
Observations	253,239	253,239	253,239	253,239	253,239	253,239
Nbr Inventors	40,806	40,806	40,806	40,806	40,806	40,806
R-squared	0.032			0.045		
Year, TITLE, COH	Yes	Yes	Yes	Yes	Yes	Yes
LR test sigma_u=0		3.0e^04***			2.1e^04***	
Wald test			4220***			1.0e^4***
LR test			3.8e^4***			7073***

*** p<0.01, ** p<0.05, * p<0.1

This table estimates the effect of being a migrant decomposed by cohort of entry in the United States, the baseline cohort being the natives, on inventors' productivity with different combinations of characteristics. Standard errors appear in parenthesis and are clustered at the inventor level. Column (1),(2) and (3) shows the effect on the number of patents produced yearly per inventor, measured as $Log(1 + NPAT_{it})$ for columns (1) and (2) and as $NPAT_{it}$ for (3). Column (4),(5) and (6) show the effect on the total number of citations received yearly per inventor, measured as $Log(1 + NCIT_{it})$ for columns (4) and (5) and as $NCIT_{it}$ for (6). All columns are estimated with a random effect model, with a maximum likelihood estimator (MLE) for column (2) and (5) as a robustness check.

Also, as a robustness check, we control for the selection (SELEC) within the categories considering not only the cap but also the number of granted H1-B visa (for more details, please refer to the appendix D Table 2.D.2). That check confirms and strengthens the previous findings of Table 2.9. Finally, Table 2.10, we are basing the cohort of entry not based on the number of caps, but on the selection variable (SELEC). This new variable (COH_SELEC) takes the value 0 if the individual is a native, the value of 1 when the individual is a migrant entering in US before 1990 (referred as COH_SELEC_1 in the Table 2.10), the value 2 when the variable SELEC is under 1 (referred as COH_SELEC_2 in Table 2.10), indicating that there are less granted visas than the maximum limit, then the value 3 when this SELEC is above 1 (referred to COH_SELEC_3 in Table 2.10), the value 4 when the variable SELEC is later under 1 (referred to COH_SELEC_4 in Table 2.10), and finally the variable COH_SELEC takes the value 5 when the variable SELEC is again above 1 (referred to COH_SELEC_5 in Table 2.10; see Appendix D, Table 2.D.1 for more details).

Table 2.10: Migrant vs. Natives productivity, Migration cohort based on selection

	GLS (1)	GLS MLE (2)	NEG. BIN. (3)	GLS (4)	GLS MLE (5)	NEG. BIN. (6)
COH_SELEC_1	0.0231 (0.039)	-0.018 (0.027)	0.904* (0.050)	-0.034 (0.10)	-0.156** (0.077)	0.898*** (0.035)
COH_SELEC_2	0.047* (0.024)	0.014 (0.016)	0.947 (0.032)	0.165*** (0.109)	0.08* (0.046)	0.991 (0.024)
COH_SELEC_3	0.014 (0.015)	0.001 (0.010)	0.996 (0.023)	0.127*** (0.038)	0.109*** (0.031)	1.02 (0.018)
COH_SELEC_4	0.025 (0.019)	0.014 (0.041)	1.041 (0.035)	0.147*** (0.048)	0.143*** (0.045)	1.122*** (0.03)
COH_SELEC_5	0.058*** (0.010)	0.049*** (0.008)	1.121*** (0.020)	0.166*** (0.025)	0.130*** (0.025)	1.197*** (0.019)
CMOV	-0.011*** (0.002)	-0.009*** (0.002)	0.979*** (0.004)	-0.012 (0.009)	-0.027*** (0.006)	0.965*** (0.004)
CMOV*CMOV	0.001*** (0.001)	0.001*** (0.001)	1.002*** (0.001)	0.004*** (0.001)	0.009*** (0.001)	1.006*** (0.001)
EDLEV	0.033*** (0.003)	0.022*** (0.002)	1.041*** (0.005)	0.084*** (0.008)	0.052*** (0.007)	1.063*** (0.004)
Observations	253,239	253,239	253,239	253,239	253,239	253,239
Nbr Inventors	40,806	40,806	40,806	40,806	40,806	40,806
R-squared	0.033			0.047		
Year, TITLE, COH	Yes	Yes	Yes	Yes	Yes	Yes
LR test sigma_u=0		3.0e^04***			2.0e^04***	
Wald test			4354***			1.0e^4***
LR test			3.8e^4***			7840***

*** p<0.01, ** p<0.05, * p<0.1

This table estimates the effect of being a migrant decomposed by cohort of entry in the United States, the baseline cohort being the natives, on inventors' productivity with different combinations of characteristics. Standard errors appear in parenthesis and are clustered at the inventor level. Column (1),(2) and (3) shows the effect on the number of patents produced yearly per inventor, measured as $\text{Log}(1 + NPAT_{it})$ for columns (1) and (2) and as $NPAT_{it}$ for (3). Column (4),(5) and (6) show the effect on the total number of citations received yearly per inventor, measured as $\text{Log}(1 + NCIT_{it})$ for columns (4) and (5) and as $NCIT_{it}$ for (6). All columns are estimated with a random effect model, with a maximum likelihood estimator (MLE) for column (2) and (5) as a robustness check.

The choice to base the cohort of entry on the selection variable (SELEC) gives a double insight. First, on the building strategy of the selection variable, we should expect a stronger and positive effect between cohorts where the number of caps was lower than the number of granted visas, hence when the variable SELEC takes a value higher than 1 (COH_SELEC_3 and COH_SELEC_5). Second, considering the dispersion of the selection variable (SELEC) Appendix D, Table 2.D.2, we observe a much higher selection in the recent years, for the last cohort (COH_SELEC_5), hence we are expecting a large and significant effect for the migrants' productivity of this last cohort compared to the natives and also among the other cohorts.

The results in Table 2.10 give similar insight as the previous table, Table 2.9. The first cohort (COH_SELEC_1) that entered the US before the H1-B visa implementation is not more productive than the natives for the number of patents and, even less productive than the natives when we account for the number of citations. Thus, when the H1-B visa policy started (COH_SELEC_2) but still with few demands for it, we observe a slickly positive effect for both numbers of patents and number of citations, nevertheless not robust. After that in a second period the selection variable is above 1, hence the demand for the number of H1-B visa exceeding the offer. For this cohort (COH_SELEC_3) we observe similar results as the previous cohort, with nonetheless a stronger positive effect on the number of citations received by the migrants entering the US during this period, compared to the natives. We do not observe a real impact on the productivity of the migrants entering the US during the third cohort due to the increase of the number of H1-B visas available in this period. We believe that the third cohort captures both the increase in the demand of H1-B visa and the relaxation of the H1-B visa policy, hence increasing the offer of H1-B visas. After that, the US government increased a second time the number of H1-B visas available, relaxing, even more, the selection of the cohort entering the US (COH_SELEC_4). We, nevertheless, observe a positive and significant effect on the number of citations received by the inventors entering the US in this cohort compared to the natives. Finally, when the number of visas has been reduced while the demand of such visas was strong (COH_SELEC_5), we observe that the impact of this policy has considerably increased the value of the selection variable (SELEC) being in average around 1.6 during the fifth cohort. We are hence expecting the inventors entering the US being highly selected compared to the previous cohorts. This last expectation is confirmed by the results in Table 2.10. We observe that this migrants' inventors are strongly more productive in both the number of patents and citations than the natives and the other cohorts.

2.4 Robustness Checks

To assess the reliability of our results, we carry out a robustness check.
We right-truncate each inventor's time-data points after the last patent filed. That is, we limit our observation of inventors from their first labour market experience in the US to their last patent filed in the US (see Appendix C). =By doing so, we homogenize the different inventors' career by disregarding the amount of time spent in retirement.

With this window of observation, we observe a negative effect of being a migrant on individual productivity, see Appendix C, Tables 2.C.1, and 2.C.2 column 1. However, when we control for the number of years spent in the US labour market (US_EXP), column 2, we find again that migrants are more productive than natives. This sign change can be explained by the time spent into the US labour market that, by construction, is higher for natives = (see Table 2.C.1).

When we break down the migrant variable by the entry channel in Table 2.C.3, we do not observe anymore that migrants entering the US in changing of a company are more productive than the natives. Besides, we do not strongly see, as in the previous robustness check or the sections 3.5.1 and 3.5.2, that migrants entering the US through the same company are more productive than natives.

2.5 Conclusion

Migrants have higher productivity than natives both in quantity, by producing more patents, and in quality, by receiving more citations, even when they occupy similar positions within their respective companies in computer and engineering enterprises based in the United States. This productivity gap between natives and migrants is partly due to migrants' superior education. Unlike the previous empirical literature findings, we observe that inter-firm mobility performed by both the natives and the migrants have a negative impact on their productivity. This different result can be explained by the unique features of our dataset and also have a theoretical grounding. While the previous literature on mobility and productivity of inventors were considering the individuals' mobility only between the first and last patent applications, we can extend this window of observation by observing inventors' inter-firm mobility from the beginning of their career. Furthermore, this result is coherent with other theoretical findings describing that companies tend to retain the best employees, hence the inventors leaving their company may not be as skilled as the stayers. An ongoing selection occurs since an individual is hired: the company can recognize the prior unobservable characteristics of their new employees and make decisions of his importance for the company, in promoting him and/or reallocating him. We find the first evidence of this process on the migrant's population by decomposing them with respect to their entry channel. Indeed, we have shown that migrants are, on average, more productive than natives, while there is heterogeneity in productivity, depending on the migrant's entry channel. We observe that migrants entering in the US through the own multinational's branches are performing better in quantity and quality than migrants entering in the US for education reasons or when changing of company[22]. This finding is another proof on the ongoing selection mechanism occurring within the company, where in that case, those who will be reallocated in another country will be the ones positively selected. However, when we control for selection via the H1-B visas policy, we do not find any productivity difference between migrants and natives. Going further in breaking down the migrant cohort of entry by the several variations in the number of visas available, we found that the most selected cohorts of migrants were performing better than both the natives and the less selected ones.

[22] Nevertheless, this result needs to be strengthened, we do not find any statistical difference across the channels of entry coefficients when we perform a Wald test for both the GLS and Negative binomial specifications.

2.6 Appendix

2.6.1 Appendix A: observations after the first and before the last patent filed in the US

Table 2.A.1: Sample descriptive statistics

	Natives	Migrants			
		Education	Changing Comp.	Within Comp.	Total
Nbr. years in U.S (after educ.)	5.36	4.51	3.5	3.16	3.8
Nbr. company move	1.33	1.40	1.46	0.74	1.35
Average year of birth	1965	1971	1971	1970	1971
Average U.S labour entry year	1999	2003	2005	2005	2004
Average selection effect	0	1.23	1.26	1.13	1.23
Company stayers	34.5%	24.8%	23%	52%	28.9%

Table 2.A.2: Migrants vs. Natives productivity, Robustness Check

Panel A: Annual Number of Patents ($NPAT_{it}$)

	(1) NEG. BIN.	(2) NEG. BIN.	(3) NEG. BIN.	(4) NEG. BIN.	(5) NEG. BIN.
MIG	1.178*** (0.014)	1.166*** (0.014)	1.076*** (0.013)	1.079*** (0.013)	1.064*** (0.013)
CMOV				0.953*** (0.005)	0.951*** (0.005)
CMOV*CMOV				1.007*** (0.001)	1.007*** (0.0001)
EDLEV					1.041*** (0.005)
Observations	253,239	253,239	253,239	253,239	253,239
Nbr Inventors	40,806	40,806	40,806	40,806	40,806
LR test	3.9e^4***	3.9e^4***	3.9e^4***	3.8e^4***	3.8e^4***
Wald test	2588***	2894***	4022***	4102***	4172
Year	Yes	Yes	Yes	Yes	Yes
TITLE	No	Yes	Yes	Yes	Yes
COH	No	No	Yes	Yes	Yes

Panel B: Annual Number of Citations ($NCIT_{it}$)

	(1) NEG. BIN.	(2) NEG. BIN.	(3) NEG. BIN.	(4) NEG. BIN.	(5) NEG. BIN.
MIG	1.267*** (0.012)	1.255*** (0.012)	1.132*** (0.011)	1.136*** (0.011)	1.11*** (0.011)
CMOV				0.960*** (0.004)	0.957*** (0.004)
CMOV*CMOV				1.005*** (0.001)	1.006*** (0.001)
EDLEV					1.062*** (0.004)
Observations	253,239	253,239	253,239	253,239	253,239
Nbr Inventors	40,806	40,806	40,806	40,806	40,806
LR test	9877***	9866***	1.0e^4***	1.0e^4***	1.0e^4***
Wald test	2871***	3436***	6595***	6694***	6924***
Year dummies	Yes	Yes	Yes	Yes	Yes
Title	No	Yes	Yes	Yes	Yes
Cohort	No	No	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1

This table estimates the effect of being a migrant on inventors' productivity with different combinations of characteristics. Panel A shows the effect on the number of patents produced yearly per inventor. Panel B shows the effect on the total number of citations received yearly per inventor. Both Panel A and B are estimated with a random effect negative binomial model. The reported coefficients are the incidence ratio rate.

Table 2.A.3: Migrant channels vs. Natives productivity, Robustness Check

Panel A: Annual Number of Patents ($NPAT_{it}$)

	(1) NEG. BIN.	(2) NEG. BIN.	(3) NEG. BIN.	(4) NEG. BIN.	(5) NEG. BIN.
EDUC_MIG	1.187*** (0.018)	1.174*** (0.018)	1.078*** (0.016)	1.083*** (0.016)	1.060*** (0.016)
COMP_MIG	1.156*** (0.026)	1.147*** (0.026)	1.054*** (0.024)	1.062*** (0.024)	1.064*** (0.024)
WITHIN_MIG	1.179*** (0.039)	1.165*** (0.039)	1.112*** (0.037)	1.094*** (0.036)	1.082** (0.036)
CMOV				0.953*** (0.005)	0.951*** (0.005)
CMOV*CMOV				1.007*** (0.001)	1.007*** (0.001)
EDLEV					1.041*** (0.005)
Observations	253,239	253,239	253,239	253,239	253,239
Nbr Inventors	40,806	40,806	40,806	40,806	40,806
LR test	3.9e^4***	3.9e^4***	3.8e^4***	3.8e^4***	3.8e^4***
Wald test	2589***	2895***	4024***	4103***	4172***
Year	Yes	Yes	Yes	Yes	Yes
Title	No	Yes	Yes	Yes	Yes
Cohort	No	No	Yes	Yes	Yes

Panel B: Annual Number of Citations ($NCIT_{it}$)

	(1) NEG. BIN.	(2) NEG. BIN.	(3) NEG. BIN.	(4) NEG. BIN.	(5) NEG. BIN.
EDUC_MIG	1.283*** (0.015)	1.268*** (0.015)	1.138*** (0.014)	1.143*** (0.014)	1.107*** (0.014)
COMP_MIG	1.218*** (0.022)	1.212*** (0.022)	1.088*** (0.020)	1.093*** (0.020)	1.095*** (0.020)
WITHIN_MIG	1.288*** (0.036)	1.280*** (0.035)	1.211*** (0.033)	1.196*** (0.033)	1.177*** (0.033)
CMOV				0.960*** (0.004)	0.957*** (0.004)
CMOV*CMOV				1.005*** (0.001)	1.006*** (0.001)
EDLEV					1.062*** (0.004)
Observations	253,239	253,239	253,239	253,239	253,239
Nbr Inventors	40,806	40,806	40,806	40,806	40,806
LR test	9874***	9864***	1.0e^4***	1.0e^4***	1.0e^4***
Wald test	2878***	3443***	6605***	6701***	6927***
Year	Yes	Yes	Yes	Yes	Yes
Title	No	Yes	Yes	Yes	Yes
Cohort	No	No	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1

This table estimates the effect of being a migrant decomposed by entry channel in the United States on inventors' productivity with different combinations of characteristics. Panel A shows the effect on the number of patents produced yearly per inventor. Panel B shows the effect on the total number of citations received yearly per inventor. Both Panel A and B are estimated with a random effect negative binomial model. The reported coefficients are the incidence ratio rate.

Table 2.A.4: Migrants vs. Natives productivity, Mixed-effects

First stage: Main results mixed-effects models

Dep. var.	Log(1 + NPATit)		Log(1 + NCITit)	
	(1)	(2)	(3)	(4)
MIG	0.033*** (0.006)		0.133*** (0.017)	
EDUC_MIG		0.035*** (0.007)		0.135*** (0.021)
COMP_MIG		0.033*** (0.011)		0.119*** (0.031)
WITHIN_MIG		0.025 (0.015)		0.159*** (0.045)
CMOV	-0.022*** (0.002)	-0.022*** (0.002)	-0.035*** (0.007)	-0.035*** (0.006)
CMOV*CMOV	0.003*** (0.0004)	0.003*** (0.0004)	0.009*** (0.009)	0.009*** (0.001)
EDLEV	0.022*** (0.002)	0.022*** (0.002)	0.053*** (0.006)	0.053*** (0.001)
Observations	253,239	253,239	253,239	253,239
Nbr Inventors	40,806	40,806	40,806	40,806
Year	Yes	Yes	Yes	Yes
Title	Yes	Yes	Yes	Yes
Cohort	Yes	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1

This table estimates the effect of being a migrant on inventors' productivity with individual characteristics. Standard errors appear in parenthesis and are clustered at the inventor level. We show the effect on the number of patents produced yearly per inventor (1) and (2) and the effect on the total number of citations received yearly per inventor (3) and (4). Both regressions are estimated with a mixed-effects model.

2.6.2 Appendix B: observations before and after the first patent filed in the US

Table 2.B.1: Sample descriptive statistics

	Natives	Migrants			
		Education	Changing Comp.	Within Comp.	Total
Nbr. years in U.S (after educ.)	22.35	15.66	14.81	10.98	14.81
Nbr. company move	3.03	2.96	2.75	2.72	2.88
Average year of birth	1965	1971	1971	1970	1971
Average U.S labour entry year	1990	1997	1997	2001	1996
Average selection effect	0	1.23	1.26	1.13	1.23
Company stayers	18.8%	18.8%	17.8%	1%	16.7%

Table 2.B.2: Migrants vs. Natives productivity, Robustness Check

Panel A: Annual Number of Patents ($NPAT_{it}$)

	(1) NEG. BIN.	(2) NEG. BIN.	(3) NEG. BIN.	(4) NEG. BIN.	(5) NEG. BIN.
MIG	1.210*** (0.017)	1.197*** (0.017)	1.204*** (0.017)	1.199*** (0.017)	1.116*** (0.016)
CMOV				1.034*** (0.005)	1.032*** (0.005)
CMOV*CMOV				0.988*** (0.001)	0.988*** (0.0001)
EDLEV					1.213*** (0.001)
Observations	912,526	912,526	912,526	912,526	912,526
Nbr Inventors	40,806	40,806	40,806	40,806	40,806
LR test	6.0e^4***	5.8e^4***	5.8e^4***	5.8e^4***	5.7e^4***
Wald test	22006***	26541***	26969***	27261***	28371
Year	Yes	Yes	Yes	Yes	Yes
TITLE	No	Yes	Yes	Yes	Yes
COH	No	No	Yes	Yes	Yes

Panel B: Annual Number of Citations ($NCIT_{it}$)

	(1) NEG. BIN.	(2) NEG. BIN.	(3) NEG. BIN.	(4) NEG. BIN.	(5) NEG. BIN.
MIG	1.262*** (0.013)	1.243*** (0.013)	1.213*** (0.013)	1.204*** (0.012)	1.114*** (0.011)
CMOV				1.270*** (0.003)	1.220*** (0.004)
CMOV*CMOV				0.970*** (0.001)	0.978*** (0.001)
EDLEV					1.062*** (0.004)
Observations	912,526	912,526	912,526	912,526	912,526
Nbr Inventors	40,806	40,806	40,806	40,806	40,806
LR test	8727***	8619***	1.0e^4***	2.3e^4***	1.0e^4***
Wald test	28457***	29245***	52837***	172638***	6924***
Year dummies	Yes	Yes	Yes	Yes	Yes
Title	No	Yes	Yes	Yes	Yes
Cohort	No	No	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1

This table estimates the effect of being a migrant on inventors' productivity with different combinations of characteristics. Panel A shows the effect on the number of patents produced yearly per inventor. Panel B shows the effect on the total number of citations received yearly per inventor. Both Panel A and B are estimated with a random effect negative binomial model. The reported coefficients are the incidence ratio rate.

