Research Policy 52 (2023) 104631

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Productivity gains from migration: Evidence from inventors

Gabriele Pellegrino [a,*], Orion Penner [b], Etienne Piguet [c], Gaétan de Rassenfosse [b]

[a] *Department of Economic Policy, Università Cattolica del Sacro Cuore, Milano, Italy*
[b] *EPFL (Ecole polytechnique fédérale de Lausanne), Lausanne, Switzerland*
[c] *University of Neuchâtel, Neuchâtel, Switzerland*

ARTICLE INFO

JEL classification:
F22
J61
O30

Keywords:
Inventor
Productivity
Skilled migration

ABSTRACT

This paper studies the relationship between migration and the productivity of high-skilled workers, as captured by inventors of international patent applications. Using machine learning techniques to uniquely identify inventors across patents, we are able to track the migration patterns of nearly one million individual inventors across countries. Migrant inventors account for more than nine percent of inventors in our sample. The econometric analysis seeks to explain the recurring finding in the literature that migrant inventors are more productive than non-migrant inventors. We find evidence that migrant inventors become about twenty-three percent more productive after having migrated. The disambiguated inventor data are openly available.

1. Introduction

Attracting high-skilled migration is a high priority on the policy agenda (Bertoli et al., 2012). Immigrants are over-represented among academics, scientists, and entrepreneurs in many high-income OECD countries. In the European Union, the 2010 Lisbon Agenda and the Europe 2020 Strategy have emphasized the urgency of attracting foreign-born skilled workers in order to promote competitiveness (European Commission, 2011). In the United States, the bipartisan Immigration Innovation Act of 2015 was introduced with the aim of increasing drastically the number of visas for temporary skilled workers. Elsewhere in the world, the Association of Southeast Asian Nations (ASEAN) has recently set the ambitious goal of creating the ASEAN Economic Community, a unified market facilitating the free flow of skilled workers, among other objectives (IOM, 2014). The rationale behind these policies is that migration allows a better allocation of human capital, thereby raising productivity at the world level—a win-win situation for origin and destination countries in the long term. The counterarguments are that departures of intrinsically more productive individuals lead to a detrimental brain drain for origin countries and may generate a crowding-out phenomenon of native workers at destination countries.

This policy emphasis has been accompanied by a growing academic interest in the topic. Of particular interest, as a recent special issue in this Journal demonstrates, is the study of the role played by high-skilled migrants in diffusing scientific and technical knowledge and, consequently, promoting innovation (Breschi et al., 2020).[1] The present paper adds to the literature on high-skilled migration by studying the productivity patterns of international migrant inventors. It builds on a recent body of work in the economics of science and labor economics that indicates that migrant scientists, a specific group of high-skilled migrants, are more productive than non-migrant scientists. Stephan and Levin (2001) show that a large proportion of the top-tier academic researchers residing in the United States are foreign-born or foreign-educated. Borjas and Doran (2012) show that mathematicians who migrated to the United States from the states of the former Soviet Union following its collapse were significantly more productive than U.S. incumbent mathematicians. Gaulé and Piacentini (2012) study the productivity performance of Chinese chemistry students enrolled in Ph.D. programs in the United States. They show that Chinese Ph.D. students are, on average, more productive than their non-Chinese counterparts. Evidence on a productivity differential between locals and migrants immediately calls the question of the source of this difference.

This paper contributes to the literature by presenting evidence that migrant inventors become more productive as a consequence of the move. It relates most closely to Franzoni et al. (2014), which provides evidence on migrant scientists. Using an instrumental variable approach on cross-sectional survey data, they find that migrant scientists seem to become more productive after a move. Unfortunately, they do not investigate whether migrants were intrinsically more productive than locals before migrating, thereby providing only part of the story. Furthermore, the behavior of scientists, who publish papers in an

---

* Corresponding author.
E-mail address: gabriele.pellegrino@unicatt.it (G. Pellegrino).

[1] A rich body of work has investigated various economic questions related to immigration in general, including low-skilled migration. See Borjas (1994), Friedberg and Hunt (1995), Gaston and Nelson (2002), Okkerse (2008), and Kerr et al. (2018) for comprehensive surveys of this literature.

https://doi.org/10.1016/j.respol.2022.104631
Received 26 March 2021; Received in revised form 30 June 2022; Accepted 17 September 2022

academic context, may differ from that of inventors who file patents in a corporate context.

Our analysis tracks the productivity patterns of inventors over time, some of which are international migrants, some of which are not. This setup allows us to estimate the intrinsic productivity level of researchers and any potential productivity bump following their migration. We exploit data on inventors listed on international patent applications filed under the Patent Cooperation Treaty (PCT) route.[2] These data contain information on inventors' countries of nationality and residence. We infer migration events by looking at differences in residence and nationality countries following Fink and Miguélez (2013).[3]

Specifically, we apply a machine-learning algorithm to disambiguate (that is, uniquely identify) all inventors recorded in the database. We identify migrant inventors through differences in country of residence and citizenship. These data are openly available in order to encourage follow-on research.[4]

The results are as follows. First, a preliminary look at the data reveals that migrant inventors make up about ten percent of total inventors in our sample. Second, the econometric analysis suggests that migrant inventors become more productive after they have moved. In our preferred setup, migrant inventors enjoy a twenty-three-percent productivity gain. The finding holds under a range of alternative specifications and robustness tests. The analysis rules out several potential explanations for the productivity gain, and we speculate that it may be driven by skill upgrading or greater fit (ala Jovanovic, 1979).

The present paper relates to the literature that studies individual-level effects of high-skilled migration. It specifically focuses on the productivity effects of migration on migrants. In doing so, it complements an important stream of work that studies the productivity effects of migration on locals. For instance, Borjas and Doran (2012) exploit the collapse of the Soviet Union and the consequent influx of ex-Soviet mathematicians to the United States. They study the impact of migration on the productivity of U.S. mathematicians, pointing to a consistent decrease in the publication rates for native workers in subfields that attracted more ex-Soviet mathematicians. In a related paper, Ganguli (2015) finds that the influx of ex-Soviet scientists in the United States increases the number of citations that U.S. papers make to Soviet-era papers, thus contributing to the diffusion of scientific knowledge. More recently, Ferrucci (2020) studies the effect of the migration to Germany of ex-Soviet Union inventors on the patent production of German inventors. He finds an increase in patent production in the technology fields in which the Soviets were most involved, due in part to the creation of new collaboration ties between local and migrant inventors. Moser et al. (2014) use a different natural experiment to investigate similar research questions. They look at the effect on the level and quality of U.S. innovation caused by the influx of German Jewish chemistry scientists and professors during the Nazi Germany period. The authors also find evidence of a positive effect of migration on the productivity level of local inventors. Along the same line, Yoon and Doran (2020) and Moser and San (2020) show that the change in the quota system introduced in the early 1920s in the United States reduced scientific discovery and patented inventions.

The present paper also relates to works that have leveraged patent data to study questions related to the international mobility of inventors. In particular, Miguelez and Noumedem Temgoua (2020) also use the data assembled by Fink and Miguélez (2013). They study the relationship between international knowledge diffusion and the migration of inventors on a global scale. The authors find evidence of knowledge diffusion for migrants' host countries and home countries. Bahar et al. (2020) come to the same conclusions by using a different set of patents/inventors data and employing an instrumental variable approach. Similarly, Breschi et al. (2017) evaluate the role of ethnic ties in the diffusion of technical knowledge using a database of patents filed by U.S.-resident inventors of foreign origin. They find evidence of a diaspora effect for Asian countries. Other relevant papers include Caviggioli et al. (2020), Useche et al. (2020) and Marino et al. (2020).

Finally, the paper also contributes to the technical literature that has provided open datasets on disambiguated scientists and inventors. Perhaps the most visible contribution in this field is the Author-ity algorithm by Torvik et al. (2005), which disambiguates author names in the MEDLINE database. Regarding patent data specifically, our work is in the tradition of Li et al. (2014) and Morrison et al. (2017), among others.

The rest of the paper is organized as follows. Section 2 introduces the data and provides an overview of the migration patterns. Section 3 presents the econometric model, the baseline regression results, and alternative specifications. Section 4 extends the analysis in several ways to test to robustness of the findings and to shed light on the underlying mechanism. Section 5 concludes.

2. Migration patterns

2.1. PCT data

We observe inventors listed in patent applications filed under the Patent Cooperation Treaty (PCT). The PCT is an international treaty that facilitates international patenting. It is administrated by the Geneva-based World Intellectual Property Organization (WIPO) and has 156 signatory member states as of April 6, 2022.[5] The database of PCT applications represents a rich and unique opportunity to study the phenomenon of skilled workers migration because it contains highly accurate information on both the *country of residence* and *nationality* of each inventor.

