Productivity gains from migration: Evidence from inventors

Gabriele Pellegrino
Orion Penner
Etienne Piguet
Gaétan de Rassenfosse

January 2023

Innovation and Intellectual Property Policy
Working Paper series no. 20

Available at: https://ideas.repec.org/p/iip/wpaper/20.html

Productivity gains from migration: Evidence from inventors

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

[a]EPFL (Ecole polytechnique fédérale de Lausanne), Lausanne, Switzerland

[b]University of Neuchâtel, Neuchâtel, Switzerland

Abstract

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

JEL codes: F22; J61; O30

Keywords: inventor; productivity; skilled migration

1 Introduction

Attracting high-skilled migration is a high priority on the policy agenda (Bertoli et al., 2012). In many high-income OECD countries, immigrants are over-represented among academics, scientists and entrepreneurs. 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 available 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 world level—a win-win situation for origin and destination countries in the long term. The counter arguments are that departures of intrinsically more productive individuals lead to a detrimental brain-drain for origin countries and may generate crowding out phenomenon of native workers at destination countries.

Despite the policy emphasis placed on high-skilled immigrants, academic research on the topic is thin, as recently emphasized by Kerr and Turner (2015). To be sure, a rich body of work has investigated various economic questions related to immigration in general (see Borjas, 1994; Friedberg and Hunt, 1995; Gaston and Nelson, 2002; Okkerse, 2008; Kerr et al., 2018, for comprehensive surveys of the literature). Previous studies confirm that skilled migration has the potential to contribute substantially to a host country's innovative capacity and productivity growth through a variety of mechanisms. Skilled migrants increase the pool of researchers in a country and generate positive spillovers through the set of skills and knowledge they embody, as well as their personal and business networks (Glaeser, 1999; Alesina and Ferrara, 2005; Breschi et al., 2010; Breschi and Lissoni, 2009; Choudhury, 2016). This knowledge capital ultimately translates into long-term welfare gains (Romer, 1994).

As documented by Docquier and Rapoport (2009), there is considerable heterogeneity among skilled workers that is certainly worth exploring. However, the existing empirical literature largely focuses on tertiary-educated workers, or on self-employed immigrant entrepreneurs (e.g., Piguet, 2010). The present paper focuses on a specific category of (highly) skilled workers that directly contribute to inventive activities, namely inventors. Only a handful of prior studies have focused on inventors, see in particular Moser et al. (2014), Breschi et al. (2014, 2017), Zheng and Ejermo (2015), and Fink and Miguélez (2017) .

This paper studies the productivity patterns of migrants inventors. It builds on a recent body of literature in economic of science and labor economics that indicates that migrant scientists, another 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 and/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 PhD programs in the United States. They show that foreign PhD 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.

We test one mechanism for the productivity differential. We study the extent to which migrant inventors become more productive as a consequence of the move. If most existing studies confirm that migrants are more productive than non-migrants, it remains difficult to pin down the source(s) of this difference. Franzoni et al. (2014) provide some preliminary evidence on the question. Applying 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 also intrinsically more productive than locals before migrating, thereby providing only part of the story. Furthermore, the behavior of scientists, who publish papers in an academic context, may differ from that of inventors, who file patents in a corporate context.

Our analysis tracks over time the productivity patterns of inventors, some of which are migrants, some of which are not. This setup allows us to estimate the intrinsic productivity level of researchers, as well as any potential productivity bump following a migration. We exploit data on inventors listed on international patent applications. These data contain information on the country of nationality and the country of residence of inventors (Fink and Miguélez, 2013). Specifically, we apply a machine learning algorithm in order to disambiguate (that is, uniquely identify) all inventors recorded in the database. We identify migrant inventors through differences in country of residence and country of citizenship. These data are openly available in order to encourage follow-on research.[1]

The results are as follows. First, a preliminary analysis of the worldwide inventor data reveals that migrant inventors make up about ten percent of total inventors. Second, the econometric analysis suggests that migrant inventors become more productive after they have moved. In our preferred setup, migrant inventors enjoy a thirty-percent productivity gain. We observe productivity gains across recipient and sending countries, and 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 (a la Jovanovic, 1979).

The rest of the paper is organized as follows. Section 2 introduces the data and provides an overview of the migration patterns. Section 3 documents the productivity patterns. Section 4 presents the econometric model and the baseline regression results. Section 5 extends the analysis in several ways in order to shed light on the underlying mechanism. Section 6 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

[1]The data will be available from the Harvard Dataverse at https://dataverse.harvard.edu/ upon publication of the paper.

has 153 signatory member states as of December 1st, 2020.[2] 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 applications to be listed also 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.[3] Accordingly, we limit the scope of our analysis to the sample of inventor observed for the period 1990–2011.