Table 2.B.3: Migrant channels vs. Natives productivity, Robustness Check

Panel A: Annual Number of Patents ($NPAT_{it}$)

	(1) NEG. BIN.	(2) NEG. BIN.	(3) NEG. BIN.	(4) NEG. BIN.	(5) NEG. BIN.
EDUC_MIG	1.246*** (0.022)	1.243*** (0.022)	1.247*** (0.022)	1.245*** (0.022)	1.118*** (0.020)
COMP_MIG	1.024 (0.028)	1.018 (0.028)	1.022 (0.028)	1.019 (0.028)	1.031 (0.028)
WITHIN_MIG	1.437*** (0.056)	1.351*** (0.052)	1.377*** (0.053)	1.355*** (0.052)	1.294*** (0.036)
CMOV				1.033*** (0.005)	1.031*** (0.005)
CMOV*CMOV				0.988*** (0.001)	0.989*** (0.001)
EDLEV					1.212*** (0.001)
Observations	912,526	912,526	912,526	912,526	912,526
Nbr Inventors	40,806	40,806	40,806	40,806	40,806
LR test	6.0e^4***	5.8e^4***	5.8e^4***	5.8e^4***	5.7e^4***
Wald test	22007***	26396***	26823***	27114***	28200***
Year	Yes	Yes	Yes	Yes	Yes
Title	No	Yes	Yes	Yes	Yes
Cohort	No	No	Yes	Yes	Yes

Panel B: Annual Number of Citations ($NCIT_{it}$)

	(1) NEG. BIN.	(2) NEG. BIN.	(3) NEG. BIN.	(4) NEG. BIN.	(5) NEG. BIN.
EDUC_MIG	1.319*** (0.016)	1.302*** (0.015)	1.138*** (0.014)	1.143*** (0.014)	1.107*** (0.014)
COMP_MIG	1.060*** (0.021)	1.053** (0.022)	1.088*** (0.020)	1.093*** (0.020)	1.095*** (0.020)
WITHIN_MIG	1.450*** (0.041)	1.370*** (0.035)	1.211*** (0.033)	1.196*** (0.033)	1.177*** (0.033)
CMOV				1.270*** (0.003)	1.220*** (0.004)
CMOV*CMOV				0.970*** (0.001)	0.978*** (0.001)
EDLEV					1.062*** (0.004)
Observations	912,526	912,526	912,526	912,526	912,526
Nbr Inventors	40,806	40,806	40,806	40,806	40,806
LR test	8716***	8608***	1.0e^4***	1.0e^4***	1.0e^4***
Wald test	21373***	29375***	6605***	6701***	6927***
Year	Yes	Yes	Yes	Yes	Yes
Title	No	Yes	Yes	Yes	Yes
Cohort	No	No	Yes	Yes	Yes

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

This table estimates the effect of being a migrant decomposed by entry channel in the United States on inventors' productivity with different combinations of characteristics. Panel A shows the effect on the number of patents produced yearly per inventor. Panel B shows the effect on the total number of citations received yearly per inventor. Both Panel A and B are estimated with a random effect negative binomial model. The reported coefficients are the incidence ratio rate.

2.6.3 Appendix C: observations before first patent and not after last patent filed in the US

Table 2.C.1: Sample descriptive statistics

	Natives	Migrants			
		Education	Changing Comp.	Within Comp.	Total
Nbr. years in U.S (after educ.)	13.6	6.67	9.33	6.36	7.62
Nbr. company move	1.66	1.64	1.65	1.99	1.70
Average year of birth	1965	1971	1971	1970	1971
Average U.S labour entry year	1990	1997	1997	2001	1998
Average selection effect	0	1.23	1.26	1.13	1.23
Company stayers	35.8%	31.5%	32.5%	1%	27.6%

Table 2.C.2: Migrants vs. Natives productivity

Panel A: Log Annual Number of Patents $Log(1 + NPAT_{it})$

	(1) GLS	(2) GLS	(3) GLS	(4) GLS	(5) GLS	(6) GLS
MIG	-0.022*** (0.004)	0.024*** (0.004)	0.024*** (0.004)	0.025*** (0.004)	0.025*** (0.004)	0.013*** (0.004)
CMOV					0.003*** (0.002)	0.003*** (0.002)
CMOV*CMOV					-0.002*** (0.001)	-0.002*** (0.001)
EDLEV						0.032*** (0.001)
US_EXP		0.007*** (0.001)	0.007*** (0.001)	0.006*** (0.001)	0.007*** (0.001)	0.007*** (0.001)
Observations	570,019	570,019	570,019	570,019	570,019	570,019
Nbr Inventors	40,806	40,806	40,806	40,806	40,806	40,806
R–squared	0.116	0.127	0.133	0.134	0.135	0.138
Year	Yes	Yes	Yes	Yes	Yes	Yes
TITLE	No	No	Yes	Yes	Yes	Yes
COH	No	No	No	Yes	Yes	Yes

Panel B: Log Annual Number of Citations $Log(1 + NCIT_{it})$

	(1) GLS	(2) GLS	(3) GLS	(4) GLS	(5) GLS	(6) GLS
MIG	-0.054*** (0.012)	0.095*** (0.011)	0.094*** (0.011)	0.096*** (0.011)	0.092*** (0.011)	0.057*** (0.011)
CMOV					0.032*** (0.004)	0.032*** (0.004)
CMOV*CMOV					-0.006*** (0.001)	-0.006*** (0.001)
EDLEV						0.096*** (0.004)
US_EXP		0.024*** (0.001)	0.023*** (0.001)	0.019*** (0.001)	0.018*** (0.001)	0.018*** (0.001)
Observations	570,019	570,019	570,019	570,019	570,019	570,019
Nbr Inventors	40,806	40,806	40,806	40,806	40,806	40,806
R-squared	0.048	0.061	0.068	0.068	0.069	0.072
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Title	No	No	Yes	Yes	Yes	Yes
Cohort	No	No	No	Yes	Yes	Yes

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

This table estimates the effect of being a migrant on inventors' productivity with different combinations of characteristics. Standard errors appear in parenthesis and are clustered at the inventor level. Panel A shows the effect on the number of patents produced yearly per inventor. Panel B shows the effect on the total number of citations received yearly per inventor. Both Panel A and B are estimated with a random effect GLS model.

Table 2.C.3 Migrant channels vs. Natives productivity

Panel A: Log Annual Number of Patents Log(1 + $NPAT_{it}$)

	(1) GLS	(2) GLS	(3) GLS	(4) GLS	(5) GLS	(6) GLS
EDUC_MIG	-0.011**	0.033***	0.033***	0.036***	0.036***	0.018***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
COMP_MIG	-0.043***	-0.001	-0.001	0.001	0.001	0.002
	(0.007)	(0.007)	(0.007)	(0.007)	(0.001)	(0.007)
WITHIN_MIG	-0.039***	0.032***	0.027**	0.019*	0.020**	0.014
	(0.011)	(0.011)	(0.011)	(0.010)	(0.010)	(0.010)
CMOV					0.002	0.002
					(0.001)	(0.001)
CMOV*CMOV					-0.002***	-0.002***
					(0.0002)	(0.002)
EDLEV						0.031***
						(0.001)
US_EXP		0.007***	0.007***	0.006***	0.007***	0.007***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	570,019	570,019	570,019	570,019	570,019	570,019
Nbr Inventors	40,806	40,806	40,806	40,806	40,806	40,806
R-squared	0.116	0.128	0.133	0.134	0.135	0.138
Year	Yes	Yes	Yes	Yes	Yes	Yes
Title	No	No	Yes	Yes	Yes	Yes
Cohort	No	No	No	Yes	Yes	Yes

Panel B: Log Annual Number of Citations Log(1 + $NCIT_{it}$)

	(1) GLS	(2) GLS	(3) GLS	(4) GLS	(5) GLS	(6) GLS
EDUC_MIG	-0.017	0.122***	0.123***	0.129***	0.128***	0.073***
	(0.015)	(0.015)	(0.014)	(0.014)	(0.014)	(0.015)
COMP_MIG	-0.133***	0.002	0.003	0.007	0.005	0.010
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.013)
WITHIN_MIG	-0.073**	0.154***	0.139***	0.112***	0.097***	0.077**
	(0.031)	(0.031)	(0.030)	(0.030)	(0.030)	(0.030)
CMOV					0.032***	0.031***
					(0.004)	(0.005)
CMOV*CMOV					-0.006***	-0.006***
					(0.001)	(0.001)
EDLEV						0.095***
						(0.004)
US_EXP		0.024***	0.023***	0.018***	0.019***	0.019***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	570,019	570,019	570,019	570,019	570,019	570,019
Nbr Inventors	40,806	40,806	40,806	40,806	40,806	40,806
R-squared	0.049	0.061	0.068	0.069	0.069	0.072
Year	Yes	Yes	Yes	Yes	Yes	Yes
Title	No	No	Yes	Yes	Yes	Yes
Cohort	No	No	No	Yes	Yes	Yes

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

This table estimates the effect of being a migrant decomposed by entry channel in the United States on inventors' productivity with different combinations of characteristics. Standard errors appear in parenthesis and are clustered at the inventor level. Panel A shows the effect on the number of patents produced yearly per inventor. Panel B shows the effect on the total number of citations received yearly per inventor. Both Panel A and B are estimated with a random effect GLS model.

2.6.4 Appendix D: with the cohort of entry and selection added

Table 2.D.1: Addressing endogeneity, variable building

<table>
<tr><td colspan="3">H-1B Historical data[23]</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Years</td><td>Granted</td><td>Cap</td><td>Selection indicator (Granted/Cap) SELEC</td><td>Nbr Indians migrant entering</td><td>Entry cohort in U.S labour market COH_MIG</td><td>Entry cohort in U.S labour market COH_MIG2</td></tr>
<tr><td>Before 1990</td><td>-</td><td>-</td><td>0</td><td>94</td><td>cohot1</td><td>cohort1</td></tr>
<tr><td>1990</td><td>794</td><td rowspan="9">65000</td><td>0.012215</td><td>27</td><td rowspan="9">cohort2</td><td rowspan="7">cohort2</td></tr>
<tr><td>1991</td><td>51882</td><td>0.798184</td><td>28</td></tr>
<tr><td>1992</td><td>44290</td><td>0.681384</td><td>35</td></tr>
<tr><td>1993</td><td>35818</td><td>0.551046</td><td>36</td></tr>
<tr><td>1994</td><td>42843</td><td>0.659123</td><td>58</td></tr>
<tr><td>1995</td><td>51832</td><td>0.797415</td><td>92</td></tr>
<tr><td>1996</td><td>58327</td><td>0.897338</td><td>130</td></tr>
<tr><td>1997</td><td>80547</td><td>1.239185</td><td>161</td><td rowspan="4">cohort3</td></tr>
<tr><td>1998</td><td>91360</td><td>1.405538</td><td>167</td></tr>
<tr><td>1999</td><td>116513</td><td rowspan="2">115000</td><td>1.013156</td><td>201</td><td rowspan="5">cohort3</td></tr>
<tr><td>2000</td><td>133290</td><td>1.159043</td><td>253</td></tr>
<tr><td>2001</td><td>161643</td><td rowspan="3">195000</td><td>0.828938</td><td>296</td><td rowspan="3">cohort4</td></tr>
<tr><td>2002</td><td>118352</td><td>0.606933</td><td>266</td></tr>
<tr><td>2003</td><td>107196</td><td>0.549723</td><td>272</td></tr>
<tr><td>2004</td><td>138965</td><td rowspan="13">85000</td><td>1.634882</td><td>303</td><td rowspan="12">cohort4</td><td rowspan="12">cohort5</td></tr>
<tr><td>2005</td><td>124099</td><td>1.459988</td><td>332</td></tr>
<tr><td>2006</td><td>135421</td><td>1.593188</td><td>286</td></tr>
<tr><td>2007</td><td>154053</td><td>1.812388</td><td>316</td></tr>
<tr><td>2008</td><td>129464</td><td>1.523105</td><td>343</td></tr>
<tr><td>2009</td><td>110367</td><td>1.298435</td><td>266</td></tr>
<tr><td>2010</td><td>117409</td><td>1.381282</td><td>301</td></tr>
<tr><td>2011</td><td>129134</td><td>1.519223</td><td>277</td></tr>
<tr><td>2012</td><td>135530</td><td>1.594470</td><td>174</td></tr>
<tr><td>2013</td><td>153223</td><td>1.802623</td><td>65</td></tr>
<tr><td>2014</td><td>161369</td><td>1.898458</td><td>16</td></tr>
<tr><td>2015</td><td>172748</td><td>2.032329</td><td>1</td></tr>
<tr><td>2016</td><td>180057</td><td>2.118317</td><td>0</td><td></td><td></td></tr>
</table>

[23] https://www.uscis.gov/tools/reports-studies/reports-and-studies

Table 2.D.2: Migrant channels vs. Natives productivity, Migration cohort + Selection

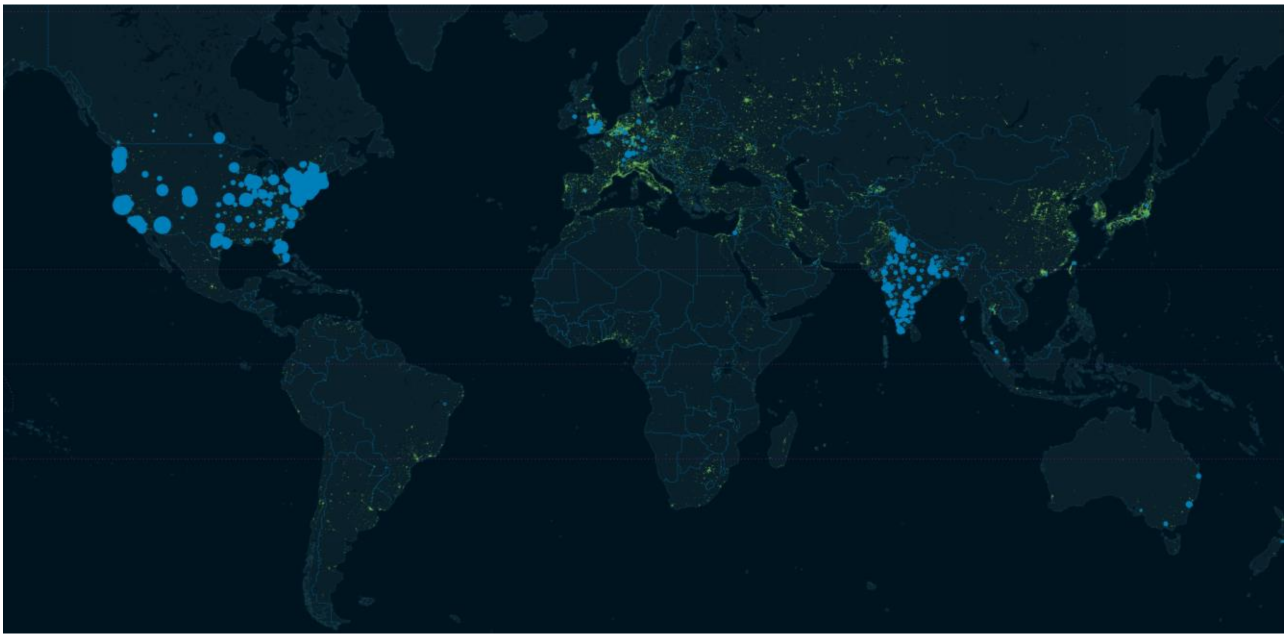
	GLS (1)	GLS MLE (2)	NEG. BIN. (3)	GLS (4)	GLS MLE (5)	NEG. BIN. (6)
COH_MIG_1	0.019 (0.039)	-0.022 (0.027)	0.896** (0.050)	-0.047 (0.11)	-0.164** (0.077)	0.885*** (0.034)
COH_MIG_2	0.106** (0.042)	0.064** (0.012)	1.037 (0.025)	0.296*** (0.109)	0.182** (0.083)	1.139*** (0.053)
COH_MIG_3	0.081** (0.014)	0.054** (0.010)	1.086* (0.054)	0.266*** (0.082)	0.209*** (0.068)	1.180*** (0.047)
COH_MIG_4	0.173*** (0.011)	0.141*** (0.041)	1.277*** (0.111)	0.427*** (0.145)	0.320*** (0.120)	1.489*** (0.102)
CMOV	-0.017*** (0.003)	-0.022*** (0.002)	0.950*** (0.005)	-0.019** (0.009)	-0.035*** (0.006)	0.957*** (0.004)
CMOV*CMOV	0.001*** (0.001)	0.003*** (0.001)	1.007*** (0.001)	0.004*** (0.001)	0.009*** (0.001)	1.006*** (0.001)
EDLEV	0.034*** (0.003)	0.023*** (0.002)	1.041*** (0.005)	0.084*** (0.008)	0.053*** (0.007)	1.063*** (0.004)
Observations	253,239	253,239	253,239	253,239	253,239	253,239
Nbr Inventors	40,806	40,806	40,806	40,806	40,806	40,806
R-squared	0.032			0.045		
Year, TITLE, COH, SELEC	Yes	Yes	Yes	Yes	Yes	Yes
LR test sigma_u=0		3.0e^04***			2.1e^04***	
Wald test			4222***			1.0e^4***
LR test			3.8e^4***			7080***

*** p<0.01, ** p<0.05, * p<0.1

This table estimates the effect of being a migrant decomposed by cohort of entry in the United States, the baseline cohort being the natives, on inventors' productivity with different combinations of characteristics. Furthermore, we include the variable that captures the selection effect ($SELEC_{it}$) described table 9. Standard errors appear in parenthesis and are clustered at the inventor level. Column (1),(2) and (3) shows the effect on the number of patents produced yearly per inventor, measured as Log(1 + $NPAT_{it}$) for columns (1) and (2) and as $NPAT_{it}$ for (3). Column (4),(5) and (6) show the effect on the total number of citations received yearly per inventor, measured as Log(1 + $NCIT_{it}$) for columns (4) and (5) and as $NCIT_{it}$ for (6). All columns are estimated with a random effect model, with a maximum likelihood estimator (MLE) for column (2) and (5) as a robustness check.

2.6.5 Appendix E: Indians localization and other stats

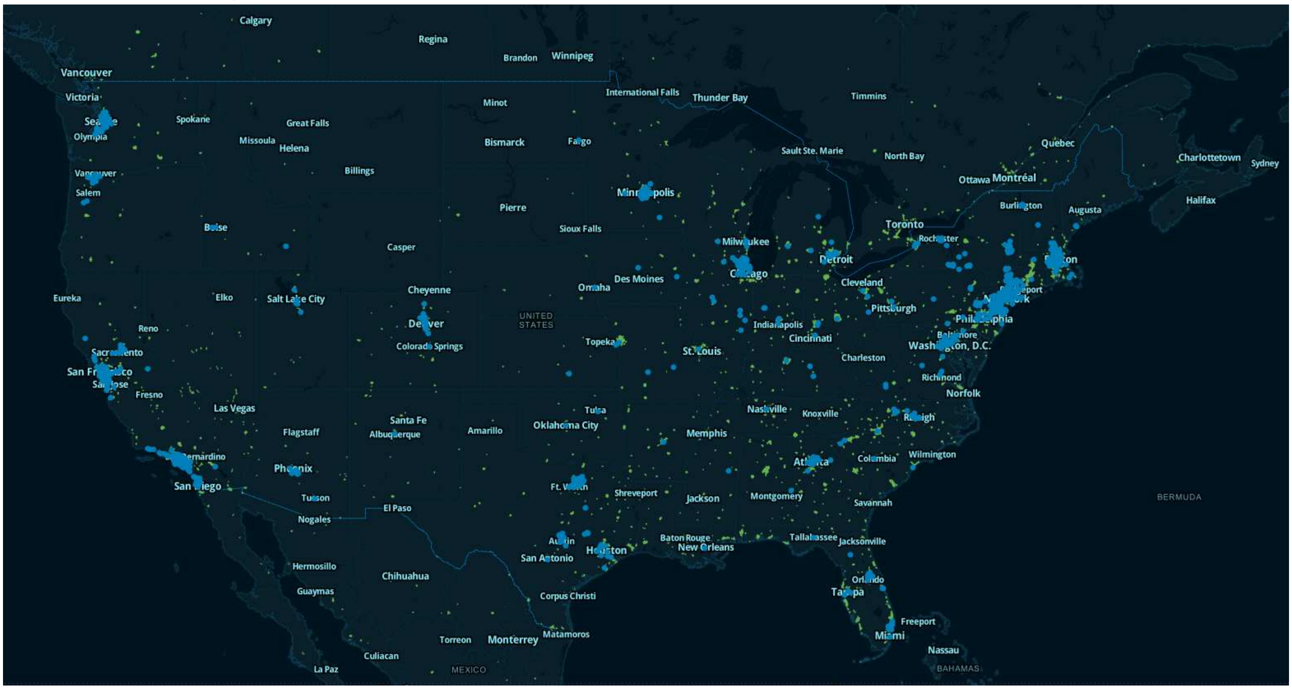
Figure 2.E.1: Worldwide repartition of Indian's inventors (into the database)



Please note that the green points are the luminosity, being both a proxy for technology and population density. The blue points are the Indians localization when they have made a patent.

(Source: Author's calculation from the Linked Inventor database)

Figure 2.E.2: Indians' localization United States (our population)



Please note that the green points are the luminosity, being both a proxy for technology and population density. The blue points are the Indians localization when they have made a patent.

(Source: Author's calculation from the Linked Inventor database)

Table 2.E.1: Natives and Migrants broken down by entry channel

Descriptive statistics (N=40.806)

	Natives	Indians' Migrants		
	All	Education channel	Changing company	Within company
	(1)	(2)	(3)	(4)
No. of inventors	36010	2815	1353	628
Total Nbr. of patents	6.42	7.49	5.59	4.97
Total Nbr. of citations	135.1	142.2	92.4	82.9
Gender (1 = male)	0.89	0.80	0.85	0.84
Level of education				
- Bachelor	0.34	0.01	0.34	0.19
- Master	0.36	0.45	0.38	0.41
- Ph.D./MBA	0.28	0.54	0.27	0.40
Title				
- Engineer	0.63	0.60	0.66	0.67
- Manager	0.17	0.18	0.18	0.17
- Company's head	0.07	0.06	0.05	0.05
- Scholar	0.06	0.10	0.07	0.07
- Founder	0.02	0.02	0.008	0.01
- Others	0.03	0.01	0.03	0.02

Chapter 3

Learning by moving: International Education and Return migration of Europeans

We study the effects of international migration on knowledge diffusion, by using an original dataset that links patent data with biographical information for a large sample of European ICT inventors, many of whom with educational or working experiences outside their home countries. First, we test whether these inventors are, upon return, more productive than native ones ("learning by moving" effect). Second, we test whether the strength of learning by moving depends on the type of international experience (educational versus working experience; and with or without patenting experience abroad). Overall, we find that return migrants are more productive than non-migrants. We do not see any significant productivity difference between inventors with and without an international education. On the other hand, we find differences related to the working experience, depending on the type of positions held (R&D-related or managerial), a result that is robust once we control for several determinants of positive and negative selection.