Information on the nationality of inventors is a unique feature of patent applications filed under the PCT. According to the treaty, only nationals or residents of a PCT contracting state are entitled to file PCT applications. Thus, to verify that each application fulfills at least one of these two requirements, the PCT application form asks for both nationality and residence (Fink and Miguélez, 2013). As a general rule, the PCT system documents the country of residence and nationality for applicants only and not for inventors. However, the U.S. patent application procedure requires all inventors listed in a PCT application to be listed as applicants—at least until 2012. Thus, if a given PCT application includes the United States as a country in which the applicant considered pursuing a patent, all inventors are listed as applicants, and their residence and nationality information are, in principle, available.[6] Accordingly, we limit the scope of our analysis to the sample of inventors observed for the period 1990–2011.

Fink and Miguélez (2013) provide a detailed account of the PCT database used in the present analysis, and we refer the reader to the original paper for a lengthy discussion. However, we discuss two elements that are key to the present analysis. First, although not all

---

[2] For an inventor to be included in the dataset, she should have declared her intention (real or not) to seek protection in the United States when applying for a PCT patent, see detailed explanations in Section 2.1.

[3] Thus, we are not in a position to track mobility events within countries (across cities or regions) since information on country of residence remains unchanged.

[4] The data are available from the Harvard Dataverse at https://doi.org/10.7910/DVN/AETFTF.

[5] Fink and Miguélez (2013) offer a detailed description of the main characteristics and the functioning of the PCT system.

[6] The United States is the most frequently designated country in PCT applications, implying that we have inventor nationality for the vast majority of applications. However, the 2011 Leahy–Smith America Invents Act (AIA) in the United States removed the requirement that inventors must also be named as applicants (Fink and Miguélez, 2013).

Table 1
Typology of migrant inventors.

	Residence always different from nationality		Residence (at least once) equal to nationality	
Change in residence and nationality	Type of inventor	No of Inventors	Type of inventor	No of Inventors
Both residence and nationality are constant over time	(1) Migrant (move not recorded)	61,686	(5)Non migrant	790,957
Residence varies over time, while nationality is constant	(2) Migrant (move recorded)	645	(6) Migrant (move recorded)	6,480
Nationality varies over time, while residence is constant	(3) Double nationality migrant (move not recorded)	238	(7) Double nationality migrant (move not recorded)	8,020
Both residence and nationality vary over time	(4) Double nationality migrant (move recorded)	140	(8) Double nationality migrant (move recorded)	2,963

Notes: The total number of immigrant inventors is 80,172, of which 11,361 have double nationality. The dataset records one (or more than one) move for 10,228 migrant inventors (cases 2, 4, 6, and 8). Time period: 1990–2011.

countries are members of the PCT, the overwhelming majority of the world population lives in PCT member states (including China, India, Indonesia, Brazil, and Nigeria). Some highly-populated countries are missing, though, notably Ethiopia, Bangladesh, and Pakistan. Second, the PCT attracts more than half of international patent applications. According to WIPO (2020), 56.7 percent of all international patent applications went through the PCT system. About 54 percent of the applications in 2020 originate from Asia, 22 percent from North America, and 22 percent from Europe. Taken together, these observations suggest that the PCT data are not well-suited to study high-skilled migration from Africa and South America. However, we note that we still capture migration events from non-member countries if nationals from these countries reside in PCT member states. For instance, a Pakistani inventor residing in the United States and filing a PCT patent will appear in the data as a Pakistani citizen with a U.S. residential address.

Although PCT data offer a unique opportunity to track the international movement of inventors, several additional features need to be highlighted. First, inventors in the raw PCT data are not uniquely identified, such that we cannot directly track their movements over time (an issuesimilar to thatfaced by Miguélez et al., 2010). To overcome this limitation, we have developed a unique machine learning disambiguation algorithm and have applied it to the PCT data, as explained at length in Appendix A. The method reaches a (cross validation) precision of about 95 percent and recall of about 80 percent. Second, migration events are imperfectly observed because the country of citizenship listed in the patent document is not equivalent to the country of birth. For instance, it may well be the case that an inventor appearing in the database as a non-migrant U.S. inventor may be an Indian-born inventor who pursued her education in the United States, obtained U.S. citizenship, and started patenting under her U.S. citizenship for the first time. We will estimate the regressions models on different data slices to alleviate these concerns to the extent possible. Third, as already alluded to, not all patent applications are filed under the PCT; international patent applications can also be filed under the Paris convention. Consequently, our data are not representative of the universe of patent applications, but we note that very few (if any) studies exploiting patent data are. We will propose robustness tests to understand potential selection effects.

2.2. Overview of migration data

The total number of disambiguated inventors, for which we have complete information regarding their country of residence and nationality, is 871,129. Defining the migration status of inventors from patent data is not trivial, as several cases can arise. We group inventors into eight categories, as shown in Table 1. The table provides a general overview of the entire sample of inventors (observed for the period 1990–2011) resulting from the disambiguation procedure broken down by the migration status of inventors.

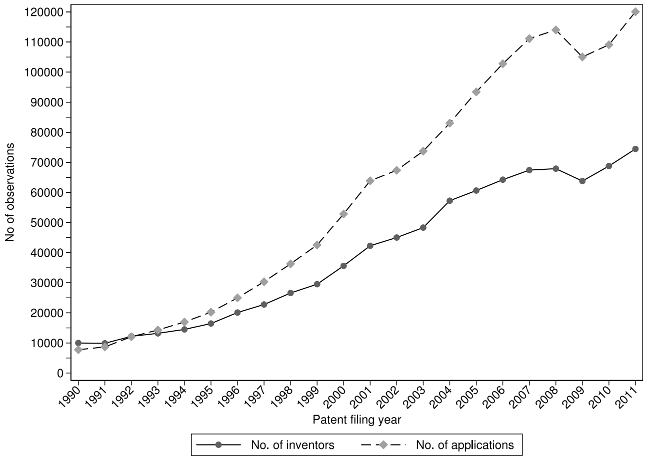

Fig. 1. Number of PCT applications and inventors by patent filing year.

Not surprisingly, the vast majority of inventors in the sample are not migrants. Indeed, 790,957 inventors, representing more than 90.8 percent of the total sample, have continuously resided in their country of nationality over the observed period (case 5). The remaining 80,172 inventors (9.2% of the total sample) are considered migrants, and we classify them into two broad groups. First, migrant inventors with double nationality (or naturalized), accounting for 14 percent of the total migrant inventors (cases 3, 4, 7 and 8; 11,361 observations). Second, migrant inventors with single nationality, amounting to 68,811 observations (cases 1, 2 and 6). As shown in Table 1, the dataset records one (or more than one) move for 10,228 migrant inventors (cases 2, 4, 6 and 8; around 13% of the migrant group). These cases correspond to situations where we observe the same inventors in two patent applications with different residence/nationality status. We do not observe the actual international move (*i.e.*, change in the country of residence) for the majority of migrant inventors (cases 1, 3 and 7, amounting to 87% of the migrant group). These inventors migrated prior to their first PCT patent application.

Fig. 1 shows the evolution of the number of inventors and applications over the period 1990–2011. From the second half of the 1990s, there is a notable increase in both the number of inventors and the number of applications. The growth reflects primarily an uptake of the PCT system, rather than a burst in inventiveness (Danguy et al., 2013). Growth temporarily halts in 2008, presumably as a result of the Global Financial Crisis.

Fig. 2 focuses on immigrant inventors. It depicts the number (left axis) and percentage (over the total sample of inventors, right axis) of immigrant inventors by patent filing year. The migration of inventors

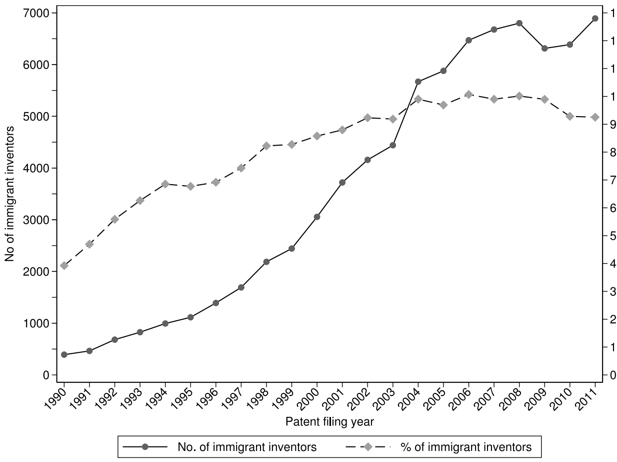

Fig. 2. Number and percentage of immigrant inventors by patent filing year.