3.1 Introduction

A longstanding tradition of emigration studies evaluates to what extent origin countries get positive returns from emigration. The first studies on this topic emphasize the role of emigrants' financial remittances on the capital formation of less-favoured regions or countries. However, with the increasing importance of highly skilled migration, the emphasis has shifted to migrants' contribution to knowledge diffusion and innovation in their countries of origin. These "knowledge remittances" may occur via three channels:

1) "Ethnic-bound" knowledge spillovers: emigrants' scientists and engineers may retain social contacts with former fellow students or educational institutions in their home countries and transmit them the scientific and technical skills they acquired abroad. These knowledge spillovers can occur through several channels, such as friendly or contractual contacts, visiting professor programs, research collaborations, or firm consultancy.
2) Diaspora networks: emigrant scientists and engineers working abroad may decide to come together as an associative platform, this to establish collaborative links with their respective home countries or regions. The main goal of such networks is to channel knowledge back home. Most of these organizations are constituted by highly skilled migrants from emerging and developing countries. Some of these networks are supported or initiated with the collaboration of sending countries governments.
3) Returnees' direct contribution: emigrant scientists and engineers, enriched from their experience in the destination country, may decide to move back to their origin countries to continue their activities.

This chapter investigates the last channel of knowledge remittances focusing on a specific group of highly skilled migrants, namely inventors. Specifically, we compare the productivity of return migrants and non-migrant colleagues at origin. Hence, we observe at least a "raw" brain gain when the population of return migrants is more productive than the non-migrants one. In fact, there is a "net" brain gain only if the return migrants become more productive, otherwise there is only a "temporary brain drain", for the home country, while they are abroad. Nevertheless, due to the composition of our sample and our main research question, that is whether return migrants are more productive than natives, we do not investigate the latter.

In the last decades a relevant stream of research on inventors has focused on mobility, both across firms (Hoisl, 2007; Agarwal, et al., 2009; Latham, et al., 2011) and across countries (Hunt & Gauthier-Loiselle, 2010; Hunt, 2011; Kerr, et al., 2017; Cappelli, et al., 2019). Most studies focus either on the learning effects that changing employer might bring along, or on whether migrant inventors contribute to knowledge diffusion, especially in their home countries. The two mobility dimensions are rarely combined, despite a clear indication on the importance of firms, especially multinationals, as channels of international migration of highly-skilled workers (Kerr, et al., 2016). Nor we find many studies that examine the international mobility of inventors since they were students, despite the recognized importance of international student mobility, especially in scientific and technical fields, as an essential source of international migration and despite a rich historical literature of the importance of such type of mobility for knowledge diffusion.

We contribute to the literature on migration and innovation by comparing, in their home country, the productivity of return migrant inventors with stayers. We test whether the former group is, on average, more productive than the latter, and whether this difference depends on the type and quality of experience that migrants acquired abroad. We consider two types of experiences, namely working versus educational. Concerning the latter, we differentiate between different levels of education (Bachelor, Master, and Ph.D./MBA) university ranking.

Although the inventors' determinants of productivity are nowadays well known (among others mobility (Hoisl, 2007; Hoisl, 2009; Latham, et al., 2011); team and company size (Schettino, et al., 2013); personality (Zwick, et al., 2017)), research comparing inventors with and without an international education or an international labour experience is still scant. One reason for the lack of individual-level studies dealing with this issue is the absence of appropriate data. Bibliographic data alone, do not provide information on crucial productivity's drivers, such as career attainment or education experiences.

We attempt to fill this gap by studying the effect of return migration on inventor's productivity, considering different past labour market and education experiences in the destination country. To do so, we control for the inventor's international education, intra-country mobility experience (changing of company), and information on education, career, and patenting experiences.

For this purpose, we use a new and original dataset on Europeans inventors in the EU-28 plus Norway, Switzerland, Ukraine, and Russia, extracted from the database described in chapter 1. The dataset includes information on education, labour market, and patenting experience for a sample of 3.372 Europeans inventors active during the period 1974 and 2016. Among these, 2.871 individuals completed their education at home and 501 have at least one experience as an international student. We can observe the worldwide localization of the patents made by all the Europeans inventor of our sample in Appendix B; in Europe, Figure 3.B.1; in North America, Figure 3.B.2; in the Middle East and Asia, Figure 3.B.3 and Oceania, Figure 3.B.4. Concerning return migration during the labour market activity, 334 inventors came back in their country of origin after spending at least one year on the labour market in the destination country.

In our empirical strategy we measure productivity with the number of patents filed at origin and the number of claims or forward citations for these patents. Our main variable of interest is a dummy variable that takes the value 1 if the inventor is a return migrant, 0 otherwise. We control for international education, mobility, and other individuals' characteristics. First, we use a GLS (Generalized Least Squares) random effect estimator, but, due to the nature of our dependent variables and the overdispersion of their distribution, we also fit a negative binomial model as a robustness check.

When investigating the difference in productivity between natives and return migrants, we consider the endogeneity issues related to the positive or negative self-selection based on observable characteristics of the individuals. Return migrants go through a two-stage selection process. First, according to the literature that argues that migrants are the *best and brightest* (Bhagwati & Hamada, 1974) migrants might be positively selected, especially if at destination the returns on human capital are higher (Borjas, 1987)[24]. However, return migrants might be negatively selected with respect to migrants who stay at destination since, according to the literature, they represent the "*worst of the best"* (Borjas & Bratsberg, 1996; Rooth & Saarela, 2007; Zaiceva & Zimmermann, 2016).

These positive and negative selection effects may bias the estimated coefficients. As discussed in Chapter 2, positively selected inventor may turn out to be the most productive at destination. Second, the scarce literature about negative selection of return migrants in the destination country suggests that there is a positive relationship between the migrants' length of stay and productivity (Breschi, et al., 2018). It is crucial for host countries to retain the best and brightest migrants, those who can contribute the most to innovation. Hence, we expect that return migrants with shorter work experiences in the destination country will be less productive than those who have spent more time there. For this purpose, we control for the past labour market time length at destination to account for negative selection.

This chapter is organized as follows: section 2 provides an overview of the background literature; section 3 presents the data and basic descriptive statistics; section 4 presents the model specification the estimation issues; section 5 the results; section 4 discusses the main findings and draws some conclusions.

3.2 Literature Review

Highly educated individuals are a consistent part of the flows of international migration (Docquier & Marfouk, 2006). The same holds for return migrants (Dustmann & Weiss, 2007). Recent evidence for the OECD member states shows that between 20 and 50 percent of migrants decide to return in their country of origin within five years (Dumont & Spielvogel, 2008). Hence the contribution of return migrants to the home country matters, based on the diversity of foreign qualifications, occupational skills and financial capital that they are going to invest in the home country (Piracha & Vadean, 2010).

3.2.1 The theories of migration selection

The analysis of the consequences of increased international mobility of high-skilled workers is often framed within a brain drain versus brain gain debate. If the early literature described the migration of skilled workers as a significant cost for sending countries (Bhagwati & Hamada, 1974; Bhagwati, 1976), the more recent one has focussed on the possibility of a positive effect, both through the return migration

[24] Borjas finds that countries with higher returns on human capital attract higher skilled migration. This hypothesis is confirmed empirically by (Moraga, 2011; Dustmann & Glitz, 2011; Parey, et al., 2017).

channel and the increased incentives to invest in education in the sending country (Stark, et al., 1997). Within these two kinds of literature, the Borjas and Bratsberg (1996) theory of optimal migration and return migration posits that individuals decide to migrate and to return based on their skills and the skill-based wage differential between the sending and receiving countries. The migration selection model predicts that, given sufficiently high portability of skills between source and destination countries, and time-equivalent migration costs, labour migrants are negatively (positively)selected on unobservable characteristics, such as abilities and productivities, if the source country has more (less) dispersion in its earnings distribution, and negatively (positively) selected on observable skills, such as education, if the returns from educational attainment is relatively higher (lower) than in the destination country. This is because it would be relatively less (more) rewarding for people with higher skills to migrate than for those with lower skills. It is essential to note that there is no relationship between the types of selection that are generated in unobserved characteristics and the types of selection that are generated in observed characteristics (Borjas, 1987). Since these two dimensions of "quality" are unrelated, negative selection in unobserved characteristics may be concurring with positive selection in observable characteristics, or vice versa. For instance, it is possible for a given country of destination to attract poorly educated persons, but these poorly educated migrants might be the most productive in the population of poorly educated workers. The theory of selection in return migration additionally incorporates reversible migration decisions. Return migration may occur for two distinct reasons. It may be the optimal residential location plan over the life cycle, which allows some workers to attain higher utility than if the migration decision was permanent, or it may result from mistakes in the initial migration decision. Either way, implications for the surviving stock of migrants in the host country are the same: return migration accentuates the selection that characterizes the initial migration flow. This implies that in the case where the migration flow is negatively selected on skills, return migrants are the "best of the worst", and if it is positively selected on skills, return migrants are the "worst of the best". Thus, the intuition is that the forces driving selection in migration also drive selection in return migration (Rooth & Saarela, 2007) (Zaiceva & Zimmermann, 2016).

Theoretically, Borjas and Bratsberg (1996) show that depending on whether migrants are positively or negatively selected at destination, emigration amplifies the original selection process. If they are positively selected, return migrants tend to be the "*worst of the best*"; if they are negatively selected, returnees are the "*best of the worst*". These theoretical findings are consistent with the empirical results of Ramos (1992), who used US census data to show that Puerto Rican migrants to the US mainland were negatively selected as a group, but returnees were the most skilled among them.

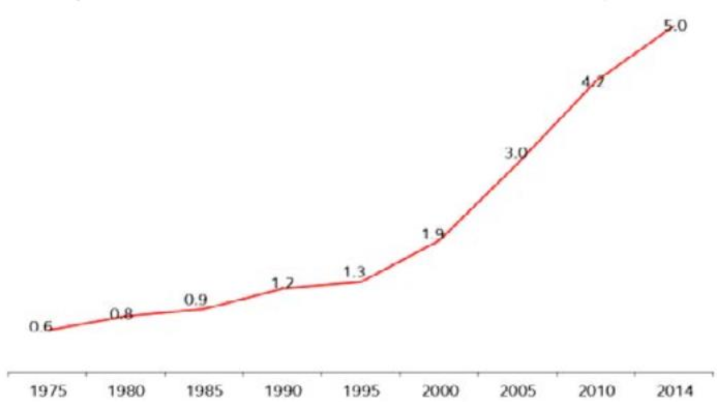
Thus, theory and evidence from the United States suggest that selection for return migration may be either positive or negative, depending on a variety of circumstances. A similarly complex picture emerges from studies carried out for other groups and countries. Barrett and Trace (1998) show that return migrants from Ireland have higher educations than those who remain abroad. In contrast, Bauer and Gang (1998) found that Egyptian return migrants were negatively selected with respect to skill and having prior migrant experience and access to social networks abroad shortened that length of stay. Sending remittances home also lengthened trips, a finding also found among unskilled Mexican immigrants to the US (Lindstrom, 1996). Dustmann (1993) finds that greater integration yields longer intended durations of stay, he shows that migrants in Germany tend to stay longer if they speak German, they are married to a German and have young children.

Recent updates of the theory stress that migrants decide to migrate to accumulate financial resources or human capital aiming to come back in their origin's country. This last finding suggests, on a theoretical framework, that return migrants should be more skilled than their native counterparts. More recently Mayr and Peri (2009) modelled the effects of the change in immigration policies between Eastern and Western Europe after 1990 on the flows of migrant and return migrant, finding a brain gain effect for Eastern Europe even after the huge emigration waves, implying a brain drain for Eastern Europe, caused by the iron curtain fall. Most empirical research on return migration has focused on the return migrants contribution to economic development in the countries of origin (Asiedu, 2005) (Rodríguez & Egea, 2006). There is nowadays, a growing literature on the individual and contextual factors which determine return migration, aiming to quantify the probability of return among migrants in Germany (Constant & Massey, 2002), among migrants in United-Kingdom (Dustmann & Weiss, 2007), among African migrants in Spain and Italy (Fokkema & De Haas, 2015), among Moroccan migrants in Europe (De Haas, et al., 2015) and among Indian inventors in United States (Breschi, et al., 2018).

3.2.2 The globalization of education

Over the last decade, the international mobility of students has increased. Foreign student enrolment worldwide has grown from 107.589 in 1950 (Barnett & Wu, 1995) to more than 3.3 million in 2008 (UNESCO, 2009) and more than 4.8 million in 2015 (UNESCO, 2018).

Figure 3.1: Worldwide raise of international student mobility (millions)



Source: UIS.Stat http://apiportal.uis.unesco.org

Two factors have contributed to this rapid growth. First, the second half of the twentieth century saw an increase in higher education enrolment around the world. Global university enrolment in 1950 was nearly six million, reaching 132 million in 2004 (Gürüz, 2008). Second, student exchange programs became integral to the international development sector and the foreign policy agendas of many countries. During the Cold War, both the United States and the Soviet Bloc used student exchanges to maintain and gather international support and to project images of their power and prosperity (Barnett & Wu, 1995).

While at the same time the composition of the major destination countries remains mostly stable, with the main host countries the United States, United Kingdom, Germany, France, and Australia, the main sending countries in terms of the number of students abroad are nowadays China and India. An increasing amount of so-called 'semi-finished human capital' (Khadria, 2012), based on an individual decision, go to upgrade themselves to institutions of higher education, with a large share of them going to the most prestigious universities. After their graduation, they are considered as an asset for the host, origin and even third-party country, with many of the newest international graduated students facing a decision which of the various options offered to them would bring the highest benefit.

Studying abroad inevitably entails several benefits for all the actors involved in the process, such as universities, country of origin, country of destination, and students that are the central beneficiaries of this process. The worldwide competition to attract more generally highly skilled workers is also operating at the regional level. The E.U. countries are less competing within the E.U., nevertheless implementing immigration policies such as the French "Scientific visa" to allow scientists from countries not a member of the European Economic Area (EEA) to work in France. Likewise, Germany introduced the “Green cards” for workers in the communication technologies from non-EEA countries; countries like Sweden and Netherlands provide tax discounts for highly skilled foreign workers including workers from the E.U. (Mahroum, 2001). Nevertheless, these policies are not always pursued with much conviction, European workers are free to move within the E.U. and the main students' mobility push is a cooperative program such as Erasmus.

Measuring student mobility and its benefits are consistently set up and studied by many world institutions. Reports[25] on the Erasmus experience by students often tell about an outstanding experience having a long-lasting impact on their lives and making them discover a new European identity. Research studying the impact of Erasmus program on International labour market mobility find a causal effect of studying abroad on later labour market mobility (Parey & Waldinger, 2011). Europeans are, nevertheless, not the forerunner for programs fostering student mobility. Findings on the US Fulbright Visiting Students relativized the importance of this program, in fact, it is small compared to more than a million of international students who were enrolled in graduate programs in the 2017-2018 academic year in United States (Baer, 2018). Studies comparing Fulbright students with the other international student in the US did not find any positive impact of this kind of scholarship program on their scientific productivity, both in the number of articles produced and citations received. A Fulbright student coming from rich countries (Western Europe, Canada) performed as well as their foreign student colleague and performed even less when comparing Fulbright and international student from poorer countries (Kahn & MacGarvie, 2011).

3.2.3 Human capital theory and economic returns through migration

The human capital theory, which posits a positive relationship between investments in education and its economic returns (Becker, 1962), suggests that human capital, as embodied in the skills, knowledge, and competences of workers, influences future income. Adding that individuals seek to maximize their utility over their lifetime by investing in human capital (education and training), the Becker's (1965) time allocation and human capital theories are therefore the pillars of the economics of education. Extending the human capital approach to international migration and education, Sjaastad (1962) applied the concept of human capital investment to migration decisions; treating migration as an investment involving costs and returns, affecting individual's decision to migrate. Empirical studies brought evidence of the previous cited theories. Dustmann and Kirchkamp (2002) show that the acquisition of tertiary education in a foreign country may yield a higher return on the home country's labour market. Nevertheless, even if at an individual level, findings show that studying abroad helps students in expanding their knowledge, skills, and boosts their labour market prospects (Knerr, et al., 2010), empirical evidence on the impact of international education on human capital enhancement is mostly

[25] Erasmus student network ESN www.esn.org

focused on academics and scientists. Corley and Sabharwal (2007) find that foreign-born scientists with a US Ph.D. are more productive than US natives. However, their salaries and job satisfaction are lower than the US natives. Other studies find a negative impact of studying abroad on productivity compared with the domestic degree. Shin and al., (2014) show that Korean, Hong Kong and Malaysian academics with a foreign degree are not more productive and possibly less productive than domestic degree holders, unless they have further experience in the destination country.

3.3 Descriptive statistics and sample description

The database we use in this chapter is the same described in chapter 1, which retrieves detailed information on the inventors matching 424.496 public LinkedIn profiles with patent data. We focus on US companies working in Semiconductor and ICT sectors and consider the period between 1950 and 2016.

In this chapter we focus our attention to inventors who are born in Europe. We define return migrants as inventors who studies and/or worked at destination. On the other hand, we define non-migrants as inventors who work and filed at least one patent for the companies in the sample, always in their home country. We gather the inventor localization from both the information on the LinkedIn profile and/or from the inventor address in the patent's documents.

The ultimate goal of this matching procedure is to enrich the inventor information we can retrieve from patent data (address at the time of the patent, name of the applicant, identity of co-inventors, and other contents) with information on the migrant vs. native status of the inventors (plus, for migrants, their country of origin and year of entry in the US) and information on education and labour market experience.

To assign a country of origin to the inventors, we create an algorithm that exploits both the information from LinkedIn (such as the country where the earliest education levels have been attained, the individual's native language, and any useful biographical detail) as well as further information such as:

- The inventor's nationality, as reported on a subset of USPTO patent applications filed according to the PCT procedure before 2011 (Miguelez & Fink, 2013).
- The results of name analysis, based on the combination of statistics on the ethno-linguistic origin of names and surnames from the IBM-GNR dataset (Breschi, et al., 2014) as well as additional linguistic analysis (Tyshchenko, 1999).

Our data also allow us to track a substantial part of the inventors' careers, most notably their mobility before the first and after the last patent filed (the two dates coincide for the vast number of inventors with just one patent over their lifetime). This solves one of the significant limitations of previous studies on mobility and productivity of inventors, which were exploiting the information on the inventors' addresses reported in patents document, and were able to track only the mobility of inventors who filed at least two patents (Hoisl & de Rassenfosse, 2014) (Hoisl, 2007) (Hunt & Gauthier-Loiselle, 2010).

We identify and classify two different types of mobility, namely education and work mobility. We consider mobile during education, referred throughout this chapter as international student, and inventor who obtained at least one university degree abroad. Then, we consider two kinds of work mobility. In general, we consider company mobile and inventor who changed of company, then we differentiate between international and within country company mobility. The categories are not mutually exclusive, as some individuals are mobile for both education and work.

The longitudinal nature of the data allows us to account for some endogeneity issues that arise when we try to establish a causal link between our variables of interest and the productivity of inventors. The final panel used in this chapter, consists of 3.372 Europeans born inventors, who patented at least once between 1974 and 2015. Of these, 2.638 individuals are non-migrants without any experience (education/labour) abroad, and 734 individuals are return migrants (as described in Table 3.1). Amongst the return migrants, we can distinguish 400 inventors who came back in their country of origin with at least one international degree completed at destination, 233 inventors who came back in their origin's country with at least one year of labour market experience at destination, and 101 with both international education and labour market experience abroad. We observe and compare the productivity of the inventors only in their home country, meaning that we are comparing inventors who work in the same technological and in the same working environment, their home country. We believe that, with a more homogenous population we can more accurately assess the effect of return migration on productivity.

Table 3.1: List by country of origin of European inventors

Country of origin	Nbr. of inventors	Nbr. of return migrants	Nbr. of return migrants International education	Nbr. of return migrants Working exp. abroad
AT	30	11	7	5
BE	43	10	7	6
BG	5	2	1	1
CH	57	23	18	10
CZ	33	6	3	3
DE	626	146	110	67
DK	88	25	16	10
EE	3	0	0	0
ES	98	27	26	5
FI	59	15	13	3
FR	444	119	74	59
GB	1043	156	77	95
GR	15	9	7	7
HR	1	0	0	0
HU	7	1	1	0
IE	207	47	33	23
IT	273	50	43	12
LU	1	1	1	1
MT	2	1	1	0
NL	86	24	15	9
NO	38	11	9	2
PL	31	8	5	4
PT	5	3	3	2
RO	21	6	6	1
RU	86	14	12	3
SE	63	19	13	10
UA	7	0	0	0
N	3372	734	501	334

The decision to patent a specific innovation is part of a company's strategy. Therefore, the productivity of inventors working in companies with weak intellectual property rights (IPRs) policy will be underestimated if compared to inventors employed in companies with a stronger IPRs policy. Also, during his lifespan inventors can change job position, being farther (or closer) to the R&D activities of the company's and holding more a managerial position rather than an engineering one. For these reasons, in our regressions we control for the inventors' company position.

In table 3.2 we compare the education levels, productivity, and positions held of non-migrants (Column 1) and return migrants (Columns 2). In particular, our results based on a t-test of mean differences at 5% significance level (Columns 3), show that return migrants are on average better educated (in terms of level and university ranking) than non-mobile inventors. Also, we find return migrants to be more likely to change company during their career. Concerning productivity, we find return migrants filed their first patent earlier, however, we do not find any significant difference for the average number of patents filed, the number of citations received, and the number of claims between non-migrants and return migrants, averages computed on the whole period observed. We do find some differences considering the positions held, according to job descriptions in their CVs. In fact, non-migrants hold more engineering positions, whereas return migrants hold more scientists' ones.