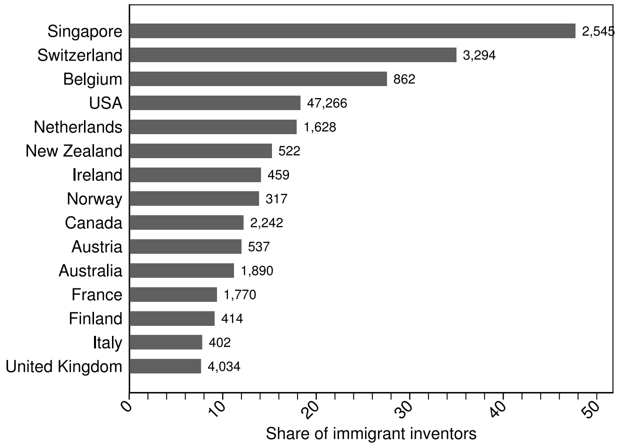

Fig. 3. Top-15 countries per share of immigrant inventors over total resident inventors. Note: Absolute numbers showed next to each bin. The same inventor can appear in more that one bin if s/he moves multiple times.

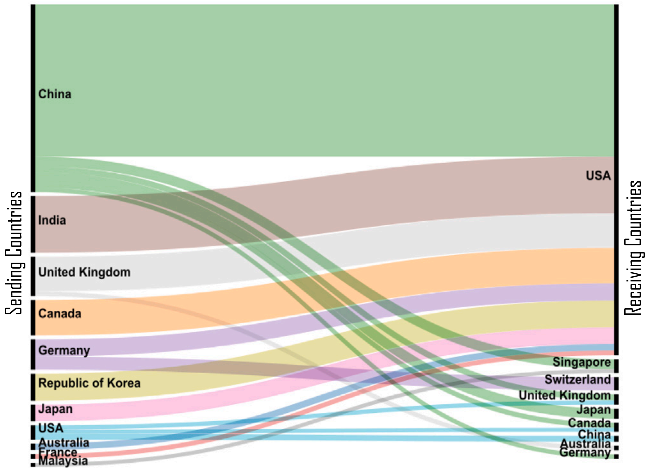

Fig. 4. Most populated corridors for the total sample of 80,172 migrant inventors.

appears to be a growing phenomenon, both in terms of absolute numbers and as a fraction of the total sample of inventors. Migrant inventors account for about ten percent of all inventors listed in PCT applications since the mid-2000s. The figure has reached a plateau since then.

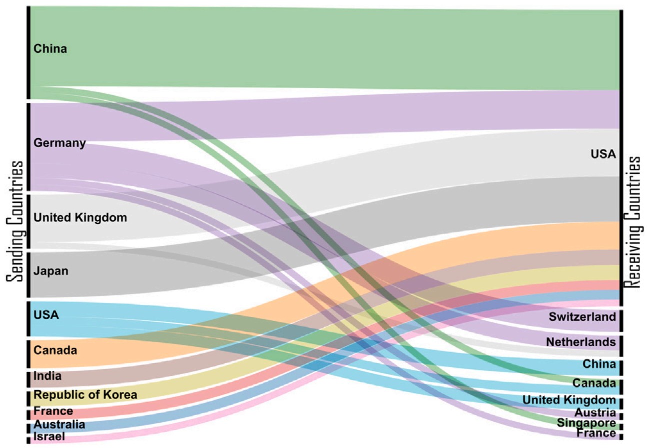

Fig. 5. Most populated corridors for the sample of 10,228 migrant inventors for which we observe an international move.

We do not claim that the migration figures we report are representative of the population of patenting inventors, let alone high-skilled migrants in general. The PCT procedure is but one way of filing patents. As mentioned previously, PCT applications target the international market and are thus presumably of higher economic value than purely national applications—not all inventions meet the profitability threshold required to cover the cost of the international patenting process (Harhoff et al., 2003; de Rassenfosse and van Pottelsberghe, 2013). If migrant inventors were more likely to produce high-value inventions than non-migrant inventors, then the migration figures we report would be inflated. It is not clear that this is the case (Guellec and van Pottelsberghe, 2000), but we are cautious to not extrapolate the findings outside the population of PCT patenting inventors.

2.3. Migration flows

Disambiguated inventor data enable the tracking of inventors across countries. Fig. 3 lists the top 15 receiving countries ranked according to the proportion of immigrant inventors over the total number of inventors residing in that particular country. For each of the 15 countries we also report the absolute number of immigrant inventors. Singapore, Switzerland and Belgium stand out with, respectively, 48 percent, 35 percent and 26 percent of their resident inventors being a foreign national. The remaining countries show significantly lower shares, ranging from 18 percent for the Netherlands and the United States to seven percent for Italy. As expected, the United States is, by far, the country with the largest pool of foreign inventors. It ranks fifth in terms of share of immigrant inventors, but it records more than 47,000 foreign nationals that have filed at least one PCT patent application during the period 1990–2011.

Fig. 4 illustrates the most frequent immigration corridors. The starting point (left-hand side) represents the country of first nationality and the end point (right-hand side) represents the country of residence. It emerges that inventor migration is a phenomenon extremely concentrated among a relatively limited number of receiving countries. The majority of migrants reside in the United States, followed by Singapore, Switzerland, the United Kingdom and Japan. On the other hand, outward migration is more fragmented. China and India represent the two most important sending countries, followed by the United Kingdom, Canada and Germany. Note that Fig. 4 relies on the total sample of 80,172 immigrant inventors. However, as previously pointed out, for most of these inventors, we do not actually observe any variation in their country of residence during the study period.

To provide some evidence on the mobility events included in our dataset, Fig. 5 reports the most frequent immigration corridors of the

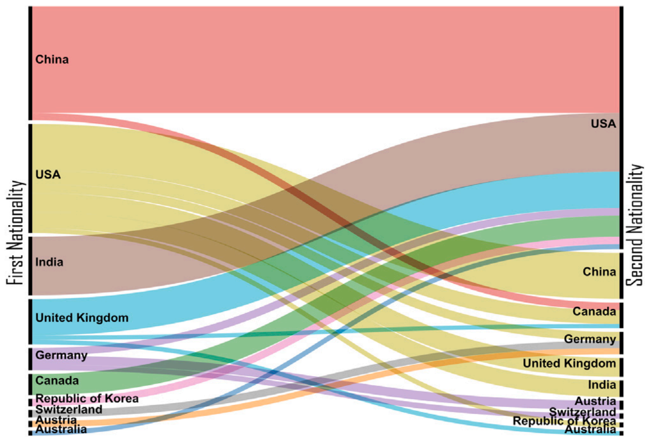

Fig. 6. Most frequent change or acquisition of nationalities.

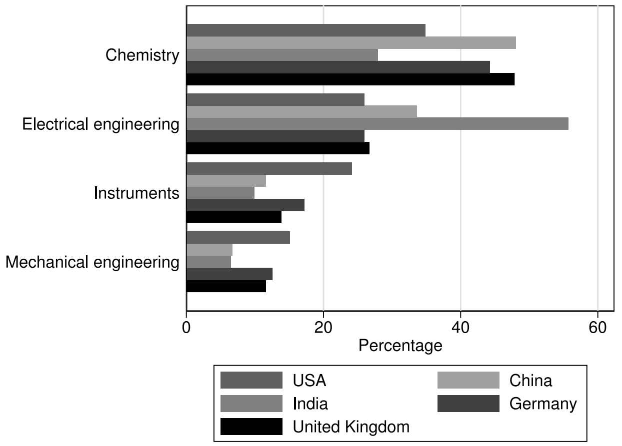

Fig. 7. Technology fields of U.S. inventors alongside those whose country of origin is one of the top-4 'sending' to the United States. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

10,228 migrant inventors for which we observe the actual move. China and the United States remain the first sending and receiving countries, respectively. However, some interesting differences with respect to the previous figure emerge. India loses relevance as a sending country in favor of other countries, such as Germany, the United Kingdom and the United States. In particular, our dataset records a notable number of moves by German inventors, with the United States, Netherlands and Switzerland being the preferred destinations.

Comparing Figs. 4 and 5 provides some insights on the potential reason for the move. We notice that German inventors typically move to the United States after their first PCT patent application, whereas Indian inventors typically move before their first PCT application. This could indicate that Indians usually come to the United States for education purposes and then stay to become inventors. An alternative explanation – though less likely in our opinion – is that Indians were already inventing at home, but they were simply not filing PCT patents.[7] Our data are silent on these issues.

As Table 1 shows, a sizable proportion of migrant inventors in our sample (around 14%) have a double nationality or have acquired a new nationality over the course of the study period. To further characterize changes in citizenship, Fig. 6 depicts the most frequent cases of inventor's change (or acquisition) of nationality. The most common cases refer to Chinese, Indian, U.K., German and Canadian inventors changing or acquiring U.S. nationality. On the other hand, many inventors from the United States have also experienced some changes in their citizenship status in favor of various countries such as China, Germany, the United Kingdom, Canada and India. As already mentioned, it is important to note that the data do not allow us to discriminate between cases of pre-existing double nationality, naturalization or acquisition of new nationality. However, a more careful inspection of the data reveals that most of the names and surnames of the American citizen inventors who, at some point, have filed a patent declaring Chinese or Indian citizenship, are typical Chinese and Indian names (such as Zhang Lu and Agrawal Avneesh). Thus, most of these inventors may be Chinese and Indian overseas born in the United States or educated in the United States.

Migrant inventors may have different areas of technological specialization than locals. Fig. 7 provides descriptive evidence in this respect. It reports the distribution of PCT applications by main technology area for the United States (the country with the largest absolute number of foreign inventors) and the top-4 giving countries to the United States, namely China, India, Germany and the United Kingdom.

Following Schmoch (2008) we identify four main areas of technology: Chemistry, Electrical engineering, Instruments, and Mechanical engineering.[8] British, German and Chinese inventors migrating to the United States are more frequently found in chemistry compared to the baseline rate of about 35 percent for U.S. inventors. Along the same lines, Indian and Chinese inventors migrating to the United States are more frequently found in electrical engineering, compared to the baseline rate of about 25 percent for U.S. inventors. On the other hand, U.S. inventors are relatively more numerous than immigrant inventors in the fields of instruments and mechanical engineering.

3. Migration and productivity

3.1. Empirical strategy

Our data on disambiguated inventors listed on PCT applications offer a unique opportunity to study the productivity effect of migration. We are able to track the patenting activity of inventors over time and across countries. We estimate the following inventor-level panel regression model with inventor fixed effects:

$$E[y_{i,t}|\mathrm{AfterMove}_{i,t}, X_{i,t}] = \exp(\beta_1 \mathrm{AfterMove}_{i,t} + \delta_i + \delta_t + \beta_2 \mathrm{X_{i,t}}), \qquad (1)$$

where the dependent variable $y_{i,t}$ identifies the number of patent families filed by inventor *i* in year *t*. We model it as an exponential function of the covariates following Hausman et al. (1984). Patent filing strategies may vary across jurisdictions, with applicants in some countries such as Japan filing multiple narrow patents and applicants in other countries such as the United States filing broad patents. To account for these institutional differences, we count the number of (INPADOC) patent families that the inventor has contributed to rather than the number of individual patents (de Rassenfosse et al., 2013). For simplicity, we will use the term 'patent' and not 'patent family' in the remainder of the document.

The variable of interest, $AfterMove$, is a binary indicator that takes value 1 starting from the year preceding the first 'move' of the migrant inventor *i* (as observed in patent documents). The one-year lag, previously used by Singh and Agrawal (2010), reflects the fact that an inventor is unlikely to produce a patentable invention right after her move. It is also consistent with survey evidence presented in Appendix E in de Rassenfosse and Jaffe (2018). This Appendix shows a histogram

---

[7] As explained below, the empirical analysis will control for potential difference in the propensity to file PCT patents with the use of country of origin fixed effects among other ways.

[8] We exclude some residual technology areas accounting for less than four percent of the sample.

on the time between initial expenditure on R&D and the first patent filing for a sample of patent applicants at the European Patent Office. Most patents are filed within one year of the start of the R&D project.

The coefficient β_1 provides an indication of whether, and to what extent, inventors who move from one country to another become more productive. If the PCT data do not capture the move but we know that the inventor is a migrant (cases 1, 3, and 7 in Table 1), then the variable $AfterMove$ takes value 1 for all t's.[9]

The variable δ_i represents the inventor fixed effects. It accounts for a potential selection effect by capturing unobservable, time invariant individual characteristics that may cause variation in patenting activity across different inventors. The use of inventor fixed effects implies that the estimates exploit within-variations in the variable of interest. Thus, identification will come only from observations corresponding to inventors for which we actually observe a move. Next, δ_t includes a vector of year fixed effects that control for systematic variations in patenting activity over time (*e.g.*, to capture the impact of the global financial crisis).

Finally, the vector X controls for confounding variables at the inventor level. First, it includes a complete set of country of inventor residence fixed effects. As explained by Griliches (1990), not all inventions are patented and patents are, therefore, an imperfect proxy of inventive output. We will obtain unbiased estimates of our variable of interest as long as variations in the propensity to patent are random with respect to the migrant status of inventors. One systematic source of variation in the propensity to patent relates to inventor's country of residence. Firms in developed countries tend to rely more on the PCT than firms in developing countries. The inclusion of country fixed effect controls for the possibility that, say, an Indian inventor who moves to the United States would be seen as more productive after her move simply because her company in India did not patent frequently—or applied for non-PCT patents.[10]

Second, following Moser et al. (2014), we control for possible variations in productivity over the life cycle of an inventor by constructing a variable that records the number of years that have passed from the last PCT patent application filed by inventor i. Third, in order to control for the size of the inventor's collaborative network, we build a variable that measures the average number of inventors that have collaborated in the inventive process with inventor i in year t (as captured by the number of inventors listed in the patent applications). Fourth, we control for the main technology class associated with patent applications by inventor i in year t, by including the set of 634 technological categories based on the Cooperative Patent Classification (CPC at 4-digit level).[11]

The initial sample is composed of all 871,129 inventors, who have filed a total of 2,597,315 PCT applications over the study period (see Table B.1 in Appendix). However, because we estimate equation (1) with fixed effects, a total of 598,429 inventors that have filed patent(s) in only one year are dropped from the original sample, forming our regression sample. Panel B of Table B.1 in Appendix reports basic descriptive statistics for the regression sample (N = 272,700 inventors and 1,890,897 applications). This restriction increases the balance between the sample of non-immigrant and immigrant inventors. Finally, because not observing a patent should not be treated as a missing value but as a zero, we create a balanced panel database by populating the missing data points at inventor/year level with zeros.

3.2. Results

Baseline findings

Table 2 presents the results of the baseline estimates obtained by applying the pseudo-maximum-likelihood Poisson technique to takes into account the count data nature of the dependent variable.[12] Column (1) reports the most parsimonious regression model. It shows a positive and statistically significant relationship between the variable $AfterMove$ and the number of patent applications filed by inventor i in year t. More specifically, the point estimate implies that inventors who move from one country to another show a sensible increase in their productivity of about 0.23 additional patent application filed each year. The gradual inclusion of the other control variables in columns (2)–(5) reduces the magnitude of this point estimate to about 0.07 in column (5), our preferred specification. Compared with an average of 0.30 patents per inventor per year between 1990–2011, this figure implies a twenty-three percent increase in patenting by immigrant inventors following their move. We add the time elapsed since the last patent application in column (2), country of residence fixed effects in column (3), year fixed effects in column (4), and technology field fixed effects in column (5).

The econometric specification is robust to selection of more productive inventors being more likely to migrate. Indeed, the fixed-effect model allows the variable δ_i to be correlated with the covariates. Thus, the fact that more productive inventors may also be more likely to move is not a concern in our setup. Nevertheless, one could argue that a migration event is such an important event that it fundamentally alters inventor's unobserved characteristics, *i.e.*, the inventor fixed effect may not be 'fixed'. Capturing such productivity shock is precisely the role of the variable $AfterMove$. Yet, we provide an alternative modeling strategy in column (6). We estimate the inventor fixed effects using only pre-move information. The predicted fixed effects are then included in an pseudo-maximum-likelihood Poisson estimations model that covers both the pre-move and the post-move period. The coefficient of interest reaches 0.088, which provides a upper-bound estimate for the effect of migration.[13]

The post-move inventor's productivity level may also be influenced by the inherent characteristics of the employer. Migrant inventors may be more likely to be employed by more productive and financially sound companies, which could positively affect their post-move propensity to patent. To control for this possibility, we extend the baseline specification by including a complete set of patent assignee fixed effects.[14] As shown in column (7) of Table 2, the inclusion of this additional set of fixed effects only slightly reduces the magnitude of the coefficient of interest.

Alternative specifications

The results reported in Table 2 are robust to a broad range of different specifications.

First, we test the sensitivity of the main results when considering different time periods. Indeed, Fink and Miguélez (2017) show that the coverage of inventors' residence and nationality information provided by PCT data increases substantially after 2004, with a change to PCT rules providing that all PCT applications automatically include all PCT member states as designated states, including the United States (for

---

[9] In an alternative specification, we will exclude inventors for which we do not record the move.

[10] In addition, in a robustness test, we impose that inventors have at least one PCT patent prior to migrating. Many firms do not file for patents but firms that do tend to do so frequently. Imposing at least one patent prior to migrating allows us to identify patent-active inventors.

[11] We identify the main technological class by considering the most frequent CPC (at 4-digit) class per inventor i in year t.

[12] We have used the Stata command *ppmlhdfe*, which allows to include multi-way fixed effects.

[13] To run the model in column (6) of Table 2, we exclude all inventors for which we do not have information on the year in which they move.

[14] The information on patent assignees provided by the WIPO dataset is of good quality. Nevertheless, we have further improved the assignees name harmonization by performing an extensive manual inspection to thoroughly check for the presence of typos, variations and abbreviations in company names.