Table 3.2: Inventors' characteristics broken down, Non-migrant vs. Return migrant

	(1) Non-migrant	(2) Return migrant	(3) Non-migrant – Return migrant
Education level[b]	2.059 (0.699)	2.351 (0.647)	-0.293*** (-10.64)
Best university rank reached[b]	5.416 (3.910)	6.448 (4.130)	-1.032*** (-6.06)
Number of company move	1.063 (1.167)	1.213 (1.144)	-0.150** (-3.12)
Age at first patent filed	34.06 (6.207)	33.37 (5.835)	0.687** (2.78)
Number of patents made	4.027 (6.342)	4.395 (8.453)	-0.368 (-1.10)
Number citations received	57.07 (138.9)	63.68 (149.7)	-6.612 (-1.07)
Number of claims made	227.5 (853.5)	240.4 (786.0)	-12.96 (-0.39)
Engineer[a]	0.729 (0.444)	0.586 (0.493)	0.144*** (7.12)
Founder[a]	0.00720 (0.0846)	0.0136 (0.116)	-0.00642 (-1.40)
CEO[a]	0.0349 (0.183)	0.0518 (0.222)	-0.0169 (-1.89)
Manager[a]	0.113 (0.317)	0.138 (0.345)	-0.0246 (-1.74)
Others[a]	0.0383 (0.192)	0.0599 (0.238)	-0.0217* (-2.27)
Scientist[a]	0.0773 (0.267)	0.151 (0.359)	-0.0739*** (-5.20)
N	2638	734	3372

Note: Column (1) describes the general characteristics of all inventor's non-migrant composing the sample as well as column (2) for the inventors return migrant with mean coefficients; sd in parentheses. Then, column (3) shows the ttest performed on each category with ttest significant at 5%, coefficient is (1) – (2) for each category, *t* statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

For more detail about the variables please refer to Table 4: Variables description.

a: value in percentage

b: categorical variable

In Table 3.3 we follow the approach proposed by Shin and al., (2014) and distinguish between return migrants with an international education (Column 2) and return migrants with a working experience abroad (Column 3). We compare, with a t-test of mean differences at 5% of significance level, non mobile inventors with return migrants with a working experience (Column 4) and return migrants with an international education (Column 5). We find a significant difference between non-migrants and both the return migrants' groups for the level of education, whereas, when we consider the university ranking, we only find a difference between non-migrants and return migrants with international education. Also, we find a significant difference between non-migrants and return migrants with a working experience abroad in terms of number of company moves. When we consider productivity, computed as the cumulative number of patents, citations or claims, we do not find any significant difference between non-migrants and the two return migrant groups. This could be explained by the different time span in which we observe return migrants and non-migrants since we only consider patents filed at origin and, by definition, return migrants start their career at home later than the non-migrant. Finally, we find significant differences for engineer and scientist positions, for both the return migrants group compared to the non-migrants.

Table 3.3: Inventors' characteristics broken down, Non-migrant vs. Return migrant with an international education vs. Return migrant with a working experience abroad

	(1) Non-migrant	(2) Return migrant Work exp. Abroad	(3) Return migrant International Educ.	(4) (1) – (2)	(5) (1) – (3)
Education level[b]	2.059 (0.699)	2.329 (0.693)	2.407 (0.602)	-0.271*** (-6.72)	-0.348*** (-11.56)
Best university rank reached[b]	5.416 (3.910)	5.817 (4.049)	6.890 (4.169)	-0.402 (-1.71)	-1.474*** (-7.33)
Number of company move	1.063 (1.167)	1.410 (1.023)	1.122 (1.181)	-0.347*** (-5.75)	-0.0588 (-1.02)
Age at first patent filed	34.06 (6.207)	33.62 (5.870)	32.95 (5.656)	0.442 (1.29)	1.107*** (3.95)
Number of patents made	4.027 (6.342)	4.485 (8.217)	4.192 (8.110)	-0.458 (-0.98)	-0.165 (-0.43)
Number citations received	57.07 (138.9)	70.72 (157.3)	58.44 (138.1)	-13.65 (-1.51)	-1.373 (-0.20)
Number of claims made	227.5 (853.5)	320.0 (938.1)	188.2 (616.8)	-92.55 (-1.72)	39.29 (1.22)
Engineer[a]	0.729 (0.444)	0.515 (0.501)	0.605 (0.489)	0.214*** (7.46)	0.125*** (5.30)
Founder[a]	0.00720 (0.0846)	0.0210 (0.143)	0.0120 (0.109)	-0.0138 (-1.72)	-0.00477 (-0.93)
CEO[a]	0.0349 (0.183)	0.0808 (0.273)	0.0379 (0.191)	-0.0460** (-2.99)	-0.00305 (-0.33)
Manager[a]	0.113 (0.317)	0.147 (0.354)	0.130 (0.336)	-0.0337 (-1.66)	-0.0168 (-1.03)
Others[a]	0.0383 (0.192)	0.0629 (0.243)	0.0579 (0.234)	-0.0246 (-1.78)	-0.0196 (-1.77)
Scientist[a]	0.0773 (0.267)	0.174 (0.379)	0.158 (0.365)	-0.0963*** (-4.50)	-0.0804*** (-4.70)
N	2638	334	501	2972	3139

Note: Column (1) describes the general characteristics of all inventor's non-migrant composing the sample as well as column (2) for the inventors return migrant with a working experience abroad and, column (3) for those with international education, with mean coefficients; sd in parentheses. Then, column (4) shows the ttest performed between the non-migrants and the return migrants with a working experience abroad. And, column (5) shows the ttest performed between non-migrants and the return migrants with an international education with ttest significant at 5%, *t* statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
For more detail about the variables please refer to Table 4: Variables description.
a: value in percentage
b: categorical variable

Since we do not find evidence of a productivity gap between return migrant and non-migrants, we now consider inventors that succeeded to patent abroad and those who that did not. We test whether return migrants who did not patent at destination are more negatively selected than those who did. Table 3.4, reports statistics for return migrants who only patented at origin (Column 3), return migrants who patented at both destination and origin (Column 2), and non-migrants (Column 1).

Table 3.4: Inventors' characteristics broken down, Non-migrant vs. Return migrant with a patent filed abroad vs. Return migrant with no patent filed abroad

	(1) Non-migrant	(2) Return migrant patent filed abroad	(3) Return migrant no patent abroad	(4) (1) – (2)	(5) (1) – (3)
Education level[b]	2.059 (0.699)	2.338 (0.719)	2.357 (0.616)	-0.279*** (-5.46)	-0.298*** (-9.86)
Best university rank reached[b]	5.416 (3.910)	5.948 (4.091)	6.653 (4.133)	-0.533 (-1.83)	-1.237*** (-6.30)
Number of company move	1.063 (1.167)	1.319 (1.117)	1.169 (1.153)	-0.256** (-3.21)	-0.106 (-1.91)
Age at first patent filed	34.06 (6.207)	32.27 (5.454)	33.82 (5.930)	1.791*** (4.56)	0.235 (0.82)
Number of patents made	4.027 (6.342)	4.803 (9.267)	4.228 (8.099)	-0.776 (-1.20)	-0.201 (-0.54)
Number citations received	57.07 (138.9)	84.26 (185.3)	55.27 (131.7)	-27.19* (-2.09)	1.801 (0.28)
Number of claims made	227.5 (853.5)	442.4 (1143.2)	157.9 (560.7)	-214.9** (-2.68)	69.61* (2.35)
Engineer[a]	0.729 (0.444)	0.488 (0.501)	0.626 (0.484)	0.241*** (6.81)	0.104*** (4.52)
Founder[a]	0.00720 (0.0846)	0.0329 (0.179)	0.00576 (0.0757)	-0.0257* (-2.08)	0.00144 (0.39)
CEO[a]	0.0349 (0.183)	0.0986 (0.299)	0.0326 (0.178)	-0.0637** (-3.07)	0.00225 (0.26)
Manager[a]	0.113 (0.317)	0.150 (0.358)	0.132 (0.339)	-0.0373 (-1.47)	-0.0195 (-1.21)
Others[a]	0.0383 (0.192)	0.0610 (0.240)	0.0595 (0.237)	-0.0227 (-1.35)	-0.0212 (-1.92)
Scientist[a]	0.0773 (0.267)	0.169 (0.376)	0.144 (0.351)	-0.0917*** (-3.49)	-0.0666*** (-4.10)
N	2638	213	521	2851	3159

Note: Column (1) describes the general characteristics of all inventor's non-migrant composing the sample, as well as column (2) for the inventors, return migrant with a patent previously made abroad and, column (3) for those with no patent filed while abroad, with mean coefficients; sd in parentheses. Then, column (4) shows the ttest performed between the non-migrants and the return migrants with a patent previously made abroad. And, column (5) shows the ttest performed between non-migrants and the return migrants with no patent filed while abroad with ttest significant at 5%, t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
For more detail about the variables please refer to Table 4: Variables description.
a: value in percentage
b: categorical variable

Since we proxy success with patents at destination, we assume that inventors who only patented at origin are more negatively selected.

This latest descriptive result suggests that the learning by moving effect has a stronger effect on productivity for post-education experiences such as labour market (here we do not find any) or patenting activity abroad, as the Table 3.4 suggested. Nevertheless, all groups and sub-groups given by the latest tables show that inventors return migrant got a better level of education but also a better education quality, the learning by moving effect for education still exists but doesn't directly affect the inventor's productivity. Nevertheless, these cumulated productivity statistics do not account for the potentially different time length of observation in the country of origin between return migrants and non-migrants. This is why, in the next section, we consider only the per year productivity measures as dependent variables in our econometric models.

3.4 Model specification

As dependent variable, we proxy productivity by using the number of patents, the number of forward citations and number of patent claims for inventor *i,* in year *t*[26] . Equation 1 is our baseline specification. We compare, the productivity of return migrants and non-migrants ($RETURN_i$) in the country of origin by investigating the role played by company mobility ($CMOV_{it}$) to explaining the difference (if any).

$$Productivity_{it} = \beta_0 + \beta_1 RETURN_i + \beta_2 CMOV_{it} + \beta_3 X_{it} + \beta_4 X_i + \varepsilon_{it} \quad (1)$$

In equation 2, we distinguish between return migrants who got an international education ($RETURN\_EDUC_i$), and those who worked at least one year in the destination country ($RETURN\_WORK_i$). Note that the two groups are not mutually exclusive, as in our sample we have inventors with both education and work experience at destination.

$$Productivity_{it} = \beta_0 + \beta_1 RETURN\_WORK_i + \beta_2 RETURN\_EDUC_i + \beta_3 CMOV_{it} + \beta_4 X_{it} + \beta_5 X_i + \varepsilon_{it} \quad (2)$$

In equation 3 we investigate international education by distinguishing between inventors who have at least one international education experience in another European country ($EDUC\_EU_i$), in the United States ($EDUC\_US_i$), or somewhere else ($EDUC\_OTHER_i$). By doing so, we investigate if the international education location is a crucial determinant of the inventor's productivity.

$$Productivity_{it} = \beta_0 + \beta_1 RETURN\_WORK_i + \beta_2 EDUC\_EU_i + \beta_3 EDUC\_OTHER_i + \beta_4 EDUC\_US_i + \beta_5 CMOV_{it} + \beta_6 X_{it} + \beta_7 X_i + \varepsilon_{it} \quad (3)$$

In equation 4 we investigate whether differences in the working experience of the return migrant at destination determine his productivity back home. We do so by decomposing the group of return migrant in two sub-groups, the inventors that succeeded to patent in the destination country ($RETURN\_PAT_i$) and those who did not ($RETURN\_NOPAT_i$).

$$Productivity_{it} = \beta_0 + \beta_1 RETURN\_PAT_i + \beta_2 RETURN\_NOPAT_i + \beta_3 CMOV_{it} + \beta_4 X_{it} + \beta_5 X_i + \varepsilon_{it} \quad (4)$$

[26] The year t referring to the application date

In all the specifications, we control for several individuals' characteristics. Table 3.5 reportsthe description for both the time invariant (the level of education ($EDLEV_i$), the gender ($GEND_i$), the cohort (COH_i) that is based on the estimated year of birth and, the quality of their education measured by the ranking of the highest university reached ($EDQUAL_i$)) and time variants (the position held in the company ($POSITION_{it}$) or the experience on the labour market ($LABEXP_{it}$)) characteristics.

As we have information on the inventors since their first labour market activity, hence before the first patent filed, we are able to mitigate the reverse causality issue stressed by the previous literature on mobility and productivity. Also, by measuring the inventors' skills based on their education type and university ranking, we partially address potential issues of positive selection[27].

[27] Migration is not a random phenomenon; migrants choose to migrate because they are more skilled or motivated than the native population in the destination country.

Table 3.5: Variables description

Variable	Definition or formula to calculate
Productivity measures	
$NPAT_{it}$	The number of patents granted by the inventor i the year t. Application date 1969-2016 (Source. EPO)
$NCIT_{it}$	The number of forwards citations received by the patent of the inventor i the year t up to 2016 (Source. EPO).
$NCLA_{it}$	The number of claims made by the patent of each inventor i the year t. (Source. EPO)
Company's position	
$POSITION_{it}$	Categorical variable describing the type of position that the inventor i was occupying the year t. We observe six positions, such as Engineer; Manager; Founder or Co-Founder; Company's head (CEO); Scientist; Others.
Mobility and Migration status	
$RETURN_{i}$	Dummy variable indicating whether inventor i is a return migrant or a non-migrant native.
$RETURN\_WORK_{i}$	Dummy variable indicating whether inventor i is a return migrant with working experience abroad or a non-migrant native.
$RETURN\_EDUC_{i}$	Dummy variable indicating whether the inventor i is a return migrant with international education or not.
$EDUC\_EU_{i}$	Dummy variable indicating whether the inventor i is a return migrant with an international education within Europe.
$EDUC\_US_{i}$	Dummy variable indicating whether the inventor i is a return migrant with international education in United-States.
$EDUC\_OTHER_{i}$	Dummy variable indicating whether the inventor i is a return migrant with international education in another region than the EU or US.
$RETURN\_PAT_{i}$	Dummy variable indicating whether inventor i is a return migrant with a patent done in the destination country.
$RETURN\_NOPAT_{i}$	Dummy variable indicating whether inventor i is a return migrant with no patent done in the destination country.
$CMOV_{it}$	Count variable indicating how many times the inventor i as changed of company up to year t.
Education Measures	
$EDLEV_{i}$	Count variable giving the education level of inventor i among three; Bachelor = 1, Master = 2 or PhD/MBA = 3.
$EDQUAL_{i}$	Count variable indicating the highest-ranked university, in function of the degree's type, where the inventor i has been during his education. See Appendix.
Labour market Measures	
$LABEXP_{it}$	Count variable giving the labour market experience, since the first experience in a company of the inventor i the year t.
Other demographic information	
COH_{i}	Categorical variable giving the inventor's cohort among five; starting with the inventors born before 1950 and ending for those born between 1980 – 1990.
$GEND_{i}$	Dummy variable giving the inventor's gender.

3.5 Discussion of the results

3.5.1 Different experience abroad

In this section, we present and discuss the results of the estimated models. Table 3.6 presents the results of the return migrant dummy on the log number of patents, Table 3.7 the log number of forward citations, and Table 3.8 the log number of claims. Each column of the three tables strictly corresponds to the equations discussed in the previous section.

In Column 1 of each table, we show a specification with the main variable of interest only and the individual characteristics reported in Table 3.5. We find a significant and positive productivity difference between return migrants and non-migrants. The number of patents filed by return migrants is on average higher by 5.1%, their number of citations by 10.7% and their number of claims by 7.2%. These results are confirmed by the negative binomial specification (Appendix A, Table 3.A.1, for the number of patents; Table 3.A.2 for the number of citations; and, table 3.A.3 for the number of claims, Column 1 of each table). Therefore, we find that return migrant inventors are more productive than non-migrant ones.

In Column 2, we decompose the return migrants' group between those who did a part of their education abroad and those who did a part of their labour market experience abroad. For all the dependent variables and models (Column 2 of Tables 3.6, 3.7, and 3.8) we do not find any significant effect on the productivity for return migrant inventors with international education. On the contrary, we do find a significant and positive difference in productivity for return migrants with a labour market experience abroad, compared to non-migrants. These results are confirmed by the Tables 3.A.1, 3.A.2 and 3.A.3 in Appendix A. This last result, on return migrant with international education, differs from the previous findings on the researcher's productivity and international education. Kahn & MacGarvie (2011) Shin and al., (2014) all found a negative impact of researchers with an international education compared to those with a home education only. They also found that for some destination country, there is no effect on the productivity, measured by the number of publications and citations received, between researchers with and without an international education.

Since we do not find any significant difference in productivity for inventors with international education, we decompose this group in three different categories (column 3): if the inventor did a part of his/her education in United States ($INTSTUD\_US_i$); a part in another European country ($INTSTUD\_EU_i$) or somewhere else ($INTSTUD\_OTHER_i$). However, in Appendix A, Table 3.A.2, Column 3, we find a positive and significant coefficient for return migrants who studied in another European country, which corresponds to a higher number of citations received by 11.9%, compared to the non-migrants. This result, mainly because it concerns only the number of citations received, can be driven by network effects, if the ties made during education between Europeans are stronger than those made in United States or somewhere else.

Finally, in column 4 of each Table 3.6; 3.7 and 3.8, we decompose the return migrant variable into two categories that depend on the different experience in the destination country. We investigate the average difference in productivity, with respect to non-migrant natives, for a return migrant who succeeded to patent in the destination country ($RETURN\_PAT_i$) and for return migrants who did not ($RETURN\_NOPAT_i$). For all the dependent variables and models, we find a positive and strongly significant difference in the productivity for return migrants who succeeded to patent in the destination country. We find that these inventors have a positive and significant coefficient for the number of patents of 17.5%, the number of forward citations of 25.8% and the number of claims made of 25.9%. Again, modelling the independent variables by a negative binomial confirms these results. When we analyse the return migrants who did not patent in the destination country we find, for the number of patents made, a positive coefficient slightly significant at 10%, confirmed by the negative binomial model Table 3.A.1, Column 4. Considering the number of claims made, we do not find any significant difference.

Nevertheless, we do find a significant result regarding the number of citations received by return migrants without a patent in the destination country. However, when considering the number of claims and the number of patents as a measure of productivity, we always find the return migrants with a patent in the destination country are more productive than those who don't have a patent in the destination country[28]. These last results suggest that within the flow of return migrants a sub-group is more negatively selected than the other. We also found that this previous sub-group could not be systematically more productive than their non-migrants' colleagues. However, when we consider the number of citations, we do not find any difference between the parameters of return migrants with and without a patent abroad.

[28] Confirmed by a Wald test that compares the equality of the parameters RETURN_PAT and RETURN_NOPAT at 5%.

Table 3.6: Return Migrants vs. Non-migrants productivity, Log Annual Number of Patents

	(1) Log (1 + $NPAT_{it}$)	(2) Log (1 + $NPAT_{it}$)	(3) Log (1 + $NPAT_{it}$)	(4) Log (1 + $NPAT_{it}$)
$RETURN_i$	0.0510^{**} (0.0211)			
$CMOV_{it}$	0.0108^{*} (0.00599)	0.0104^{*} (0.00596)	0.0105^{*} (0.00596)	0.0113^{*} (0.00594)
$RETURN\_WORK_i$		0.135^{***} (0.0311)	0.137^{***} (0.0312)	
$RETURN\_EDUC_i$		-0.0102 (0.0238)		-0.009 (0.0237)
$EDUC\_EU_i$			0.00246 (0.0292)	
$EDUC\_OTHER_i$			-0.0128 (0.0418)	
$EDUC\_US_i$			-0.0570 (0.0441)	
$RETURN\_PAT_i$				0.175^{***} (0.0420)
$RETURN\_NOPAT_i$				0.0661^{*} (0.0375)
Nbr. observations	14077	14077	14077	14077
Nbr. inventors	3372	3372	3372	3372
R-squared	0.0630	0.0631	0.0635	0.0638
Controls	YES	YES	YES	YES

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table estimates the effect of being a return migrant on inventors' productivity with different combinations of characteristics. Coefficient with reported marginal effects. Standard errors appear in parenthesis and are clustered at the inventor level. Table 3.6 shows the effect on the number of patents (*$NPAT_{it}$*) produced yearly per inventor. All columns are estimated with a random effect GLS model. Besides the listed variables, we control for all specifications by country of origin and time fixed effect; position in the company (*$POSITION_{it}$*); gender (*$GEND_i$*); cohort (*COH_i*); labour market experience (*$LABEXP_{it}$*) and education quantity (*$EDLEV_i$*) and quality (*$EDQUAL_i$*).

Table 3.7: Return Migrants vs. Non-migrants productivity, Log Annual Number of Citations

	(1) Log (1 + $NCIT_{it}$)	(2) Log (1 + $NCIT_{it}$)	(3) Log (1 + $NCIT_{it}$)	(4) Log (1 + $NCIT_{it}$)
$RETURN_i$	0.107** (0.0453)			
$CMOV_{it}$	0.0673*** (0.0146)	0.0663*** (0.0146)	0.0668*** (0.0146)	0.0671*** (0.0147)
$RETURN\_WORK_i$		0.225*** (0.0651)	0.230*** (0.0655)	
$RETURN\_EDUC_i$		0.0242 (0.0542)		0.0247 (0.0542)
$EDUC\_EU_i$			0.116 (0.0711)	
$EDUC\_OTHER_i$			0.00229 (0.0813)	
$EDUC\_US_i$			-0.159 (0.106)	
$RETURN\_PAT_i$				0.258*** (0.0869)
$RETURN\_NOPAT_i$				0.179** (0.0871)
Nbr. observations	14077	14077	14077	14077
Nbr. inventors	3372	3372	3372	3372
R-squared	0.0630	0.0631	0.0635	0.0638
Controls	YES	YES	YES	YES

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table estimates the effect of being a return migrant on inventors' productivity with different combinations of characteristics. Coefficient with reported marginal effects. Standard errors appear in parenthesis and are clustered at the inventor level. Table 3.7 shows the effect on the total number of citations ($NCIT_{it}$) received yearly per inventor. All columns are estimated with a random effect GLS model. Besides the listed variables in each panel, we control for all specifications by country of origin and time fixed effect; position in the company ($POSITION_{it}$); gender ($GEND_i$); cohort (COH_i); labour market experience ($LABEXP_{it}$) and education quantity ($EDLEV_i$) and quality ($EDQUAL_i$). Furthermore, we control for the age of the patent up to 2016 and the inventor's patent stock.