Table 2
Pseudo-maximum-likelihood Poisson estimations.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
After Move	0.233*** (0.009)	0.273*** (0.008)	0.284*** (0.009)	0.107*** (0.005)	0.077** (0.005)	0.088*** (0.005)	0.069*** (0.004)
Inv. fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Av. no of inv.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time since last patent	No	Yes	Yes	Yes	Yes	Yes	Yes
Countr. of Res. fixed effects	No	No	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	No	No	Yes	Yes	Yes	Yes
CPC fixed effects	No	No	No	No	Yes	Yes	Yes
Inv. fix. eff. (before move)	No	No	No	No	No	Yes	No
Assignee fixed effects	No	No	No	No	No	No	Yes
Number of observations	5,999,400	5,999,400	5,999,400	5,999,400	5,999,400	5,994,582	5,994,582
Number of migrant inventors	33,865	33,865	33,865	33,865	33,865	33,646	33,646
R^2	0.11	0.17	0.17	0.30	0.32	0.35	0.42

Notes: The dependent variable measures the number of patents filed by inventor *i* in year *t*. The variable $AfterMove$ is a binary indicator that takes value one starting from the year preceding the first move of the migrant inventor *i*. The sample is composed of cases 1 to 8 as defined in Table 1. Coefficients are expressed as average marginal effects (AME). Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

Table 3
Pseudo-maximum-likelihood Poisson estimations: Different time periods.

	(1990–2011)	(1990–2003)	(2004–2011)	(2000–2011)
	(1)	(2)	(3)	(4)
After Move	0.077*** (0.005)	0.101*** (0.007)	0.087*** (0.003)	0.104*** (0.009)
Inv. fixed effects	Yes	Yes	Yes	Yes
Av. no of inv.	Yes	Yes	Yes	Yes
Time since last patent	Yes	Yes	Yes	Yes
Countr. of Res. fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
CPC fixed effects	Yes	Yes	Yes	Yes
Number of observations	5,999,400	3,817,800	2,181,600	3,272,400
Number of migrant inventors	33,865	33,865	33,865	33,865
R^2	0.32	0.30	0.33	0.29

Notes: The dependent variable measures the number of patents filed by inventor *i* in year *t*. The variable $AfterMove$ is a binary indicator that takes value one starting from the year preceding the first move of the migrant inventor *i*. The sample is composed of cases 1 to 8 as defined in Table 1. Coefficients are expressed as average marginal effects (AME). Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

which information on inventors' residence and nationality is systematically requested, see Section 2.1). Accordingly, we estimate the full specification on two distinct time periods: from 1990 to 2003 and from 2004 to 2011. The estimated coefficients of these two models are reported in Table 3 columns (2) and (3). For comparative purposes, we also report the coefficient of Table 2, column (5), in the first column. The sign and significance of the coefficient of interest are in line with the baseline results. We also extend the second time period to 2000–2011 in column (4), with quantitatively similar results.

Second, we test the robustness of our main results to the definition of immigrant inventors. In particular, we re-estimate equation (1) by considering two restrictive definitions of immigrant inventors. In the first case, reported in Table 4, we exclude from the sample inventors with a declared double nationality (categories 3, 4, 7 and 8 in Table 1). We are concerned by the possibility that some inventors with double nationality may not be immigrants, but rather second-generation migrants or native inventors who have acquired another nationality. The coefficient of interest remains practically unchanged (0.069). In the second case, reported in Table 5, we use an even more stringent definition of immigrant inventors. We exclude immigrant inventors for which we cannot observe any actual move (*i.e.*, change in country of residence, cases 1, 3 and 7 in Table 1). Once again, the results confirm the main findings, with the coefficient of interest reaching 0.088.

In a final specification, we have restricted the sample to patents in the field of chemistry, in which patent protection is known to be an effective appropriation mechanism (Cohen et al., 2000). The propensity to patent is higher in this field than in others, and patent data, therefore, measure an inventor's output more accurately. The results, reported in Table B.2 in Appendix B, confirm the positive impact of migration on productivity (coefficient about 0.05 in column 5).

Taken together, the results confirm the presence of a boost to productivity following a move. In our preferred specification, inventors experience a twenty-three-percent increase in productivity.

4. Additional analyses

4.1. Exploring country-level heterogeneity

As mentioned in Section 2, one of the distinctive features of the PCT database is its international coverage. In this subsection, we take advantage of this unique feature of our database and explore heterogeneity at the country level.

Disaggregation by receiving and sending country

We first run the main model on a selected number of receiving countries. We focus on the top five countries in terms of total number of applications filed by immigrants during the period 1990–2011, namely the United States, Germany, the United Kingdom, China, Switzerland, and all the remaining countries pooled together. Table 6 presents the estimation results separately for each country. In all the models, the coefficients are positive and statistically significant. They range between 0.062 for Great Britain and 0.175 for China. It would be erroneous to interpret the differences across countries as saying that some countries offer a more productive environment for migrant inventors than others. Indeed, the composition of migrant inventors, technological fields, and yearly number of patents per inventor differ from one sample to the next.

Table 4
Pseudo-maximum-likelihood Poisson estimations: Excluding immigrant inventors with double nationality.

	(1)	(2)	(3)	(4)	(5)
After Move	0.214***	0.261***	0.273***	0.095***	0.069***
	(0.013)	(0.013)	(0.013)	(0.007)	(0.003)
Inv. fixed effects	Yes	Yes	Yes	Yes	Yes
Av. no of inv.	Yes	Yes	Yes	Yes	Yes
Time since last patent	No	Yes	Yes	Yes	Yes
Countr. of Res. fixed effects	No	No	Yes	Yes	Yes
Year fixed effects	No	No	No	Yes	Yes
CPC fixed effects	No	No	No	No	Yes
Number of observations	5,751,152	5,751,152	5,751,152	5,751,152	5,751,152
Number of immigrant inventors	22,581	22,581	22,581	22,581	22,581
R^2	0.10	0.16	0.17	0.30	0.33

Notes: The dependent variable measures the number of patents filed by inventor *i* in year *t*. The variable $AfterMove$ is a binary indicator that takes value one starting from the year preceding the first move of the migrant inventor *i*. The sample is composed of cases 1, 2, 5, and 6 as defined in Table 1. Coefficients are expressed as average marginal effects (AME). Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

Table 5
Pseudo-maximum-likelihood Poisson estimations: Excluding immigrant inventors for which we do not observe the actual move.

	(1)	(2)	(3)	(4)	(5)
After Move	0.226***	0.253***	0.263***	0.109***	0.088***
	(0.026)	(0.026)	(0.022)	(0.011)	(0.005)
Inv. fixed effects	Yes	Yes	Yes	Yes	Yes
Av. no of inv.	Yes	Yes	Yes	Yes	Yes
Time since last patent	No	Yes	Yes	Yes	Yes
Countr. of Res. fixed effects	No	No	Yes	Yes	Yes
Year fixed effects	No	No	No	Yes	Yes
CPC fixed effects	No	No	No	No	Yes
Number of observations	5,478,462	5,478,462	5,478,462	5,478,462	5,478,462
Number of migrant inventors	10,186	10,186	10,186	10,186	10,186
R^2	0.10	0.16	0.16	0.30	0.32

Notes: The dependent variable measures the number of patents filed by inventor *i* in year *t*. The variable $AfterMove$ is a binary indicator that takes value one starting from the year preceding the first move of the migrant inventor *i*. The sample is composed of cases 2, 4, 5 6 and 8 as defined in Table 1. Coefficients are expressed as average marginal effects (AME). Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

Next, we replicate the previous exercise by focusing on the top five sending countries (i.e., nationalities). We re-estimate equation (1) by focusing, in turn, on Chinese, American, German, British, and Indian immigrant inventors. As in the previous case, we also estimate the model on a sample that pools all the remaining migrant inventors together. The results, reported in Table 7, are fairly consistent across countries. The coefficient of interest ranges from 0.022 for Chinese inventors to 0.077 for Indian inventors. Again, it would be erroneous to interpret Table 7 as saying that some migrant inventors are more productive than others—inventors specialize in different fields and move to different countries such that comparison across countries is not warranted.

Accounting for productivity differences within country

The regression models estimated so far are quite demanding, with a large number of control variables, fixed effects, and data slicing. They allow us to rule out several explanations for the productivity increase such as moving to a more productive country. All specifications point to the conclusion that migrant inventors experience a genuine increase in productivity following a move.

In this section, we test the extent to which these productivity gains might be driven by inventors moving to particularly productive regions. To illustrate, if all migrant inventors in the United States were moving to the Silicon Valley, we might be capturing a Silicon Valley effect rather than a general migration effect. To address this concern, we focus on the United States as a receiving country (for the sheer size of the country and because we have fine-grained location data) and control for a full set of 1,114 county dummies. We have been able to extract the information for 54,819 inventors that have resided in the United States for at least one year.

Table 8 reports the estimated coefficients. Column (5) presents the same regression model than that in column (1) of Table 6; except that sample size is now smaller because of data availability on inventor regional location. The coefficient of interest has a similar magnitude (0.085 vs. 0.094) such that we are confident that the difference in sample sizes does not drive our results. Controlling for regional dummies, column (6), leaves the coefficient essentially unchanged.