Table 3.8: Return Migrants vs. Non-migrants productivity, Log Annual Number of Claims

	(1) Log (1 + $NCLA_{it}$)	(2) Log (1 + $NCLA_{it}$)	(3) Log (1 + $NCLA_{it}$)	(4) Log (1 + $NCLA_{it}$)
$RETURN_i$	0.0723^{**} (0.0340)			
$CMOV_{it}$	0.0460^{***} (0.0115)	0.0455^{***} (0.0115)	0.0454^{***} (0.0115)	0.0469^{***} (0.0115)
$RETURN\_WORK_i$		0.170^{***} (0.0522)	0.168^{***} (0.0528)	
$RETURN\_EDUC_i$		0.0145 (0.0380)		
$EDUC\_EU_i$			0.0203 (0.0508)	
$EDUC\_OTHER_i$			0.0221 (0.0577)	
$EDUC\_US_i$			0.0331 (0.0841)	
$RETURN\_PAT_i$				0.259^{***} (0.0665)
$RETURN\_NOPAT_i$				0.0265 (0.0761)
Nbr. observations	14077	14077	14077	14077
Nbr. inventors	3372	3372	3372	3372
R-squared	0.0630	0.0631	0.0635	0.0638
Controls	YES	YES	YES	YES

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table estimates the effect of being a return migrant on inventors' productivity with different combinations of characteristics. Coefficient with reported marginal effects. Standard errors appear in parenthesis and are clustered at the inventor level. Table 3.8 shows the effect on the total number of claims ($NCLA_{it}$) produced yearly per inventor. All columns are estimated with a random effect GLS model. Besides the listed variables in each panel, we control for all specifications by country of origin and time fixed effect; position in the company ($POSITION_{it}$); gender ($GEND_i$); cohort (COH_i); labour market experience ($LABEXP_{it}$) and education quantity ($EDLEV_i$) and quality ($EDQUAL_i$). Furthermore, we control for the inventor's patent stock.

3.5.2 Accounting for selection at destination

The last statement of section 3.3.3.1 suggests that, while at destination, some return migrants are more negatively selected than others. For this reason, we propose four additional controls to capture the endogeneity due to the positive and negative selection occurring at destination. In fact, within the return migrants who entered in the host labour market, some are more productive than others and stay in the host labour market longer. As Table 3.6, 3.7 and 3.8, Column 4 suggest, return migrants who enter in the host labour market and succeed to patent, perform better (number of patents, citations and claims) than both the non-migrants and the return migrants who did not patent in the host labour market, this results are confirmed by Appendix A, Table 3.A.1, 3.A.2 and 3.A.3. Furthermore, in Chapter 2, we have shown that the positively selected migrants are more productive than both the other migrants at destination and the non-migrants. Hence, to account for positive selection we control for the return migrants' productivity ($HOST\_PATENT_i$, $HOST\_CITATION_i$, $HOST\_CLAIM_i$) while in the host country.

For negative selection, we control for the time spent on the host labour market ($HOST\_LABOUR_i$). We assume that the more (less) a migrant stay at destination, the more (less) he/she is productive since the host country tries to retain the best (similar findings in the literature strengthen the previous statement (Breschi, et al., 2018)). For this matter, we control for negative selection using the time spent on the host labour market. Appendix A, Figure 3.A.4 shows the descriptive statistics of the productivity and length of stay in the destination country. However, we would like to stress that our proxy for negative selection is not perfect. If one may think that migrants that early leave the host country to come back home are the less productive ones, the migrants that stay longer may not even be negatively selected at all. Personal motives may drive the decision to leave the destination country rather than any failure at destination.

Table 3.9: Return Migrants vs. Non-migrants productivity, Self-selection, Log Annual Number of Patents

	(1) Log (1 + $NPAT_{it}$)	(2) Log (1 + $NPAT_{it}$)	(3) Log (1 + $NPAT_{it}$)	(4) Log (1 + $NPAT_{it}$)
$RETURN_i$	0.020 (0.0218)	0.026 (0.0227)		
$CMOV_{it}$	0.0115* (0.00599)	0.0115* (0.00596)	0.0112* (0.00597)	0.0106* (0.00594)
$RETURN\_WORK_i$			0.072* (0.037)	0.137*** (0.042)
$RETURN\_EDUC_i$			-0.011 (0.023)	-0.010 (0.023)
$HOST\_PATENT_i$	0.009 (0.006)	0.009 (0.006)	0.008 (0.006)	0.009 (0.006)
$HOST\_CITATION_i$	0.0001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)
$HOST\_CLAIM_i$	-0.0002 (0.0004)	-0.0002 (0.0004)	-0.0002 (0.0004)	-0.0001 (0.0004)
$HOST\_LABOUR_i$		-0.001 (0.003)		-0.006** (0.003)
Nbr. observations	14077	14077	14077	14077
Nbr. inventors	3372	3372	3372	3372
R-squared	0.066	0.066	0.067	0.068
Controls	YES	YES	YES	YES
Positive selection	YES	YES	YES	YES
Negative selection	NO	YES	NO	YES

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table estimates the effect of being a return migrant on inventors' productivity with different combinations of characteristics. Coefficient with reported marginal effects. Standard errors appear in parenthesis and are clustered at the inventor level. Table 3.9 shows the effect on the number of patents ($NPAT_{it}$) produced yearly per inventor. All columns are estimated with a random effect GLS model. Besides the listed variables in each panel, we control for all specifications by country of origin and time fixed effect; position in the company ($POSITION_{it}$); gender ($GEND_i$); cohort (COH_i); labour market experience ($LABEXP_{it}$) and education quantity ($EDLEV_i$) and quality ($EDQUAL_i$). To account for positive selection, we control for the number of patents made at destination ($HOST\_PATENT_i$), the number of citations received of these patents ($HOST\_CITATION_i$) and the number of claims of these patents ($HOST\_CLAIM_i$). To account for negative selection, we control for the length of stay on the host labour market ($HOST\_LABOUR_i$).

Table 3.10: Return Migrants vs. Non-migrants productivity, Self-selection, Log Annual Number of Citations

	(1)	(2)	(3)	(4)
	Log (1 + $NCIT_{it}$)	Log (1 + $NCIT_{it}$)	Log (1 + $NCIT_{it}$)	Log (1 + $NCIT_{it}$)
$RETURN_i$	0.073	0.074		
	(0.057)	(0.061)		
$CMOV_{it}$	0.044***	0.044***	0.043***	0.042***
	(0.0156)	(0.0156)	(0.0156)	(0.0155)
$RETURN\_WORK_i$			0.207**	0.239***
			(0.091)	(0.114)
$RETURN\_EDUC_i$			-0.009	-0.008
			(0.063)	(0.063)
$HOST\_PATENT_i$	0.0122	0.0123	0.009	0.0118
	(0.0126)	(0.0136)	(0.0124)	(0.0132)
$HOST\_CITATION_i$	0.001***	0.001***	0.001***	0.001***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
$HOST\_CLAIM_i$	-0.001	-0.001	-0.001	-0.001
	(0.0008)	(0.0008)	(0.0008)	(0.0008)
$HOST\_LABOUR_i$		-0.0002		-0.012
		(0.007)		(0.007)
Nbr. observations	14077	14077	14077	14077
Nbr. inventors	3372	3372	3372	3372
R-squared	0.069	0.069	0.070	0.070
Controls	YES	YES	YES	YES
Positive selection	YES	YES	YES	YES
Negative selection	NO	YES	NO	YES

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table estimates the effect of being a return migrant on inventors' productivity with different combinations of characteristics. Coefficient with reported marginal effects. Standard errors appear in parenthesis and are clustered at the inventor level. Table 3.10 shows the effect on the total number of citations ($NCIT_{it}$) received yearly per inventor. All columns are estimated with a random effect GLS model. Besides the listed variables in each panel, we control for all specifications by country of origin and time fixed effect; position in the company ($POSITION_{it}$); gender ($GEND_i$); cohort (COH_i); labour market experience ($LABEXP_{it}$) and education quantity ($EDLEV_i$) and quality ($EDQUAL_i$). Furthermore, we control for the age of the patent up to 2016 and the inventor's patent stock. To account for positive selection, we control for the number of patents made at destination ($HOST\_PATENT_i$), the number of citations received of these patents ($HOST\_CITATION_i$) and the number of claims of these patents ($HOST\_CLAIM_i$). To account for negative selection, we control for the length of stay on the host labour market ($HOST\_LABOUR_i$).

Table 3.11: Return Migrants vs. Non-migrants productivity, Self-selection, Log Annual Number of Claims

	(1) Log (1 + $NCLA_{it}$)	(2) Log (1 + $NCLA_{it}$)	(3) Log (1 + $NCLA_{it}$)	(4) Log (1 + $NCLA_{it}$)
$RETURN_i$	0.032 (0.035)	0.011 (0.037)		
$CMOV_{it}$	0.041^{***} (0.0114)	0.041^{***} (0.0115)	0.041^{***} (0.0115)	0.041^{***} (0.0115)
$RETURN\_WORK_i$			0.094 (0.058)	0.068 (0.076)
$RETURN\_EDUC_i$			0.018 (0.038)	0.018 (0.038)
$HOST\_PATENT_i$	-0.008 (0.007)	-0.011 (0.007)	-0.011 (0.007)	-0.011 (0.007)
$HOST\_CITATION_i$	0.0002 (0.0003)	0.0002 (0.0003)	0.0002 (0.0003)	0.0002 (0.0003)
$HOST\_CLAIM_i$	0.002^{***} (0.001)	0.002^{***} (0.001)	0.002^{***} (0.001)	0.002^{***} (0.001)
$HOST\_LABOUR_i$		0.005 (0.004)		0.002 (0.005)
Nbr. observations	14077	14077	14077	14077
Nbr. inventors	3372	3372	3372	3372
R-squared	0.165	0.165	0.165	0.165
Controls	YES	YES	YES	YES
Positive selection	YES	YES	YES	YES
Negative selection	NO	YES	NO	YES

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table estimates the effect of being a return migrant on inventors' productivity with different combinations of characteristics. Coefficient with reported marginal effects. Standard errors appear in parenthesis and are clustered at the inventor level. Table 3.11 shows the effect on the total number of claims ($NCLA_{it}$) produced yearly per inventor. All columns are estimated with a random effect GLS model. Besides the listed variables in each panel, we control for all specifications by country of origin and time fixed effect; position in the company ($POSITION_{it}$); gender ($GEND_i$); cohort (COH_i); labour market experience ($LABEXP_{it}$) and education quantity ($EDLEV_i$) and quality ($EDQUAL_i$). Furthermore, we control for the inventor's patent stock. To account for positive selection, we control for the number of patents made at destination ($HOST\_PATENT_i$), the number of citations received of these patents ($HOST\_CITATION_i$) and the number of claims of these patents ($HOST\_CLAIM_i$). To account for negative selection, we control for the length of stay on the host labour market ($HOST\_LABOUR_i$).

The results show that, when we control for these proxies of the selection mechanism, return migrants and non-migrants are equally productive, both in terms of the number of patents (Table 3.9, Column 1 and 2), citations (Table 3.10, Column 1 and 2), and claims (Table 3.11, Column 1 and 2). When we break down return migrants between those who experience at least one year in the host labour market and those who studied abroad, we observe intriguing results. When we first control for positive selection, we find a decrease in the coefficient and the level of significance for the labour market return migrant's variable ($RETURN\_WORK_i$). This shows that the positive selection mechanism can partially explain the difference in productivity between return migrants and non-migrants. When we control for positive selection using the proxies for productivity at destination, we still observe a positive and significant at 10% productivity difference for return migrants with working experience at destination on the number of patents produced of about 7.2%. For the number of citations received, we still observe a positive and significant coefficient at 5%, for return migrants with working experience at destination, the average number of citations received when back home is 20% higher than non-migrants. We, nevertheless, obtain no significant differences when using the number of claims as dependent variable. When we account for both positive and negative selection (Table 3.9, 3.10 and 3.11, Column 4), we find return migrants with an experience on the host labour market are more productive than both non-migrants and return migrants with an international education only. The coefficients, both significant at the 1% level, suggest a difference of 13,7% in number of patents and 23.9% in number of citations received. These last results confirm the previous ones, showing that working experience seems to matter more than education at destination.

Table 3.12: Return Migrants vs. Non-migrants productivity, Self-selection, Log Annual Number of Patents and Number of Patents

	(1) Log (1 + $NPAT_{it}$)	(2) Log (1 + $NPAT_{it}$)	(3) $NPAT_{it}$	(4) $NPAT_{it}$
$CMOV_{it}$	0.011^{*} (0.006)	0.011^{*} (0.006)	1.024^{**} (0.013)	1.024^{**} (0.122)
$RETURN\_EDUC_i$	-0.010 (0.023)	-0.009 (0.023)	0.996 (0.042)	0.996 (0.042)
$RETURN\_PAT_i$	0.105^{**} (0.049)	0.215^{***} (0.071)	1.266^{***} (0.086)	1.439^{***} (0.150)
$RETURN\_NOPAT_i$	0.062^{*} (0.038)	0.105^{***} (0.040)	1.130 (0.087)	1.188^{**} (0.099)
HOST_CITATION	0.0002 (0.0002)	0.0002 (0.0002)	1.0002 (0.0002)	1.0002 (0.0002)
HOST_CLAIM	0.0002 (0.0004)	0.0002 (0.0004)	1.0003 (0.004)	1.0004 (0.005)
HOST_LABOUR		-0.008^{**} (0.003)		0.991 (0.005)
Nbr. observations	14077	14077	14077	14077
Nbr. inventors	3372	3372	3372	3372
R-squared	0.066	0.067		
Controls	YES	YES	YES	YES
Positive selection	YES	YES	YES	YES
Negative selection	NO	YES	NO	YES
LR test			1589^{***}	1578^{***}

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table estimates the effect of being a return migrant with or without a patent abroad on inventors' productivity with different combinations of characteristics. Coefficient with reported marginal effects for Column 1 and 2; and with incidence ratio rate for Column 3 and 4. Standard errors appear in parenthesis and are clustered at the inventor level. Table 3.12 shows the effect on the number of patents ($NPAT_{it}$) produced yearly per inventor. All columns are estimated with a random effect GLS model column 1 and 2, and, with a negative binomial model column 3 and 4. Besides the listed variables in each panel, we control for all specifications by country of origin and time fixed effect; position in the company ($POSITION_{it}$); gender ($GEND_i$); cohort (COH_i); labour market experience ($LABEXP_{it}$) and education quantity ($EDLEV_i$) and quality ($EDQUAL_i$). To account for positive selection, we control for the number of citations received of these patents ($HOST\_CITATION_i$) and the number of claims of these patents ($HOST\_CLAIM_i$). To account for negative selection, we control for the length of stay on the host labour market ($HOST\_LABOUR_i$).

Table 3.13: Return Migrants vs. Non-migrants productivity, Self-selection, Log Annual Number of Citations and Number of Citations

	(1) Log (1 + $NCIT_{it}$)	(2) Log (1 + $NCIT_{it}$)	(3) $NCIT_{it}$	(4) $NCIT_{it}$
$CMOV_{it}$	0.044***	0.043***	1.051***	1.048***
	(0.016)	(0.015)	(0.012)	(0.011)
$RETURN\_EDUC_i$	-0.007	-0.005	1.005	1.005
	(0.063)	(0.063)	(0.038)	(0.037)
$RETURN\_PAT_i$	0.304**	0.553***	1.345***	1.495***
	(0.126)	(0.195)	(0.083)	(0.140)
$RETURN\_NOPAT_i$	0.127	0.223**	1.184**	1.236***
	(0.098)	(0.108)	(0.081)	(0.094)
HOST_CITATION	0.0013***	0.0013***	1.0005***	1.0005***
	(0.0004)	(0.0002)	(0.0002)	(0.0002)
HOST_CLAIM	-0.0008	-0.0006	0.9997	0.9998
	(0.0009)	(0.0008)	(0.0005)	(0.0003)
HOST_LABOUR		-0.017*		0.992
		(0.009)		(0.005)
Nbr. observations	14077	14077	14077	14077
Nbr. inventors	3372	3372	3372	3372
R-squared	0.069	0.070	0.070	0.070
Controls	YES	YES	YES	YES
Positive selection	YES	YES	YES	YES
Negative selection	NO	YES	NO	YES
LR test			628***	624***

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table estimates the effect of being a return migrant with or without a patent abroad on inventors' productivity with different combinations of characteristics. Coefficient with reported marginal effects for Column 1 and 2; and with incidence ratio rate for Column 3 and 4. Standard errors appear in parenthesis and are clustered at the inventor level. Table 3.13 the effect on the total number of citations ($NCIT_{it}$) received yearly per inventor. All columns are estimated with a random effect GLS model column 1 and 2, and, with a negative binomial model column 3 and 4. Besides the listed variables in each panel, we control for all specifications by country of origin and time fixed effect; position in the company ($POSITION_{it}$); gender ($GEND_i$); cohort (COH_i); labour market experience ($LABEXP_{it}$) and education quantity ($EDLEV_i$) and quality ($EDQUAL_i$). Furthermore, we control for the age of the patent up to 2016. To account for positive selection, we control for the number of citations received of these patents ($HOST\_CITATION_i$) and the number of claims of these patents ($HOST\_CLAIM_i$). To account for negative selection, we control for the length of stay on the host labour market ($HOST\_LABOUR_i$).

Table 3.14: Return Migrants vs. Non-migrants productivity, Self-selection, Log Annual Number of Claims and Number of Claims

	(1) Log (1 + $NCLA_{it}$)	(2) Log (1 + $NCLA_{it}$)	(3) $NCLA_{it}$	(4) $NCLA_{it}$
$CMOV_{it}$	0.042***	0.042***	1.078***	1.078***
	(0.011)	(0.011)	(0.020)	(0.019)
$RETURN\_EDUC_i$	0.018	0.18	1.066	1.066
	(0.039)	(0.038)	(0.059)	(0.058)
$RETURN\_PAT_i$	0.126*	0.155	1.556***	1.565***
	(0.071)	(0.111)	(0.131)	(0.195)
$RETURN\_NOPAT_i$	0.004	0.015	1.104	1.107
	(0.077)	(0.082)	(0.127)	(0.136)
HOST_CITATION	-0.0001	-0.0001	0.9997	0.9997
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
HOST_CLAIM	0.0019**	0.0019***	1.002***	1.0018***
	(0.0007)	(0.0004)	(0.0005)	(0.005)
HOST_LABOUR		-0.002		0.999
		(0.006)		(0.058)
Nbr. observations	14077	14077	14077	14077
Nbr. inventors	3372	3372	3372	3372
R-squared	0.161	0.165	0.165	0.165
Controls	YES	YES	YES	YES
Positive selection	YES	YES	YES	YES
Negative selection	NO	YES	NO	YES
LR test			943***	940***

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table estimates the effect of being a return migrant with or without a patent abroad on inventors' productivity with different combinations of characteristics. Coefficient with reported marginal effects for Column 1 and 2; and with incidence ratio rate for Column 3 and 4. Standard errors appear in parenthesis and are clustered at the inventor level. Table 3.14 shows the effect on the total number of claims ($NCLA_{it}$) produced yearly per inventor. All columns are estimated with a random effect GLS model column 1 and 2, and, with a negative binomial model column 3 and 4. Besides the listed variables in each panel, we control for all specifications by country of origin and time fixed effect; position in the company ($POSITION_{it}$); gender ($GEND_i$); cohort (COH_i); labour market experience ($LABEXP_{it}$) and education quantity ($EDLEV_i$) and quality ($EDQUAL_i$). Furthermore, we control for the inventor's patent stock. To account for positive selection, we control for the number of citations received of these patents ($HOST\_CITATION_i$) and the number of claims of these patents ($HOST\_CLAIM_i$). To account for negative selection, we control for the length of stay on the host labour market ($HOST\_LABOUR_i$).

In Table 3.12, 3.13 and 3.14, we control for positive and negative selection focusing on return migrants who have experience in the host labour market and decompose them between those who patented abroad and those who did not. We show that, when we control for positive selection only (column 1 and 3 of each Table), we obtain a positive and significant difference in productivity for the return migrants with at least one patent at destination. Even when we control for negative selection (Table 3.12, 3.13 and 3.14, Column 2 and 4), we observe that the inventors' return migrants who patented abroad are more productive than both the non-migrants and the inventors that did not succeed to patent in their former host country. In particular, we find that, compared to non-migrants, they produce from 21% to 44% more patents[29], and they receive from 55% to 49% more citations[30].

[29] Without controlling for selection, we found 17% (Table 3.6, Column 4) to 39% (Appendix A Table 3.A.1, Column 4).
[30] Without controlling for selection, we found 26% (Table 3.7, Column 4) to 43% (Appendix A Table 3.A.2, Column 4).

These last results suggest that, even when we control for positive and negative selection, return migrants that succeeded to make patents abroad are those contributing the most to the brain gain when back home.

3.6 Conclusion

In this Chapter we analyze the effect of return migration on inventors' productivity at origin by breaking down the flow of return migrants by migration reasons, between education and work.