4.2. Dealing further with selection

One possible remaining threat to the validity of our estimates relates to the fact that immigrant inventors may be inherently more productive than non migrant inventors. The use of longitudinal data together with the inclusion of inventor fixed effects allow us to account, at least partially, for the existence of positive selection among migrant inventors. Furthermore, the use of country of residence and assignee fixed effects account for the possibility that migrant inventors became more productive simply because they moved to a better environment. However, there may be other pre-move factors that may play a decisive role in determining the post-move productivity differentials experienced by migrant inventors.

In this section we carry out an additional test to control further for some potential selection issues. We seek to reduce the heterogeneity among migrant and non-migrant inventors by adopting a matching strategy. We create two groups of treated (migrant) and control (non-migrant) inventors by exact matching on some relevant pre-move inventor-level characteristics. More in details, we consider: the cumulative number of patents filed by inventor i up to the year of the move; the filing year of her first patent, her country of residence before move, and the CPC (4 digit) mode of her patent portfolio. As shown

Table 6

Pseudo-maximum-likelihood Poisson estimations: Selected receiving countries.

	US	DE	GB	CN	CH	Other countries
	(1)	(2)	(3)	(4)	(5)	(6)
After Move	0.094***	0.106***	0.062***	0.175***	0.101***	0.062***
	(0.003)	(0.012)	(0.008)	(0.011)	(0.009)	(0.004)
Inv. fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Av. no of inv.	Yes	Yes	Yes	Yes	Yes	Yes
Time since last patent	Yes	Yes	Yes	Yes	Yes	Yes
Countr. of Res. fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
CPC fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	1,846,427	685,094	328,823	307,752	63,465	1,565,804
Number of immigrant inventors	22,268	3,354	2,870	2,008	1,669	11,867
R^2	0.31	0.36	0.33	0.47	0.40	0.36

Notes: The dependent variable measures the number of patents filed by inventor i in year t. The variable $AfterMove$ is a binary indicator that takes value one starting from the year preceding the first move of the migrant inventor i. The sample is composed of cases 1 to 8 as defined in Table 1. Coefficients are expressed as average marginal effects (AME). Robust standard errors in parentheses.

***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

Table 7

Pseudo-maximum-likelihood Poisson estimations: Selected sending countries.

	CN	US	DE	GB	IN	Other countries
	(1)	(2)	(3)	(4)	(5)	(6)
After Move	0.022***	0.038***	0.048	0.076***	0.077***	0.075***
	(0.006)	(0.008)	(0.030)	(0.005)	(0.010)	(0.004)
Inv. fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Av. no of inv.	Yes	Yes	Yes	Yes	Yes	Yes
Time since last patent	Yes	Yes	Yes	Yes	Yes	Yes
Countr. of Res. fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
CPC fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	5,431,553	5,352,105	5,324,769	5,332,236	5,311,453	5,519,134
Number of immigrant inventors	9,500	8,902	3,838	4,490	3,049	14,437
R^2	0.32	0.32	0.32	0.32	0.33	0.33

Notes: The dependent variable measures the number of patents filed by inventor i in year t. The variable $AfterMove$ is a binary indicator that takes value one starting from the year preceding the first move of the migrant inventor i. The sample is composed of cases 1 to 8 as defined in Table 1. Coefficients are expressed as average marginal effects (AME). Robust standard errors in parentheses.

***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

Table 8

Pseudo-maximum-likelihood Poisson estimations: Controlling for regional heterogeneity (United States only).

	(1)	(2)	(3)	(4)	(5)	(6)
After Move	0.257***	0.297***	0.298***	0.111***	0.085***	0.083***
	(0.008)	(0.007)	(0.007)	(0.004)	(0.003)	(0.003)
Inv. fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Av. no of inv.	Yes	Yes	Yes	Yes	Yes	Yes
Time since last patent	No	Yes	Yes	Yes	Yes	Yes
Countr. of Res. fixed effects	No	No	Yes	Yes	Yes	Yes
Year fixed effects	No	No	No	Yes	Yes	Yes
CPC fixed effects	No	No	No	No	Yes	Yes
Regional fixed effects	No	No	No	No	No	Yes
Number of observations	1,206,018	1,206,018	1,206,018	1,206,018	1,206,018	1,206,018
Number of immigrant inventors	13,637	13,637	13,637	13,637	13,637	13,637
R^2	0.13	0.19	0.19	0.34	0.37	0.37

Notes: The dependent variable measures the number of patents filed by inventor i in year t. The variable $AfterMove$ is a binary indicator that takes value one starting from the year preceding the first move of the migrant inventor i. The sample is composed of cases 1 to 8 as defined in Table 1. Coefficients are expressed as average marginal effects (AME). Robust standard errors in parentheses.

***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

in Panel C of Table B.1, this matching procedure reduces substantially the heterogeneity between the two groups of migrant and non-migrant inventors.

Table 9 reports the estimation results obtained using the matched sample. The coefficients of the variable of interest are positive and statistically significant in all the models, and they are similar in magnitude to the coefficients of the baseline model. These results provide additional evidence that migrant inventors seem to enjoy a genuine boost in productivity.

5. Conclusions

Existing literature has generally provided evidence that migrant inventors are more productive than their non-migrant peers. This paper moves the field forward by showing that migrant inventors seem to become more productive after they have migrated.

Overall, we find that migrants enjoy a twenty-three-percent gain in productivity following a move. This result is robust to a range of specifications and can be observed across destination and sending

Table 9
Pseudo-maximum-likelihood Poisson estimations: Matched sample.

	(1)	(2)	(3)	(4)	(5)
After Move	0.240*** (0.012)	0.300*** (0.011)	0.302*** (0.011)	0.118*** (0.006)	0.113*** (0.004)
Inv. fixed effects	Yes	Yes	Yes	Yes	Yes
Av. no of inv.	Yes	Yes	Yes	Yes	Yes
Time since last patent	No	Yes	Yes	Yes	Yes
Countr. of Res. fixed effects	No	No	Yes	Yes	Yes
Year fixed effects	No	No	No	Yes	Yes
CPC fixed effects	No	No	No	No	Yes
Number of observations	2,595,274	2,595,274	2,595,274	2,595,274	2,595,274
Number of immigrant inventors	18,413	18,413	18,413	18,413	18,413
R^2	0.11	0.16	0.17	0.32	0.34

Notes: The dependent variable measures the number of patents filed by inventor *i* in year *t*. The variable $AfterMove$ is a binary indicator that takes value one starting from the year preceding the first move of the migrant inventor *i*. The sample is composed of cases 1 to 8 as defined in Table 1. Coefficients are expressed as average marginal effects (AME). Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

countries. We have arrived at this conclusion thanks to the careful disambiguation of the approximately 1 million inventors listed in PCT patent documents using a machine-learning-based approach. These data have allowed us to track the migratory movements of inventors over time. An initial analysis of the disambiguated data reveals that migrant inventors account for more than nine percent of the population of inventors since the early 2000s. However, the majority of migrant inventors are already residing in the host country *prior* to their first PCT patent—for example, because they may have migrated for education.

We have arrived at the conclusion of a productivity boost by exploiting panel data with inventor fixed effects, and also considering matching estimators. As to the reason for the boost, the econometric analysis considers – and has ruled out – several potential explanations. We have investigated the possibility that the effect is caused by differences in propensity to patent or productivity across countries, firms, and regions within a country. We have also accounted for changes in productivity over inventors' lifecycle as well as changes in the collaboration network of inventors. Although the effect is very robust (also holding under various alternative specifications and data slicing), we note that the inherent nature of our data does not allow us to establish a strong causal relationship between migration and the productivity of inventors. We do not have a shock that 'forces' migration or an instrument that induces as-if random variations in migration events. Consequently, we cannot rule out the possibility that inventors decided to move because they knew they would be more productive for some reason that we cannot observe. Nevertheless, the empirical finding remains.

We discuss four possible explanations as to *why* migrant inventors become more productive after they have moved. First, migrants may work harder given their (presumably) more precarious visa status. Should this be the case, the mobility effect we observe would diminish over time. A specification that controls for the effect of the passing of time left the coefficient of interest essentially unchanged, suggesting a different cause for the effect.[15] Second, migrant inventors may simply move to a more productive environment relative to their home country. This situation could arise, for example, because migrant inventors work in labs that have better equipment than at home. However, we note that the econometric regression models broadly account for this explanation with the inclusion of country of residence, region, and assignee fixed effects. Leaving two other potential mechanisms responsible for this productivity increase: human capital upgrading and better fit.

By moving to a new country, migrants are exposed to new knowledge, new colleagues, new practices, representing opportunities to upgrade their human capital. Alternatively, skilled inventors usually migrate to a new country by choice, possibly with the knowledge that they will be more productive—for example, because they know they will fit particularly well in the new environment (Jovanovic, 1979). In the same vein, migrants might have escaped situations where they were not in a position to exploit their creativity fully, for example, due to ethnic or gender considerations (e.g., Sugimoto et al., 2015). Our data are silent on the individual motives for migration, and we leave it to further research to tease out these possible remaining mechanisms.

The analysis also offers policy implications. The central finding is that migration raises inventors' productivity and, consequently, increases the overall patenting activity worldwide. This leads us back to the tension between the risk of brain drain associated with liberal migration policies and the overall benefits of free migration through better human capital allocation. Do emigration countries lose precious inventors if they are welcomed elsewhere? If, as we have shown, migration by itself is the booster of productivity, then the brain drain effect appears relatively small. This explanation would speak in favor of the free movement of inventors to maximize global innovation outcomes.

CRediT authorship contribution statement

Gabriele Pellegrino: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software programming, Validation, Visualization, Writing – original draft, Writing – review & editing. **Orion Penner:** Data curation, Funding acquisition, Investigation , Methodology, Software programming, Validation, Writing – review & editing. **Etienne Piguet:** Writing – review & editing. **Gaétan de Rassenfosse:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We gratefully acknowledge the financial support of the Swiss Network for International Studies (SNIS). The paper has benefited from comments and suggestions from participants at the GEOINNO 2018 conference (Barcelona).

Appendix A. Disambiguation algorithm

As inventors receive no unique identifier in the PCT applications dataset, we must disambiguate all inventors in order to track migration flows. Inventor ambiguity may arise from two sources. First, many distinct inventors may possess the same name, thus creating uncertainty as to which application belongs to which individual. Second, each

---

[15] The results are available upon request from the authors.

inventor's name may not be reported in the same way across each of his or her patents, due to typographic error, transliteration error, incomplete reporting, *etc*. Our disambiguation approach follows in the vein of Fleming and colleagues in that it employs a (trained) classifier to predict which patents may belong to the same inventor (Balsmeier et al., 2015). Below, we outline the basic principles of the disambiguation approach developed and applied herein.

Before describing our disambiguation approach in detail, let us start with the lowest level element of the process – an inventor-patent application pair – that we will refer to as a disambiguation-id. For example disambiguation-id "2_US2011033658" refers to the second inventor (Ron Hadar) on the PCT application corresponding to patent document US2011033658. The final output of the entire process is a set of clusters of disambiguation-ids, each corresponding to the best estimate of the output of an individual inventor. The approach proceeds through the following steps:

1. Blocking
2. Feature vector calculation
3. Feature vector classification
4. Cluster extraction and violation resolution

A.1. Blocking

Blocking is the process by which we establish which pairs of disambiguation-id may end up clustered. For example, it is reasonable to believe that two disambiguation-ids belonging to an inventor named "John Smith" may, in fact, belong to the same inventor, and hence, are eligible to be assigned to the same cluster. On the other hand, it is not reasonable to think that a disambiguation-id corresponding to an inventor named "John Smith" and another corresponding to an inventor named "Borislav Trifonov" may belong to the same inventor, and hence are not eligible to be assigned to the same cluster.

In our disambiguation we take a hard line on blocking and do not consider typographic errors. For two disambiguation-ids to be "blocked" together (eligible for assignment to the same cluster) they must have the exact same family name. They must further have the exact same first given name, or at least the corresponding initial (*e.g.,* John and J are eligible). Similarly for second given names. We do not consider given names beyond the first and second. We implement such a strict blocking criterion due to the high quality of the underlying PCT data. It is important to understand that because we consider given name initials, blocking is not transitive *i.e.,* "John" and "J" will be blocked, as will "James" and "J" but *not* "John" and "James". Indeed, such cases are specifically resolved in Step 4 through violation resolution.

A.1.1. Feature vector calculation

After blocking is complete, we calculate a feature vector for each allowed pair of disambiguation-ids. The goal of each feature vector is to summarize the similarity, and dissimilarity, of the two disambiguation-ids across a number of dimensions. As these vectors will later be fed to a (trained) classifier, the exact functional form of each element is not critical. What is critical is that as much relevant information as possible is present for the classifier to use in order to discriminate between patent documents that belong to the same inventor, and those that do not. Herein, features include: the overlap of cited patents; a flag for whether one directly cites the other; various cosine similarity measures of the patent classes, various cosine similarity measures of the classes of the patents *cited* by each; overlap of assignees; overlap of assignee countries; flags for shared country of residence or nationality for the focal inventors; number of years between the two priority fillings; overlap between co-inventors on the patents; and overlap between countries of residence of the co-inventors.

A.2. Cluster extraction and violation resolution

Once the feature vectors are collected, we apply a previously trained classifier to them. The goal of the classifier is straightforward: to correctly select those disambiguation-id pairs that belong to the same inventor. Clusters are then extracted from that output in the next stage. The construction and training of the classifier does merit description however.

Our classifier is a Neural Network model implemented in Keras (with Theano backend). It consists of a dense linear input layer, four dense hidden layers with rectified linear unit activation (node counts decreasing from 20 to 5), and an output layer with sigmoid activation. It can be noted that, while this model slightly outperforms similar ones, we found that with due care and attention almost any neural network configuration can produce similar results.

To establish the training set we rely on an approach previously used to study samples of researchers (Milojević, 2013) and for similar disambiguation efforts (Balsmeier et al., 2015). Specifically, we use the set of given names for which only one, unique, given name appears for each. The logic of this approach being that if we observe a family name with only one specific given name, it is highly unlikely that the only two inventors in the entire data set with that family name would, coincidentally, also have to same given name. Thus to construct the true positive portion of our training set we collect the set of all family names for which only one unique given name appears and that appear on at least 5 patents. For each of those family names we then sample pairs of disambiguation-ids in a fashion proportional to the number of patents found for that family name. True negatives are much easier to find, as any pair of patents for which not a single pair of inventors shares a family can be very easily assumed to not have been invented by the same person. In the training set the ratio of true negative pairs to true positive pairs is 3:1 and the set, as a whole, consists of approximately 1.1 million disambiguation-id pairs. Once trained the classifier reaches a (cross-validation) precision of about 95% and recall of about 80%.

A.3. Feature vector classification

While each blocked disambiguation-id pair is assigned a value of 1 or 0 by the classifier, we are still left with the problem of extracting clusters from that data. Initially this is quite straightforward: we construct a network of disambiguation-ids (a 1 indicates an edge) and pull out the connected components. For each connected component, if there exists no violations of the kind mentioned above (*i.e.* "John" connected to "J" connected to "James") then that connected component is accepted as a cluster corresponding to an individual inventor. But if there are violations, we follow a secondary procedure. We apply a network community detection algorithm. Each community that does not contain a violation is taken as a cluster. Each community that does contain a violation is then recursively subjected to mincut until no violations remain.

As a sanity check on the results of our disambiguation approach, we cross reference, where possible, individual patents to the USPTO, and in turn, to the USPTO's in-house inventor disambiguation. While results do fluctuate across family names, we note a relatively high level of agreement with the USPTO (90+% precision, 75+% recall). But we also note that when disagreements arise between USPTO and our disambiguation, ours often ends up being correct following inspection. In particular we note that there are several inventor profiles in the USPTO disambiguation that contain many (some times more than 100) highly distinct given names.

Appendix B. Additional summary statistics and robustness checks

See Tables B.1 and B.2.

Table B.1
Summary statistics of sample used for the econometric regression.

	Total	*Non Immigrant*	*Immigrant*
Panel A: Whole sample			
No. of inventors	871,129	790,957	80,172
No. of applications	2,597,315	2,258,672	338,643
Average no. of applications x inventor	2.98	2.86	4.22
No. of applications with one inventor	662,627	582,666	79,961
No. of applications with at least two inventors	1,272,850	1,103,904	168,946
Panel B: Sample used for the main estimations			
No. of inventors	272,700	238,835	33,865
No. of applications	1,890,897	1,608,060	282,837
Average no. of applications x inventor	6.93	6.73	8.35
No. of applications with one inventor	559,407	480,681	78,726
No. of applications with at least two inventors	817,498	690,831	126,667
Panel C: Matched sample			
No. of inventors	117,967	99,590	18,377
No. of applications	858,629	722,103	136,526
Average no. of applications x inventor	7.28	7.25	7.43
No. of applications with one inventor	340,349	288,539	51,810
No. of applications with at least two inventors	274,824	229,211	45,613

Notes: Immigrant inventors are defined as inventors who have applied for at least one patent while residing in a country different from their country of nationality, see Table 1. Time period:1990–2011.

Table B.2
Pseudo-maximum-likelihood Poisson estimations: Focus on chemistry.