In our analysis, we consider a double mechanism of selection. First, when they leave the country of origin, return migrants are positively selected, since more productive than stayers and the natives at destination, as Chapter 2 of the present dissertation shows. Then, when they return home, a second mechanism of selection is in play, since return migrants might be the ones who did not succeed in the destination country, being less productive than migrants who stay in the host country and the natives there.

When we compare return migrants to non-migrants, we find a strong and significant difference in productivity between the two groups, with return migrants being more productive than non-migrants. This result suggests the existence of a brain gain effect for the home country, as the returning workforce is more productive and may consistently contribute to the innovation system of the country. Nevertheless, it is not clear whether the return migrants increase their productivity because of the experience abroad, otherwise there is only a "temporary brain loss" for the home country while its workforce is abroad.

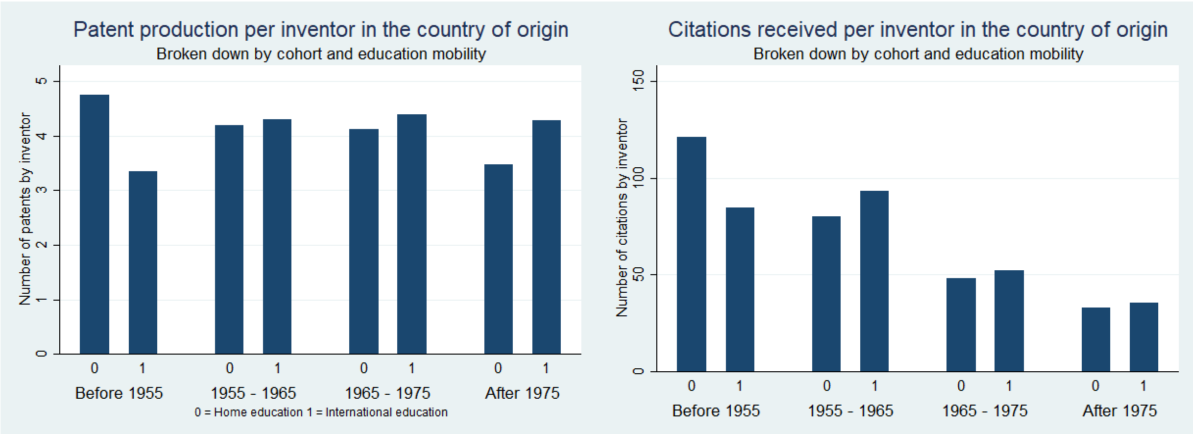
This result does not depend on the education that return migrants obtained abroad. In fact, we have also shown that returnees who have studied in a more prestigious university and attained a higher degree abroad are more skilled than those with home education and work experience only. Nevertheless, when we decompose the return migrant group between those who have an international education and those who have a work experience abroad, having studied in another country does not bring any productivity premium when being measured by the number of patents, citations, and claims, even when disentangling by country of education attainment.

However, the composition of the return migrant flow seems to be more complex. By decomposing it between return migrants who patented in the destination country and those who did not, we find that the formers are far more productive than the latter, whether return migrants or non-migrants. Moreover, we also find that, in the absence of patents abroad, differences between return migrants and non-migrants are non-significant. This last result shows that some migrants are more negatively selected than others, explaining the productivity difference within the population of return migrants. Furthermore, the possible brain gain for the home country is also conditional to the population that is returning. For this reason, we control for the length of stay and the productivity made at destination by the return migrant who experienced at least one year on the host labour market. When doing so, we strengthen the previous results by finding that return migrants with work experience in the destination country are more productive than the international students.

3.7 Appendix

3.7.1 Appendix A: Additional descriptive statistics

Figure 3.A.1: International/Home education productivity by cohort



The cohort year on the -axis represents the inventor's estimated year of birth. We estimated the year of birth based on the starting year of the first degree given by the inventor.

Figure 3.A.2: International/Home education productivity by education

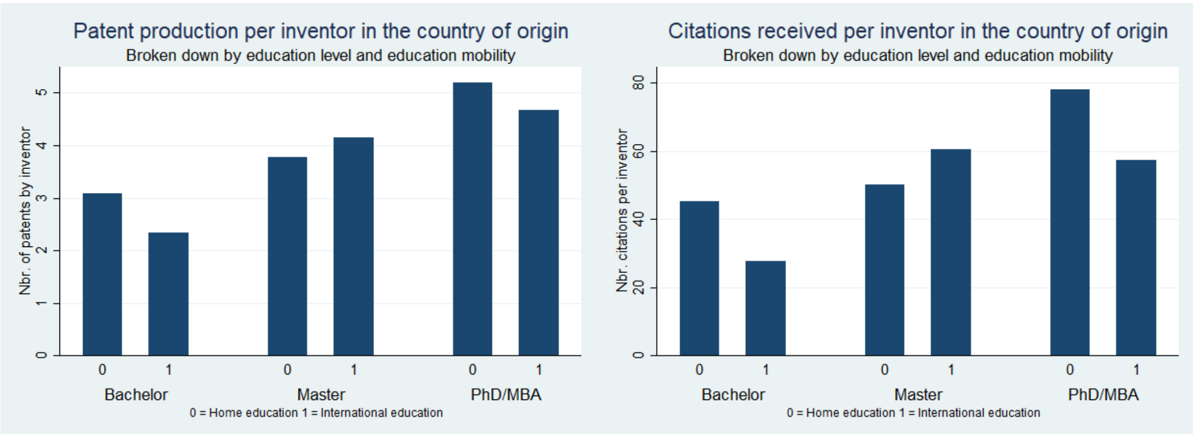


Figure 3.A.3: Dependent variables distribution

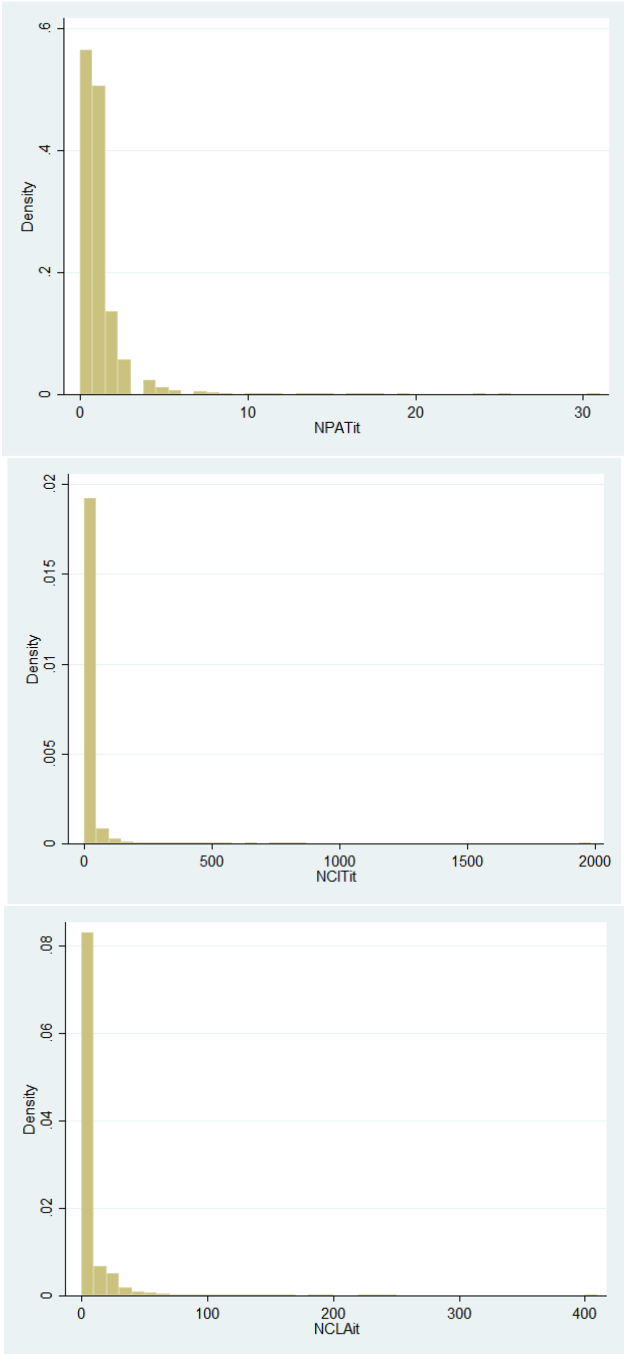


Figure 3.A.4: Return migrant productivity and length of stay, decomposed by experience at destination

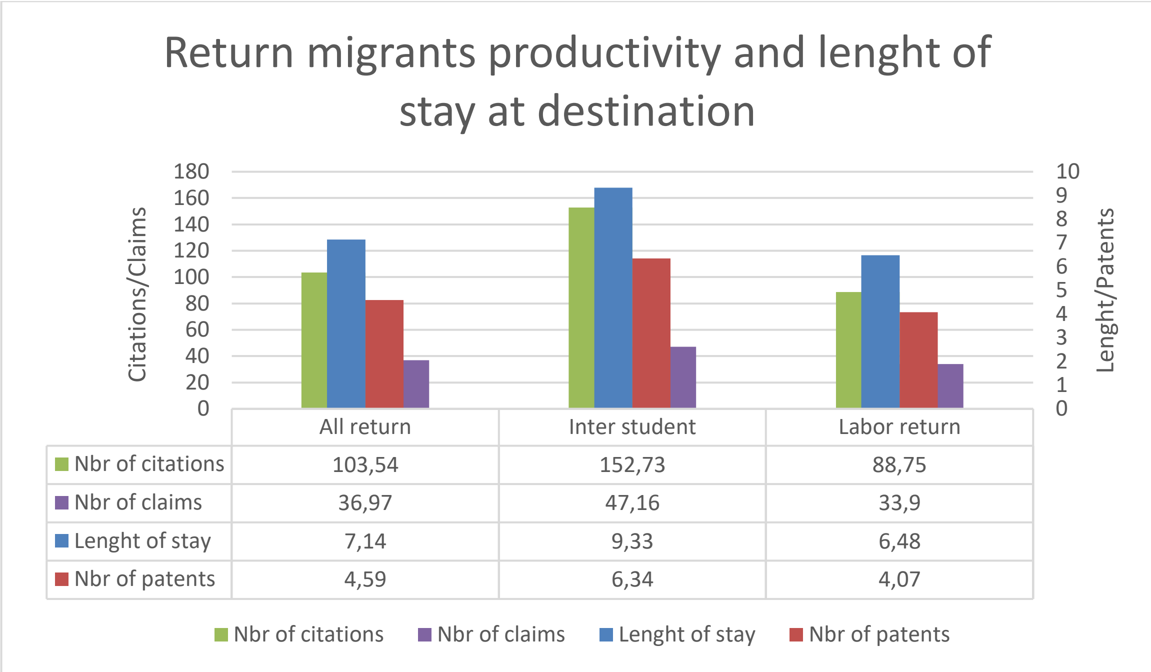


Table 3.A.1: Return Migrants vs. Non-migrants productivity, Number of Patents

	(1) $NPAT_{it}$	(2) $NPAT_{it}$	(3) $NPAT_{it}$	(4) $NPAT_{it}$
$RETURN_i$	1.120^{***} (0.040)			
$CMOV_{it}$	1.023^{*} (0.012)	1.024^{**} (0.012)	1.024^{**} (0.00596)	1.025^{**} (0.012)
$RETURN\_WORK_i$		1.296^{***} (0.063)	1.297^{***} (0.063)	
$RETURN\_EDUC_i$		0.995 (0.042)		
$EDUC\_EU_i$			1.044 (0.058)	
$EDUC\_OTHER_i$			0.928 (0.059)	
$EDUC\_US_i$			0.947 (0.083)	
$RETURN\_PAT_i$				1.394^{***} (0.082)
$RETURN\_NOPAT_i$				1.138^{*} (0.088)
Nbr. observations	14077	14077	14077	14077
Nbr. inventors	3372	3372	3372	3372
LR test	1583***	1581***	1579***	1574***
Wald test	387***	391***	393***	397***
Controls	YES	YES	YES	YES

$^{*}\ p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

This table estimates the effect of being a return migrant on inventors' productivity with different combinations of characteristics. Coefficient with reported incidence rate ratio. Standard errors appear in parenthesis and are clustered at the inventor level. Panel A shows the effect on the number of patents ($NPAT_{it}$) produced yearly per inventor. Panel B shows the effect on the total number of citations ($NCIT_{it}$) received yearly per inventor, and Panel C shows the effect on the total number of claims ($NCLA_{it}$) produced yearly per inventor. All Panels are estimated with a random effect Negative Binomial model to account for data overdispersion. Besides the listed variables in each panel, we control for all specifications by country of origin and time fixed effect; position in the company ($POSITION_{it}$); gender ($GEND_i$); cohort (COH_i); labour market experience ($LABEXP_{it}$) and education quantity ($EDLEV_i$) and quality ($EDQUAL_i$). Furthermore, for Panel B, we control for the age of the patent up to 2016 and the patent stock; for Panel C, we control only for the patent stock.

Table 3.A.2: Return Migrants vs. Non-migrants productivity, Number of citations

	(1) $NCIT_{it}$	(2) $NCIT_{it}$	(3) $NCIT_{it}$	(4) $NCIT_{it}$
$RETURN_i$	1.110^{**} (0.035)			
$CMOV_{it}$	1.156^{***} (0.020)	1.157^{***} (0.059)	1.160^{***} (0.020)	1.156^{***} (0.012)
$RETURN\_WORK_i$		1.256^{***} (0.063)	1.250^{***} (0.090)	
$RETURN\_EDUC_i$		1.044 (0.059)		1.050 (0.044)
$EDUC\_EU_i$			1.191^{**} (0.058)	
$EDUC\_OTHER_i$			0.977 (0.083)	
$EDUC\_US_i$			0.787^{**} (0.088)	
$RETURN\_PAT_i$				1.431^{***} (0.087)
$RETURN\_NOPAT_i$				1.232^{**} (0.105)
Nbr. observations	14077	14077	14077	14077
Nbr. inventors	3372	3372	3372	3372
LR test	474***	476***	476***	475***
Wald	1015***	1043***	1043***	1044***
Controls	YES	YES	YES	YES

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table estimates the effect of being a return migrant on inventors' productivity with different combinations of characteristics. Coefficient with reported incidence rate ratio. Standard errors appear in parenthesis and are clustered at the inventor level. Panel A shows the effect on the number of patents ($NPAT_{it}$) produced yearly per inventor. Panel B shows the effect on the total number of citations ($NCIT_{it}$) received yearly per inventor, and Panel C shows the effect on the total number of claims ($NCLA_{it}$) produced yearly per inventor. All Panels are estimated with a random effect Negative Binomial model to account for data overdispersion. Besides the listed variables in each panel, we control for all specifications by country of origin and time fixed effect; position in the company ($POSITION_{it}$); gender ($GEND_i$); cohort (COH_i); labour market experience ($LABEXP_{it}$) and education quantity ($EDLEV_i$) and quality ($EDQUAL_i$). Furthermore, for Panel B, we control for the age of the patent up to 2016 and the patent stock; for Panel C, we control only for the patent stock.

Table 3.A.3: Return Migrants vs. Non-migrants productivity, Number of claims

	(1) $NCLA_{it}$	(2) $NCLA_{it}$	(3) $NCLA_{it}$	(4) $NCLA_{it}$
$RETURN_i$	0.997 (0.089)			
$CMOV_{it}$	1.260*** (0.041)	1.267*** (0.041)	1.266*** (0.041)	1.268*** (0.041)
$RETURN\_WORK_i$		1.347** (0173)	1.376** (0.018)	
$RETURN\_EDUC_i$		0.847 (0.089)		
$EDUC\_EU_i$			0.875 (0.133)	
$EDUC\_OTHER_i$			0.974 (0.151)	
$EDUC\_US_i$			0.729 (0.152)	
$RETURN\_PAT_i$				1.543*** (0.244)
$RETURN\_NOPAT_i$				0.949 (0.195)
Nbr. observations	14077	14077	14077	14077
Nbr. inventors	3372	3372	3372	3372
LR test	791***	791***	832***	832***
Wald	834***	834***	895***	896***
Controls	YES	YES	YES	YES

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table estimates the effect of being a return migrant on inventors' productivity with different combinations of characteristics. Coefficient with reported incidence rate ratio. Standard errors appear in parenthesis and are clustered at the inventor level. Panel A shows the effect on the number of patents ($NPAT_{it}$) produced yearly per inventor. Panel B shows the effect on the total number of citations ($NCIT_{it}$) received yearly per inventor, and Panel C shows the effect on the total number of claims ($NCLA_{it}$) produced yearly per inventor. All Panels are estimated with a random effect Negative Binomial model to account for data overdispersion. Besides the listed variables in each panel, we control for all specifications by country of origin and time fixed effect; position in the company ($POSITION_{it}$); gender ($GEND_i$); cohort (COH_i); labour market experience ($LABEXP_{it}$) and education quantity ($EDLEV_i$) and quality ($EDQUAL_i$). Furthermore, for Panel B, we control for the age of the patent up to 2016 and the patent stock; for Panel C, we control only for the patent stock.

Appendix A.4: Ranked University building

To rank to universities, we use the QS World University Rankings[31] that allows us to decompose the ranking by specialization (Computer Science, Electrical/Electronic, Mechanic, Mathematics…). Hence, based on the specialty and the university ranking by specialization, we create a categorical variable that resums the education quality as follow:

Table 3.A.4: Inventor A, resume reporting education history

School	Degree	Degree category $EDTYPE_i$	University ranked in Electronic	Education quality $EDQUAL_i$
University of Twente, Enchede, The Netherlands	Doctor of Philosophy (Ph.D.), Electronics	Electronic	151-200	4 (5)*
Eindhoven University of Technology	MEEE, Electrical, Electronics, and Communications Engineering	Electronic	101-150	4 (4)*

*The value in parenthesis is the value associated with the University ranked, we are keeping the best ranking for the education quality variable, here 4.

Finally, the EDQUAL variable is decomposed into 12 categories; taking 1 if the university is ranked between 1 and 20; 2 between 21 and 50; taking one additional unit every 50 ranks up to 500, finally the EDQUAL variable takes 12 if the university is not ranked before 500 in the specialization did by the inventor.

[31] https://www.topuniversities.com/qs-world-university-rankings/methodology

Appendix A.5: Classify the positions

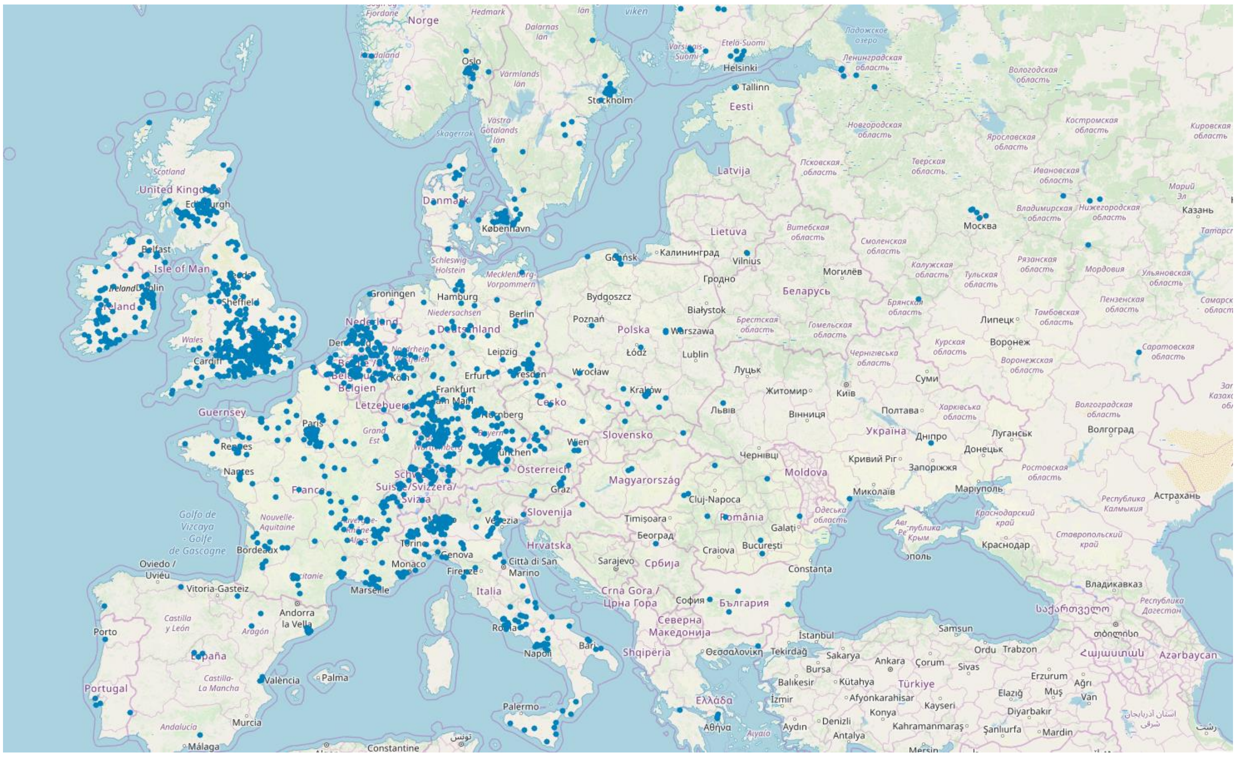
To classify the different titles given by the labour market history, we use text analysis and organize the different titles into 6 Positions. Founder refers to inventors that created or co-created a company sometimes called partners too. Manager refers to inventors that manage a team, called themselves staff, manager, chief team. Company's head relates to inventors that are director, president, CEO, chairman of the board. Engineer refers to inventors that called themselves as such. Scientist refers to inventors that are doing research. Finally, if none of these categories were assigned to one inventor at time t, we classify him as Others.