	(1)	(2)	(3)	(4)	(5)
After Move	0.195***	0.224***	0.222***	0.096***	0.051***
	(0.012)	(0.012)	(0.014)	(0.009)	(0.006)
Inv. fixed effects	Yes	Yes	Yes	Yes	Yes
Av. no of inv.	Yes	Yes	Yes	Yes	Yes
Time since last patent	No	Yes	Yes	Yes	Yes
Countr. of Res. fixed effects	No	No	Yes	Yes	Yes
Year fixed effects	No	No	No	Yes	Yes
CPC fixed effects	No	No	No	No	Yes
Number of observations	1,382,348	1,382,348	1,382,348	1,382,348	1,382,348
Number of migrant inventors	9,983	9,983	9,983	9,983	9,983
R^2	0.12	0.17	0.18	0.27	0.31

Notes: The dependent variable measures the number of patents filed by inventor i in year t. The variable $AfterMove$ is a binary indicator that takes value one starting from the year preceding the first move of the migrant inventor i. The sample is composed of cases 1 to 8 as defined in Table 1. Coefficients are expressed as average marginal effects (AME). Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

References

Bahar, D., Choudhury, P., Rapoport, H., 2020. Migrant inventors and the technological advantage of nations. Res. Policy 49 (9), 103947.

Balsmeier, B., Chavosh, A., Li, G.-C., Fierro, G., Johnson, K., Kaulagi, A., O'Reagan, D., Yeh, B., Fleming, L., 2015. Automated disambiguation of us patent grants and applications. Unpublished Working Paper, Fung Institute for Engineering Leadership.

Bertoli, S., Brücker, H., Facchini, G., Mayda, A.M., Peri, G., 2012. Understanding highly skilled migration in developed countries: The upcoming battle for brains. In: Boeri, T., Bruecker, H., Docquier, F., Rapoport, H. (Eds.), Brain Drain and Brain Gain. The Global Competition to Attract High-Skilled Migrants. Oxford University Press.

Borjas, G.J., 1994. The Economics of Immigration. J. Econ. Lit. 32 (4), 1667–1717.

Borjas, G.J., Doran, K.B., 2012. The Collapse of the Soviet Union and the Productivity of American mathematicians. Q. J. Econ. 1143–1203.

Breschi, S., Lawson, C., Lissoni, F., Morrison, A., Salter, A., 2020. STEM migration, research, and innovation. Res. Policy 49 (9), 104070.

Breschi, S., Lissoni, F., Miguelez, E., 2017. Foreign-origin inventors in the USA: Testing for diaspora and brain gain effects. J. Econ. Geogr. 17 (5), 1009–1038.

Caviggioli, F., Jensen, P., Scellato, G., 2020. Highly skilled migrants and technological diversification in the US and Europe. Technol. Forecast. Soc. Change 154, 119951.

Cohen, W.M., Nelson, R.R., Walsh, J.P., 2000. Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not). Tech. rep., National Bureau of Economic Research.

Danguy, J., de Rassenfosse, G., van Pottelsberghe de la Potterie, B., 2013. On the origins of the worldwide surge in patenting: an industry perspective on the R&D–patent relationship. Ind. Corp. Chang. 23 (2), 535–572.

European Commission, 2011. Horizon 2020—The Framework Programme for Research and Innovation. Tech. rep., Luxembourg.

Ferrucci, E., 2020. Migration, innovation and technological diversion: German patenting after the collapse of the soviet union. Res. Policy 49 (9), 104057.

Fink, C., Miguélez, E., 2013. Measuring the International Mobility of Inventors: A New Database. WIPO.

Fink, C., Miguélez, E., 2017. The International Mobility of Talent and Innovation. Cambridge University Press.

Franzoni, C., Scellato, G., Stephan, P., 2014. The mover's advantage: the superior performance of migrant scientists. Econom. Lett. 122 (1), 89–93.

Friedberg, R.M., Hunt, J., 1995. The impact of immigrants on host country wages, employment and growth. J. Econ. Perspect. 9 (2), 23–44.

Ganguli, I., 2015. Immigration and ideas: What did Russian scientists "bring" to the United States? J. Labor Econ. 33 (S1 part 2), 257–288.

Gaston, N., Nelson, D., 2002. The employment and wage effects of immigration: trade and labour economics perspectives. In: Greenaway, D., Upward, R., Wakelin, K. (Eds.), Trade, Investment, Migration and Labour Market Adjustment. In: The International Economic Association, Palgrave Macmillan UK, pp. 201–235.

Gaulé, P., Piacentini, M., 2012. Chinese graduate students and U.S. scientific productivity. Rev. Econ. Stat. 95 (2), 698–701.

Griliches, Z., 1990. Patent statistics as economic indicators: A survey. J. Econ. Lit. 28 (4), 1661–1707.

Guellec, D., van Pottelsberghe, B., 2000. Applications, grants and the value of patent. Econom. Lett. 69 (1), 109–114.

Harhoff, D., Scherer, F.M., Vopel, K., 2003. Citations, family size, opposition and the value of patent rights. Res. Policy 32 (8), 1343–1363.

Hausman, J., Hall, B.H., Griliches, Z., 1984. Econometric models for count data with an application to the patents-R&D relationship. Econometrica 909–938.

IOM, 2014. A 'Freer' Flow of Skilled Labour within ASEAN: Aspirations, Opportunities and Challenges in 2015 and Beyond. Issue in Brief No. 11.

Jovanovic, B., 1979. Job Matching and the Theory of Turnover. J. Polit. Econ. 87 (5), 972–990.

Kerr, W., Ozden, C., Rapoport, H., 2018. Editorial: Foreword by the guest editors. J. Econ. Geogr. 18 (4), 691–693.

Li, G.-C., Lai, R., D'Amour, A., Doolin, D.M., Sun, Y., Torvik, V.I., Amy, Z.Y., Fleming, L., 2014. Disambiguation and co-authorship networks of the US patent inventor database (1975–2010). Res. Policy 43 (6), 941–955.

Marino, A., Mudambi, R., Perri, A., Scalera, V.G., 2020. Ties that bind: Ethnic inventors in multinational enterprises' knowledge integration and exploitation. Res. Policy 49 (9), 103956.

Miguélez, E., Moreno, R., Suriñach, J., 2010. Inventors on the move: Tracing inventors' mobility and its spatial distribution. Pap. Reg. Sci. 89 (2), 251–274.

Miguelez, E., Noumedem Temgoua, C., 2020. Inventor migration and knowledge flows: A two-way communication channel? Res. Policy 49 (9), 103914.

Milojević, S., 2013. Accuracy of simple, initials-based methods for author name disambiguation. J. Informetr. 7 (4), 767–773.

Morrison, G., Riccaboni, M., Pammolli, F., 2017. Disambiguation of patent inventors and assignees using high-resolution geolocation data. Sci. Data 4 (1), 1–21.

Moser, P., San, S., 2020. Immigration, science, and invention. lessons from the quota acts. SSRN Scholarly Paper, (ID 3558718), Social Science Research Network, Rochester, NY.

Moser, P., Voena, A., Waldinger, F., 2014. German Jewish Émigrés and US invention. Amer. Econ. Rev. 104 (10), 3222–3255.

Okkerse, L., 2008. How to measure labour market effects of immigration: a review. J. Econ. Surv. 22 (1), 1–30.

de Rassenfosse, G., Dernis, H., Guellec, D., Picci, L., van Pottelsberghe, B., 2013. The worldwide count of priority patents: A new indicator of inventive activity. Res. Policy 42 (3), 720–737.

de Rassenfosse, G., Jaffe, A.B., 2018. Econometric evidence on the depreciation of innovations. Eur. Econ. Rev. 101, 625–642.

de Rassenfosse, G., van Pottelsberghe, B., 2013. The role of fees in patent systems: Theory and evidence. J. Econ. Surv. 27 (4), 696–716.

Schmoch, U., 2008. Concept of a technology classification for country comparisons. In: Final Report to the World Intellectual Property Organisation.

Singh, J., Agrawal, A., 2010. Recruiting for ideas: how firms exploit the prior inventions of New Hires. Manage. Sci. 57 (1), 129–150.

Stephan, P.E., Levin, S.G., 2001. Exceptional contributions to US science by the foreign-Born and foreign-educated. Popul. Res. Policy Rev. 20 (1), 59–79.

Sugimoto, C.R., Ni, C., West, J.D., Larivière, V., 2015. The academic advantage: Gender disparities in patenting. PLoS One 10 (5), e0128000.

Torvik, V.I., Weeber, M., Swanson, D.R., Smalheiser, N.R., 2005. A probabilistic similarity metric for Medline records: A model for author name disambiguation. J. Am. Soc. Inf. Sci. Technol. 56 (2), 140–158.

Useche, D., Miguelez, E., Lissoni, F., 2020. Highly skilled and well connected: Migrant inventors in cross-border m&as. J. Int. Bus. Stud. 51 (5), 737–763.

WIPO, 2020. Patent Cooperation Treaty Yearly Review. World Intellectual Property Organization.

Yoon, C., Doran, K., 2020. Immigration and Invention: Evidence from the Quota Acts. Working Paper.