Table 3.A.5: Inventor A, resume reporting labour market history

Title	Employer	$Position_{it}$	Period
Senior Research Scientist	Philips	Scientist	2000-2007
Research Staff Member	IBM	Scientist	2008-2014

3.7.2 Appendix B: Some maps

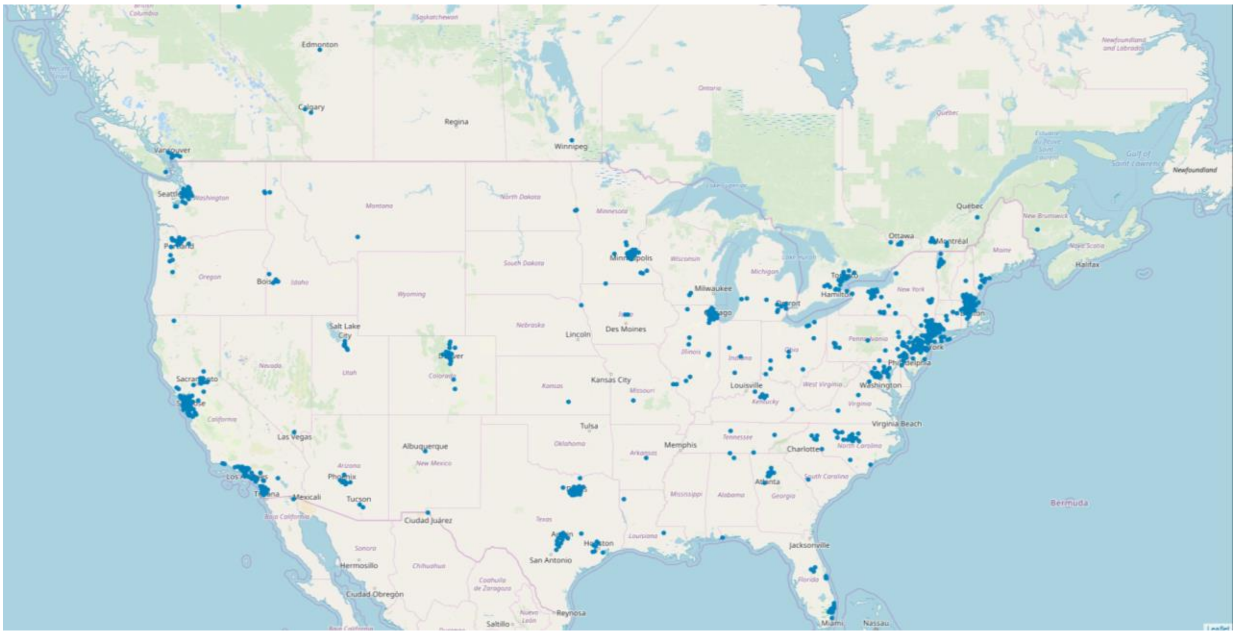
Figure 3.B.1: Patents' localization made by European inventors in Europe



The blue points are the Europeans localization when they have made a patent.

(Source: Author's calculation from the Linked Inventor database)

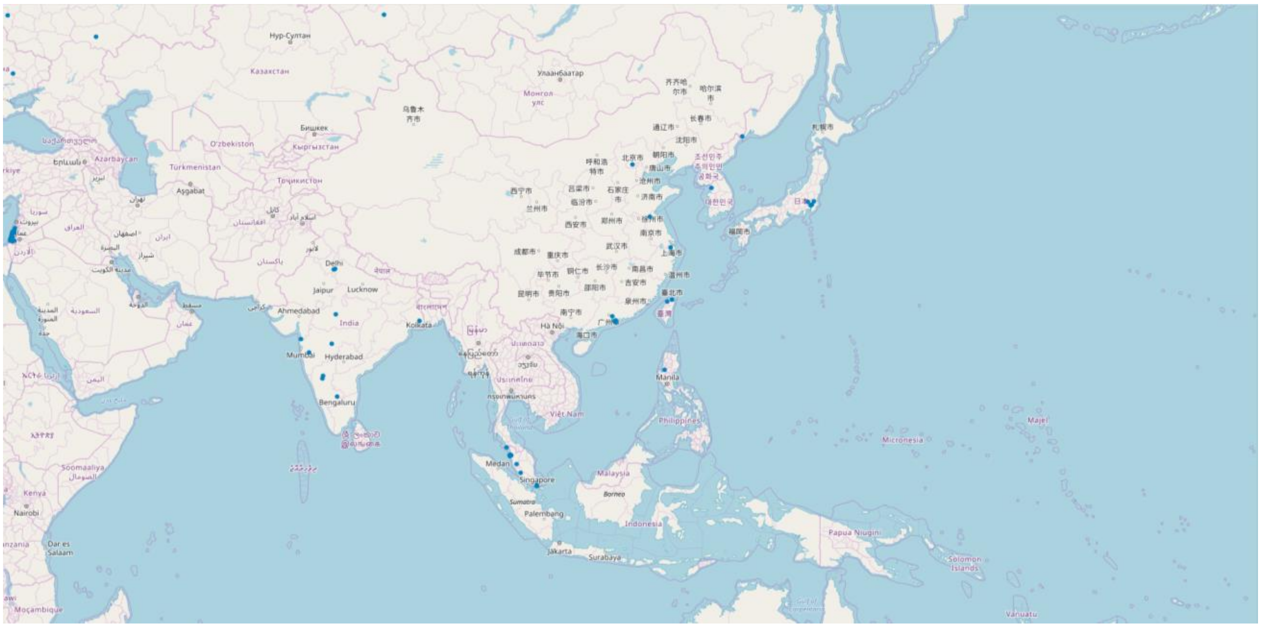
Figure 3.B.2: Patents' localization made by European inventors in North America



The blue points are the Europeans localization when they have made a patent.

(Source: Author's calculation from the Linked Inventor database)

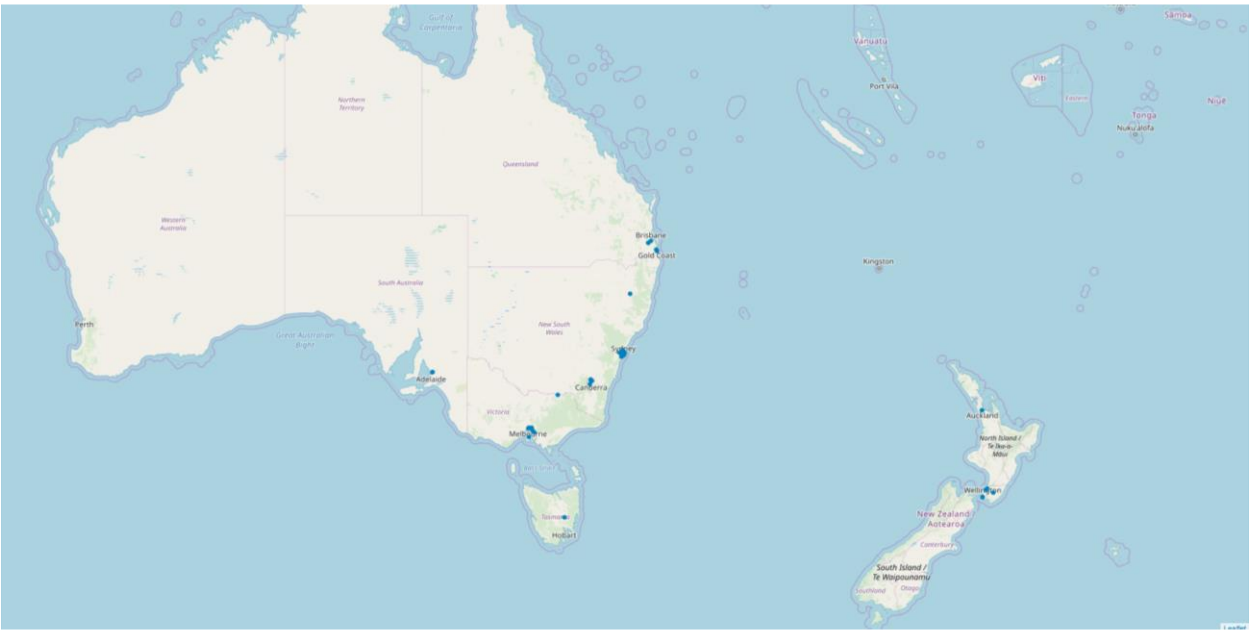
Figure 3.B.3: Patents' localization made by European inventors in the Middle East and Asia



The blue points are the Europeans localization when they have made a patent.

(Source: Author's calculation from the Linked Inventor database)

Figure 3.B.4: Patents' localization made by European inventors in Oceania



The blue points are the Europeans localization when they have made a patent.

(Source: Author's calculation from the Linked Inventor database)

Conclusion

This dissertation explores the different inventors' experiences and the impact on their productivity. Thanks to a new, original database, we foster our understanding of the productivity difference between inventors based on their migration experience or lack thereof, in both the host and home countries. By focusing on company mobility and migration, we also provide first evidence of European return migrants' contribution in their country of origin, controlling for positive and negative selection.

Overall, we show the importance of migration for innovation in the destination and origin countries through the return of the migrant population. We also stress the importance of the entry channel into the destination country, finding the educational and the within-firm mobility as the most important drivers for positive selection. Differently from the previous literature on company's mobility, we observe an opposite sign for the difference in the average productivity "at destination" between migrants and natives. When we extend the period of analysis before the first and after the last patent filed, we find the inventors that changed of company during their patenting period to be the less productive ones, while the inventors that changed of company before filing their first patent being the most productive ones. We also show that international education alone does not bring any productivity premium when the inventor returns in his home country, while working experience on the host labour market does. Also, we find that some return migrants are more negatively selected than others at destination. In fact, we observe a significant productivity gap across return migrants between the ones that succeeded to file at least one patent in the host country and the ones that did not. Finally, we find heterogeneity for the productivity at destination between the migrants that entered the United States under the most restrictive period of the H1-B visa policy and the ones that entered the US before the policy implementation or during a less restrictive one. From this last result, we show that reducing the flow of high-skilled migrants has a positive impact on the productivity of the most selected ones, migrants entering the US during the most restrictive period of the H1-B visa policy, compared to the natives and the other migrants. Nevertheless, during a restrictive period, the destination country has a smaller number of skilled individuals that can enter and contribute positively to the innovation. Hence, if we can conclude on the impact of the H1-B visa policy at the individual level, we cannot discuss the overall effect on the innovation at the country one, a macroeconomic approach should be favoured.

So far, we just scratched the surface of this database, and its potential to improve our understanding of the nexus between migration and innovation seems vast. The primary findings show a great potential in the use of this dataset to investigate the micro determinant of inventor's productivity. Further work is now ongoing to study the differences in the inventor's network formation between migrants and natives, and its impact on the productivity and the likelihood to move (changing of company and migrating). Furthermore, this database doesn't aim only at studying migration or innovation, thanks to the detailed information from LinkedIn data, it can be also used to investigate more general questions in labour economics, such as: does education mobility influences the labour market career path?

Résumé en français

De tout temps, l'humanité est en mouvement. Certaines personnes se déplacent à la recherche d'opportunités économiques, pour rejoindre leur famille ou pour étudier ; d'autres se déplacent pour fuir des conflits, la persécution, le terrorisme ou des violations des droits de l'homme. De nos jours, de plus en plus de personnes fuient les effets néfastes du changement climatique, des catastrophes naturelles ou d'autres facteurs environnementaux.

Entre 2000 et 2017, le nombre de migrants dans le monde est passé de 173 à 258 millions d'individus et la part des migrants dans la population mondiale est passée de 2,8% à 3,4%. Ainsi, aujourd'hui, plus de personnes que jamais vivent dans un pays différent de celui dans lequel elles sont nées. Parmi les flux migratoires, celui des migrants hautement qualifiés a particulièrement augmenté. La part des personnes hautement qualifiées, définies dans les statistiques officielles comme des personnes ayant achevées des études supérieures, dans la population totale migrante est passée de 27% à 50% entre 1990 et 2010 (OIM, 2018). Cela peut être dû à l'augmentation des niveaux d'éducation dans le monde et à la demande croissante de main-d'œuvre qualifiée, en particulier dans les économies développées. En outre, de meilleurs salaires et conditions d'emploi, une meilleure information, le recrutement et des coûts de transport moins élevés encouragent les migrants qualifiés à rechercher un emploi dans les économies développées. Les diplômés en sciences, technologie, ingénierie et mathématiques (STIM) contribuent de manière significative à ce flux, en particulier les flux reliant la Chine, l'Inde et l'Europe de l'Est aux États-Unis et à d'autres pays anglophones (Docquier & Rapoport, 2012; Freeman, 2013).

Cette question soulève plusieurs questions sur le rôle des migrants dans le processus d'innovation dans leurs pays de destination et d'origine. Cette thèse se compose de trois chapitres liés qui examinent autant de questions. Le chapitre 1 traite du manque de données sur la migration des travailleurs STIM et décrit la méthodologie utilisée pour créer la source de données principale pour les chapitres suivants, à savoir le jeu de données « Linked Inventor ». La base de données Linked Inventor associe des informations sur les inventeurs, extraites de LinkedIn (un média social à vocation professionnelle), à des données de brevets collectées auprès de l'Office des brevets et des marques de commerce des États-Unis (USPTO), de l'Office européen des brevets (OEB) et de l'Organisation mondiale de la propriété intellectuelle. (OMPI). Le résultat est un ensemble d'informations très détaillées sur les inventeurs, grâce auquel nous pouvons étudier les différentes caractéristiques des inventeurs migrants et autochtones travaillant dans plusieurs pays.

Dans le chapitre 2, nous examinons la différence de productivité entre les inventeurs indiens et les inventeurs nés aux États-Unis. Parmi les migrants, nous distinguons par deux voies d'entrée possibles à

destination, à savoir l'éducation et le travail. En outre, parmi les migrants entrant sur le marché du travail, nous distinguons s'ils changent d'entreprise ou migrent au sein d'une même entreprise, le plus souvent entre différents pays dans la même entreprise multinationale. Nous examinons également la différence de productivité des migrants entrant aux États-Unis à différentes périodes de la politique de visas H1-B, que nous utilisons comme indicateur indirect du degré de sélection des migrants. Enfin, nous étudions la mobilité des migrants et des natifs entre les entreprises dans le pays de destination, en découvrant un effet différent de la mobilité au cours de l'activité de brevetage et avant (après) le premier (dernier) brevet déposé. Nous montrons que les migrants sont plus mobiles en changeant d'entreprise que les natifs dans le pays de destination.

Dans le chapitre 3, nous nous concentrons sur l'Europe et comparons la productivité, dans le pays d'origine, des migrants de retour et des natifs non-migrants. Parmi les migrants de retour, nous distinguons les inventeurs ayant travaillé à l'étranger de ceux qui n'ont qu'une expérience éducative internationale. Nous ne trouvons un impact positif sur la productivité que pour les migrants de retour ayant une expérience professionnelle à l'étranger.

Cette thèse confirme qu'il existe un potentiel de gain de cerveaux à la fois pour le pays de destination et le pays d'origine du migrant, mais qu'il varie selon certaines conditions, en fonction des expériences dans le pays de destination pour les migrants de retour ou des canaux d'entrée des migrants. Grâce à nos nouvelles données, nous renforçons les résultats précédents en contrôlant pour des caractéristiques individuelles cruciales et pouvons tenir compte des politiques nationales telles que le visa H1-B aux États-Unis. À notre connaissance, il s'agit de la première tentative de relier la littérature sur la migration des entreprises et la mobilité inter-entreprises, tant pour les natifs que pour les migrants. En outre, notre projet constitue également la première tentative d'enquête et de prise en compte de la sélection positive et négative des migrants européens de retour dans les domaines des STIM. Enfin, nous sommes en mesure de contrôler pour un ensemble de caractéristiques « précoces » des inventeurs qui ne sont généralement pas observées dans d'autres études, telles que : le pays où le plus haut niveau d'éducation a été obtenu, les différents types de mobilité et l'expérience professionnelle acquise avant le premier brevet déposé.

Postfazione in italiano

Sin dai primi tempi, l'umanità è stata in movimento. Alcune persone si spostano in cerca di opportunità economiche, per riunirsi alla famiglia o studiare; altre per sfuggire a conflitti, persecuzioni, terrorismo o violazioni dei diritti umani. Altre ancora, migrano in risposta agli effetti dei cambiamenti climatici, disastri naturali o altri fattori ambientali. Ad oggi, più persone che mai vivono in un paese diverso da quello in cui sono nate: tra il 2000 e il 2017 il numero di migranti internazionali è passato da 173 a 258 milioni di individui e la percentuale di migranti sulla popolazione mondiale è passata dal 2,8% al 3,4%.

Tra i flussi migratori, quello dei migranti altamente qualificati è particolarmente aumentato. La percentuale di individui altamente qualificati, definita nelle statistiche ufficiali come coloro che hanno un diploma d'istruzione terziaria, sul totale dei migranti è cresciuta dal 27% al 50% tra il 1990 e il 2010 (OIM, 2018). Una possibile spiegazione può essere l'aumento dei livelli di istruzione in tutto il mondo e la crescente domanda di manodopera qualificata, in particolare da parte delle economie sviluppate. Inoltre, migliori salari, condizioni di lavoro, informazioni, assunzioni e minori costi di trasporto incoraggiano i più qualificati a cercare lavoro nelle economie sviluppate. I laureati in Scienze, Tecnologia, Ingegneria e Matematica (STEM) contribuiscono in modo significativo a questo flusso, in particolare ai flussi che collegano la Cina, l'India e l'Europa orientale agli Stati Uniti e ad altri paesi di lingua inglese (Docquier & Rapoport, 2012; Freeman, 2013).

Questo fenomeno solleva domande su quale sia il ruolo dei migranti nel processo di innovazione nei loro paesi di destinazione e di origine. Questa tesi è composta da tre capitoli correlati che cercano di dare una risposta a queste domande. Il capitolo 1 riguarda la mancanza di dati sulla migrazione dei lavoratori STEM e descrive la metodologia utilizzata per creare la principale fonte di dati su cui i successivi capitoli sono basati, ovvero il database "Linkedin Inventor". Questo database abbina informazioni sugli inventori, recuperate da LinkedIn (un social media orientato ai professionisti), con i dati sui brevetti raccolti dall'Ufficio brevetti e marchi degli Stati Uniti (USPTO), dall'Ufficio europeo dei brevetti (EPO) e dall'Organizzazione mondiale della proprietà intellettuale (OMPI). Il risultato è un insieme di dettagliate informazioni sugli inventori, grazie al quale siamo in grado di analizzare le diverse caratteristiche degli inventori, migranti e non, che lavorano in diversi paesi.

Nel capitolo 2 indaghiamo la differenza di produttività tra inventori migranti (indiani) e non (statunitensi) che lavorano negli Stati Uniti. Tra i migranti, differenziamo per due possibili canali di ingresso a destinazione, ovvero istruzione e lavoro. Tra i migranti che accedono grazie a quest’ultimo, distinguiamo ulteriormente tra coloro che cambiano azienda o si trasferiscono all'interno della stessa, più tipicamente tra diverse sedi, in diversi Paesi, della stessa impresa multinazionale. Analizziamo inoltre la differenza di produttività dei migranti che entrano negli Stati Uniti durante diversi periodi di restrizione sui visti H1-B, che utilizziamo per un'approssimazione sul grado di selezione dei migranti. Infine, studiamo la mobilità lavorativa dei migranti e non a destinazione, trovando un diverso effetto della mobilità durante l'attività di brevetto e prima (dopo) del primo (ultimo) brevetto depositato. Mostriamo che i migranti sono più mobili dell'azienda rispetto ai nativi di destinazione.

Nel capitolo 3, ci concentriamo sull'Europa e confrontiamo la produttività, nel paese di origine, dei migranti di ritorno e dei residenti. Tra i migranti di ritorno, distinguiamo tra gli inventori che hanno lavorato all'estero e quelli che hanno avuto solo un'esperienza di studio. I nostri risultati mostrano un premio di produttività solo per i migranti di ritorno con esperienza lavorativa all'estero.

Questa tesi conferma che esiste un potenziale effetto di *brain gain* sia per il paese ospitante che per il paese di origine dei migranti, che varia in base alle esperienze, i canali di ingresso, e le esperienze di ritorno. Grazie ai nostri dati, possiamo controllare per importanti caratteristiche individuali e prendere in considerazione politiche nazionali come il visto H1-B negli Stati Uniti. Ad oggi, il nostro è il primo tentativo di combinare le letterature sulla migrazione e la mobilità aziendale. Inoltre, è anche il primo tentativo di analizzare la selezione positiva e negativa dei migranti europei che ritornano nel loro Paese di origine. Infine, siamo in grado di controllare per una serie di caratteristiche iniziali degli inventori che di solito non vengono osservate in altri studi, come ad esempio il Paese in cui è stato ottenuto il più alto livello di istruzione, i modelli di mobilità, e l'esperienza di lavoro antecedente al primo brevetto depositato.

Bibliography

Adams, J., H., R. & Page, J., 2005. Do international migration and remittances reduce poverty in developing. *World development,* pp. 1645-1669.

Agarwal, R., Ganco, M. & Ziedonis, R. H., 2009. Reputations for toughness in patent enforcement: Implications for knowledge spillovers via inventor mobility. *Strategic Management Journal,* pp. 30(13), 1349-1374.

Agrawal, A., Kapur, D., McHale, J. & O. A., 2011. Brain drain or brain bank? The impact of skilled emigration on poor-country innovation.. *Journal of Urban Economics,* pp. 69(1), 43-55.

Alesina, A., Harnoss, J. & Rapoport, H., 2016. Birthplace diversity and economic prosperity. *Journal of Economic Growth,* pp. 21(2), 101-138.

Allan, K., 2003. Linguistic metatheory.. *Language Sciences,* pp. 25(6), 533-560.

Almeida, P. & Kogut, B., 1999. Localization of knowledge and the mobility of engineers in regional networks. *Management science,* pp. 45(7), 905-917.

Asiedu, A., 2005. Some benefits of migrants' return visits to Ghana. *Population, Space and Place,* pp. 11(1), 1-11.

Baer, J., 2018. *Fall 2018 Intenational Student Enrollment Hot Topics Survey,* s.l.: IIE Center for Academic Mobility.

Barham, B. & Boucher, S., 1998. Migration, remittances, and inequality: estimating the net effects of migration on income distribution. *Journal of development economics,* pp. 55(2), 307-331.

Barnett, G. A. & Wu, R. Y., 1995. The international student exchange network: 1970 & 1989. *Higher Education,* pp. 30(4), 353-368.

Batalova, J. & Fix, M., 2017. *New brain gain: Rising human capital among recent immigrants to the United States..* Washington, DC: Migration Policy Institute.

Becker, G. S., 1962. Investment in human capital: A theoretical analysis.. *Journal of political economy,* pp. 70(5, Part 2), 9-49.

Becker, G. S., 1965. A Theory of the Allocation of Time. *The economic journal,* pp. 493-517.

Beine, M., Docquier, F. & Rapoport, H., 2007. Measuring international skilled migration: a new database controlling for age of entry.. *The World Bank Economic Review,* pp. 21(2), 249-254.

Bertoli, S. et al., 2009. The battle for brains: how to attract talent.. *CEPR project, Fondazione Rodolfo Debenedetti.*

Bhagwati, J. & Hamada, K., 1974. The brain drain, international integration of markets for professionals and unemployment: a theoretical analysis.. *Journal of Development Economics,* pp. 1(1), 19-42.

Bhagwati, J. N., 1976. Taxing the brain drain. Challenge. pp. 19(3), 34-38.

Borjas, G. & Bratsberg, B., 1996. Who leaves? The outmigration of the foreign. *Review of Economics and Statistics.*

Borjas, G. J., 1982. The earnings of male Hispanic immigrants in the United States. *ILR Review,* pp. 35(3), 343-353.

Borjas, G. J., 1985. Assimilation, changes in cohort quality, and the earnings of immigrants. *Journal of labor Economics,* pp. 3(4), 463-489.

Borjas, G. J., 1987. Self-selection and the earnings of immigrants.

Borjas, G. J., Bronars, S. G. & Trejo, S. J., 1992. Self-selection and internal migration in the United States. *Journal of urban Economics,* pp. 32(2), 159-185.

Borjas, G. J. & Doran, K. B., 2012. The collapse of the Soviet Union and the productivity of American mathematicians.. *The Quarterly Journal of Economics,* pp. 127(3), 1143-1203.

Breschi, S. & Lenzi, C., 2010. Spatial patterns of inventors' mobility: Evidence on US urban areas. *Papers in Regional Science,* pp. 89(2), 235-250.

Breschi, S., Lenzi, C., Lissoni, F. & Vezzulli, A., 2010. The geography of knowledge spillovers: the role of inventors' mobility across firms and in space. *The Handbook of Evolutionary Economic Geography, Cheltenham: Edward Elgar,* pp. 353-369.

Breschi, S. & Lissoni, F., 2005. Mobility and social networks: localised knowledge spillovers revisited.. *Annales d'economie et de statistique.*

Breschi, S. & Lissoni, F., 2009. Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of economic geography,* pp. 9(4), 439-468.

Breschi, S., Lissoni, F. & Miguelez, E., 2018. Return Migrants' Self-selection: Evidence For Indian Inventors.. *In The Role of Immigrants and Foreign Students in Science, Innovation, and Entrepreneurship.,* p. University of Chicago Press.

Breschi, S., Lissoni, F. & Montobbio, F., 2007. The scientific productivity of academic inventors: new evidence from Italian data.. *Econ. Innov. New Techn,* pp. 16(2), 101-118.

Breschi, S., Lissoni, F. & Tarasconi, G., 2014. Inventor Data for Research on Migration and innovation: A Survey and a Pilot. *WIPO.*

Breschi, S., Lissoni, F. & Tarasconi, G., 2017. Inventor data for research on migration and innovation.

Campbell, B. A., Coff, R. & Kryscynski, D., 2012. Rethinking sustained competitive advantage from human capital. *Academy of Management Review,* pp. 37(3), 376-395.

Campbell, B. A., Ganco, M., Franco, A. M. & Agarwal, R., 2012. Who leaves, where to, and why worry? Employee mobility, entrepreneurship and effects on source firm performance.. *Strategic Management Journal,,* pp. 33(1), 65-87.

Cappelli, R., Czarnitzki, D., Doherr, T. & Montobbio, F., 2019. Inventor mobility and productivity in Italian regions. *Regional Studies,* pp. 53(1), 43-54.

Carliner, G., 1980. Wages, earnings and hours of first, second, and third generation American males. *Economic Inquiry,* pp. 18(1), 87-102.

Carrington, M. W. & Detragiache, M. E., 1998. How big is the brain drain?. *International Monetary Fund.,* pp. No. 98-102.

Chellaraj, G., Maskus, K. E. & Mattoo, A., 2005. The contribution of skilled immigration and international graduate students to US innovation.. *The World Bank.*

Chiquiar, D. & Hanson, G. H., 2005. International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States. *Journal of political Economy,* pp. 113(2), 239-281.

Chiswick, B. R., 1978. The effect of Americanization on the earnings of foreign-born men. *Journal of political Economy,* pp. 86(5), 897-921.

Chiswick, R., 2011. High-skilled immigration in a global labor market.

Constant, A. & Massey, D. S., 2002. Return migration by German guestworkers: Neoclassical versus new economic theories. *International migration,* pp. 40(4), 5-38.

Corley, E. A. & Sabharwal., M., 2007. Foreign-born academic scientists and engineers: producing more and getting less than their US-born peers?. *Research in Higher Education,* pp. 48(8): 909-940.

Cuecuecha, A., 2005. The Characteristics of the Immigrants from Mexico to the US. *ITAM, CIE, JEL F 22,* p. C25.

Czaika, M., 2018. *High-skilled Migration: Drivers and Policies..* Oxford University Press.: s.n.

Czaika, M. & Parsons, C. R., 2017. The gravity of high-skilled migration policies. *Demography,* pp. 54(2), 603-630.

De Coulon, A. & Piracha, M., 2005. Self-selection and the performance of return migrants: the source country perspective.. *Journal of Population Economics,* pp. 18(4), 779-807.

De Haas, H., Fokkema, T. & Fihri, M. F., 2015. Return migration as failure or success?. *Journal of international migration and integration,* pp. 16(2), 415-429.

Defoort, C., 2008. Tendances de long terme des migrations internationales: analyse à partir des six principaux pays receveurs.. *Population,* pp. 63(2), 317-351.

Di Lorenzo, F. & Almeida, P., 2017. The role of relative performance in inter-firm mobility of inventors. *Research Policy,* pp. 46(6), 1162-1174.

Docquier, F., Lowell, B. L. & Marfouk, A., 2009. A gendered assessment of highly skilled emigration.. *Population and Development Review,* pp. 35(2), 297-321.

Docquier, F. & Marfouk, A., 2006. International migration by education attainment, 1990–2000.. *International migration, remittances and the brain drain,* pp. 151-199.

Docquier, F. & Rapoport, H., 2012. Globalization, brain drain, and development.. *Journal of Economic Literature,* pp. 50(3), 681-730.

Dumont, J. C. & Spielvogel, G., 2008. *Return migration: A new perspective.,* s.l.: rganization for Economic Cooperation and Development (OECD), International Migration Outlook, Annual Report.

Dustmann, C. & Glitz, A., 2011. Migration and education. *Handbook of the Economics of Education,* pp. 327-439.

Dustmann, C. & Kirchkamp, O., 2002. The optimal migration duration and activity choice after re-migration.. *Journal of development economics,* pp. 67(2), 351-372.

Dustmann, C. & Weiss, Y., 2007. Return migration: theory and empirical evidence from the UK.. *British Journal of Industrial Relations,* pp. 236-256.

Eddy, L., 2011. *The effect of student mobility on student achievement.* s.l.:s.n.

Ejermo, O. & Jung, T., 2011. Team experience of inventors and inventive performance.. *In Paper submitted for the 6th Annual Conference of the EPIP Association.*

Elms, T., 2008. Lexical distance among the languages of europe.

Feldman, M. P., Link, A. N. & Siegel, D. S., 2012. The economics of science and technology: an overview of initiatives to foster innovation, entrepreneurship, and economic growth. *Springer Science & Business Media.*

Fernandez-Zubieta, A., Geuna, A. & Lawson, C., 2015. What do we know of the mobility of research scientists and impact on scientific production. *Global mobility of research scientists.*

Ferrucci, E. & Lissoni, F., 2019. Foreign inventors in Europe and the United States: Diversity and Patent Quality. *Research Policy.*

Fokkema, T. & De Haas, H., 2015. Pre-and Post-Migration Determinants of Socio-Cultural Integration of African Immigrants in Italy and Spain.. *International Migration,* pp. 53(6), 3-26.

Foley, C. F. & Kerr, W. R., 2013. Ethnic innovation and US multinational firm activity.. *Management Science,* pp. 59(7), 1529-1544.

Freeman, R. B., 2010. Globalization of scientific and engineering talent: international mobility of students, workers, and ideas and the world economy. *Economics of Innovation and New Technology,* pp. 19(5), 393-406.

Gambardella, A., Harhoff, D. & Verspagen, B., 2012. The value of patented inventions at the extensive and intensive margins, mimeo.

Gambardella, A., H. D. & Verspagen, B., 2006. The value of patents. *In EPIP Conference, Munich.*

Genuer, R., Poggi, J. M. & & Tuleau-Malot, C., 2010. Variable selection using random forests. *Pattern Recognition Letters,* Volume 2225-2236, p. 31(14).

Griliches, Z., 1998. Patent statistics as economic indicators: a survey. *R&D and productivity: the econometric evidence,* Volume 287-343, p. University of Chicago Press.

Gürüz, K., 2008. *Higher Education and International Student Mobility in the Global Knowledge Economy.* Albany, NY: State: University of New York Press.

Hall, B. H. J. A. B. &. T. M., 2001. The NBER patent citation data file: Lessons, insights and methodological tools. *National Bureau of Economic Research.,* p. No. w8498.

Hall, B. H., J. M. & Turner, L., 2007. Identifying age, cohort, and period effects in scientific research productivity: discussion and illustration using simulated and actual data on French physicists. *Econ. Innov. New Techn. 16.2,* pp. 159-177.

Hall, B., Jaffe, A. & Trajtenberg, M., 2001. The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools.. *NBER Working paper ,* Volume No. 8498.

Hanson, G. H. & Liu, C., 2017. High-Skilled Immigration and the Comparative Advantage of Foreign-Born Workers across US Occupations. In: N. B. o. E. Research, ed. *High-Skilled Migration to the United States and its Economic Consequences.* s.l.:University of Chicago Press, pp. p. 7 - 40.

Haque, N. U. & Kim, S. J., 1995. "Human capital flight": Impact of migration on income and growth. *Staff Papers,* pp. 42(3), 577-607.

Hoisl, K., 2007. Tracing mobile inventors—the causality between inventor mobility and inventor productivity. *Study of Inventors,* pp. 65-119.

Hoisl, K., 2009. Does mobility increase the productivity of inventors?. *The Journal of Technology Transfer,* pp. 34(2), 212-225.

Hoisl, K. & de Rassenfosse, G., 2014. Knowledge fit and productivity gains from mobility. *Druid Society Conference.*

Horta, H., 2009. Holding a post-doctoral position before becoming a faculty member: does it bring benefits for the scholarly enterprise?. *Higher Education,* pp. 58(5), 689-721.

Hunt, J., 2004. Are migrants more skilled than non-migrants? Repeat, return, and same-employer migrants. *Canadian Journal of Economics/Revue canadienne d'économique,* pp. 37(4), 830-849.

Hunt, J., 2011. Which immigrants are most innovative and entrepreneurial? Distinctions by entry visa. *Journal of Labor Economics,* pp. 29(3), 417-457.

Hunt, J. & Gauthier-Loiselle, M., 2010. How much does immigration boost innovation?. *American Economic Journal: Macroeconomics,* pp. 2(2), 31-56.

Iacus, S. M., King, G. & Porro, G., 2012. Causal inference without balance checking: Coarsened exact matching. *Political analysis,* pp. 20(1), 1-24.

Ibarraran, P. & Lubotsky, D., 2007. Mexican immigration and self-selection: New evidence from the 2000 Mexican census. *In Mexican immigration to the United States,* Volume University of Chicago Press., pp. pp. 159-192.

IOM, 2018. *Global migration indicators,* Berlin: Global Migration Data Analysis Centre (GMDAC) International Organization for Migration.

Kahn, S. & MacGarvie, M., 2011. he effects of the foreign Fulbright program on knowledge creation in science and engineering.. *In The rate and direction of inventive activity revisited (pp. 161-197),* pp. pp. 161-197.

Kerr, S. P., Kerr, W., Ozden, C. & Parsons, C., 2016. Global talent flows. *The World Bank.*

Kerr, S. P., Kerr, W., Özden, Ç. & Parsons, C., 2017. High-skilled migration and agglomeration.. *Annual Review of Economics,* pp. 9, 201-234.

Kerr, W. R., 2007. The Ethnic Composition of US Inventors. *Harvard Business School Working Paper,* pp. No. 08-006.

Kerr, W. R., 2008. Ethnic scientific communities and international technology diffusion. *The Review of Economics and Statistics,* pp. 90(3), 518-537.

Kerr, W. R., 2018. *The Gift of Global Talent: How Migration Shapes Business, Economy & Society..* s.l.:Stanford University Press.

Kerr, W. R. & Lincoln, W. F., 2010. The supply side of innovation: H-1B visa reforms and US ethnic invention.. *Journal of Labor Economics,* pp. 28(3), 473-508.

Khadria, B., 2012. Demography,Diaspora and Development: Migration in India's Transformation to a "Superpower" in the 21st Century. *International Labor Migration,* p. 246.

Knerr, B., Tlatlik, R. & Xi, Z., 2010. The benefits of international student mobility. A comparison between Chinese and Indian students in Germany.

Koser, K. & Salt, J., 1997. The geography of highly skilled international migration. *International Journal of Population Geography,* pp. 3(4), 285-303.

Latham, W., Le Bas, C. & Volodin, D., 2011. Value of invention, prolific inventor productivity and mobility: evidence from five countries, 1975-2002.

Lenzi, C., 2010. Workers' mobility and patterns of knowledge diffusion: evidence from Italian data. *The Journal of Technology Transfer,* pp. 651-670.

Lissoni, F., Llerena, P., McKelvey, M. & Sanditov, B., 2008. Academic patenting in Europe: New evidence from the KEINS database.. *Research Evaluation,* pp. 17(2), 87-102.

Lissoni, F., Montobbio, F. & Zirulia, L., 2013. Inventorship and Authorship as Attribution Rights: An Equiry in the Economics of Scientific Credit. *Journal of Economic Behavior & Organization,* pp. 95, 49-69.

Mahroum, S., 2001. Europe and the immigration of highly skilled labour.. *International Migration,* pp. 39(5), 27-43.

Mariani, M. & Romanelli, M., 2007. “Stacking” and “picking” inventions: The patenting behavior of European inventors.. *Research Policy,* pp. 36, 1128-1142.

Marx, M., Strumsky, D. & Fleming, L., 2009. Mobility, skills, and the Michigan non-compete experiment. *Management Science,* pp. 55(6), 875-889.

Mayr, K. & Peri, G., 2009. Brain drain and brain return: theory and application to Eastern-Western Europe.. *The BE Journal of Economic Analysis & Policy,* p. 9(1).

Miguelez, E. & Fink, C., 2013. Measuring the international mobility of inventors: A new database. *World Intellectual Property Organization-Economics and Statistics Division.*

Miyagiwa, K., 1991. Scale economies in education and the brain drain problem. *International Economic Review,* pp. 743-759.

Moraga, J. F.-H., 2011. New Evidence on Emigrant Selection. *The Review of Economics and Statistics,* pp. 93(1), 72-96.

NBER, 2016. *Immigrants Play a Key Role in STEM Fields,* Cambridge: s.n.

Ottaviano, G. I. & Peri, G., 2006. The economic value of cultural diversity: evidence from US cities. *Journal of Economic geography,* pp. 6(1), 9-4.

Ozgen, C., Nijkamp, P. & Poot, J., 2012. Immigration and innovation in European regions. *Migration impact assessment: New horizons,* pp. 261-300.

Parey, M., Ruhose, J., Waldinger, F. & Netz, N., 2017. The selection of high-skilled emigrants. *Review of Economics and Statistics,* pp. 99(5), 776-792.

Parey, M. & Waldinger, F., 2011. Studying abroad and the effect on international labour market mobility: Evidence from the introduction of ERASMUS.. *The Economic Journal,* pp. 121(551), 194-222.

Parrotta, P., Pozzoli, D. & Pytlikova, M., 2014. The nexus between labor diversity and firm’s innovation. *Journal of Population Economics,* pp. 27(2), 303-364.

Peri, G., 2007. Higher Education, Innovation and Growth. *In Giorgio Brunello, Pietro Garibaldi and Etienne Wasmer eds. Education and Training in Europe, Oxford: Oxford University Press.*

Pezzoni, M., Lissoni, F. & Tarasconi, G., 2014. ow to kill inventors: testing the Massacrator algorithm for inventor disambiguation.. *Scientometrics,* pp. 101(1), 477-504.

Pinheiro, J. et al., 2007. Linear and nonlinear mixed effects models. *R package version,* pp. 3(57), 1-89.

Piracha, M. & Vadean, F., 2010. Return migration and occupational choice: Evidence from Albania.. *World Development,* pp. 38(8), 1141-1155.

Rodríguez, V. & Egea, C., 2006. Return and the social environment of Andalusian emigrants in Europe.. *Journal of Ethnic and Migration Studies,* pp. 32(8), 1377-1393.

Rosenkopf, L. & Almeida, P., 2003. Overcoming local search through alliances and mobility. *Management science,* pp. 49(6), 751-766.

Ruggles, S. et al., 2010. Integrated Public Use Microdata Series (IPUMS): Version 5.0. *Minneapolis: University of Minnesota.*

Saxenian, A., 2006. The new argonauts: Regional advantage in a global economy.. *Harvard University Press, Cambridge, MA.*

Saxenian, A. M. Y. & Quan, X., 2002. Local and global networks of immigrant professionals in Silicon Valley. *Public Policy Instit. of CA.*

Schettino, F., Sterlacchini, A. & Venturini, F., 2013. Inventive productivity and patent quality: evidence from Italian inventors.. *Journal of Policy Modelling,* pp. 35(6),1043-1056.

Shin, J. C., Jung, J., Postiglione, G. A. & Azman, N., 2014. Research productivity of returnees from study abroad in Korea, Hong Kong, and Malaysia.. *Minerva,* pp. 52(4), 467-487.

Sjaastad, L. A., 1962. The costs and returns of human migration.. *Journal of political Economy,* pp. 70(5, Part 2), 80-93.

Skirgard, H., Roberts, S. G. & Yencken, L., 2017. Why are some languages confused for others? Investigating data from the great language game.. *PloS one,* p. 12(4).

Somaya, D., Williamson, I. O. & Lorinkova, N., 2008. Gone but not lost: The different performance impacts of employee mobility between cooperators versus competitors. *Academy of Management Journal,* pp. 51(5), 936-953.

Stark, O., Helmenstein, C. & Prskawetz, A., 1997. A brain gain with a brain drain.. *Economics letters,* pp. 55(2), 227-234.

Stephan, P. E. & Levin, S. G., 2001. Exceptional contributions to US science by the foreign-born and foreign-educated. *Population research and Policy review,* pp. 20(1-2), 59-79.

Straubhaar, T. & Wolburg, M., 1998. Brain drain and brain gain in Europe: An evaluation of the East-European migration to Germany. *Univ., of California, Center for German and European Studies.*

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society,* pp. 267-288.

Topel, R. & Ward, M., 1992. Job mobility and the careers of yound men.. *The Quarterly journal of Economics,* pp. 107, 439-479.

Trajtenberg, M., 2005. The mobility of inventors and the productivity of research. *Conference presentation, NBER Summer Institute.*

Trajtenberg, M. & Shalem, R., 2009. Software patents, inventors and mobility. Inventors and Mobility. *SSRN paper.*

Turner, L. & Mairesse, J., 2005. Individual productivity differences in public research: How important are non-individual determinants? An econometric study of French physicists' publications and citations (1986-1997).. *Centre National de la Recherche Scientifique.*

Tyshchenko, K., 1999. Metatheory of Linguistics.

UNESCO, 2009. *Repport on the 2009 World Conference on Higher Education,* s.l.: s.n.

UNESCO, 2018. *Facts and Figures, Mobility in higher education.,* s.l.: s.n.

Wadhwa, V., Rissing, B., Saxenian, A. & Gereffi, G., 2007. Education, entrepreneurship and immigration: America's new immigrant entrepreneurs, Part II.

Walsh, J. P. & Nagaoka, S., 2009. Who invents? Evidence from the Japan&US inventor survey. *RIETI Discussion papers.*

Widmaier, S. & Dumont, J. C., 2011. Are recent immigrants different? A new profile of immigrants in the OECD based on DIOC 2005/06.

Wurm, C., 2013. On the Metatheory of Linguistics. *PhD thesis, University of Bielefeld.*

Zwick, T., Frosch, K., Hoisl, K. & Harhoff, D., 2017. The power of individual-level drivers of inventive performance. *Research policy,* pp. 46(1), 121-137.