Although PCT data offer a unique opportunity to track the international movement of inventors, several 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. To overcome this limitation, we have developed a 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 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.

[2]See Fink and Miguélez (2013) for a detailed description of the main characteristics and the functioning of the PCT system.

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

citizenship and started patenting under her U.S. citizenship for the first time. We will estimate the regressions models on different slices of the data to alleviate these concerns to the extent possible. Third, 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/or 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 full sample of inventors (observed for the period 1990–2011) resulting from the disambiguation procedure broken down by the migration status of inventors.

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) moves for 10,228 migrant inventors (cases 2, 4, 6 and 8). Time period: 1990-2011

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 always resided in their country of nationality over the observed period (case 5). The remaining 80,172 inventors (9.2% of the total sample) can be considered migrants and we classify them in 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.

Figure 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.

Figure 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 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.

We do not claim that the migration figures we report are representative of the population of patenting inventors. 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). 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.

Figure 1: Number of PCT applications and inventors by patent filing year

results_2020/flows-eps-converted-to.pdf

2.3 Migration flows

Disambiguated inventor data enable the tracking of inventors across countries. Figure 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, 47 percent, 34 percent and 25 percent of their resident inventors being a foreign national. The remaining countries show significantly lower shares, ranging from 19 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 45,800 foreign nationals that have filed at least one PCT patent application during the period 1990–2011.

Figure 2: Number and percentage of immigrant inventors by patent filing year

results_2020/flows_imm-eps-converted-to.pdf

Figure 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 Figure 4 is computed using 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, Figure 5 reports the most frequent immigration corridors of the 10,228 migrant inventors for which

Figure 3: Top-15 countries per share of immigrant inventors over total resident inventors

results_2020/top15_dest_cou_per-eps-converted-to.pdf

Note: Absolute numbers showed next to each bin. The same inventor can appear in more that one bin if s/he moves multiple times.

we observe the actual move. China and the Unites States remain the first sending and receiving countries, respectively. However, we can notice some interesting differences with respect to the previous figure. India looses 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 Figure 4 and Figure 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

Figure 4: Most populated corridors for the total sample of 80,172 migrant inventors

results_2020/alluvial_corridors-eps-converted-to.pdf

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.[4] Our data are silent on these issues.

As shown in Table 1, 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 better characterize changes in citizenship, Figure 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

[4]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.

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

results_2020/move_alluvial-eps-converted-to.pdf

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.

3 Productivity patterns

This section presents a descriptive overview of the productivity patterns of inventors. Table 2 (Panel A) suggests that the 80,172 migrant inventors account for 13 percent of total PCT

Figure 6: Most frequent change or acquisition of nationalities

results_2020/ch_nat_alluv-eps-converted-to.pdf

applications in our sample (338,643 out of 2,597,315 inventor-patent pairs, associated with 1,310,573 unique patent applications).

These figures are a first hint that migrant inventors could be more ‘productive’ than non-migrants—indeed, they have filed on average 4.22 PCT applications compared to 2.86 for non-migrants (the difference is statistically significantly different from zero at the 1% probability threshold). Migrant inventors also seem to be more involved in collaborative inventions than non-migrants. They invent less frequently alone (23.6% vs. 25.7%) and more often appear in teams of three or more inventors (50% vs. 48.8%).

Figure 7 depicts the frequency distribution of the number of PCT applications for inventors, by migration status. Compared to migrants, native inventors are more concentrated in the first category: more than 60 percent of non-immigrant inventors have filed just one patent application compared with 50 percent in the case of immigrant inventors. Conversely,

Table 2: 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 x inventor	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 three 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 x inventor	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 three inventors	817,498	690,831	126,667
Panel C: Matched sample			
No. of inventors	232,556	211,218	21,338
No. of applications x inventor	1,575,977	1,418,024	157,953
Average no. of applications x inventor	6.78	6.71	7.40
No. of applications with one inventor	516,547	466,248	50,299
No. of applications with at least three inventors	616,116	554,153	61,963

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.

the proportion of immigrant inventors is always greater than the proportion of native ones in all other categories (2–4; 5–10; 11–20; and more than 21 patents).

Figure 7: Number of PCT applications produced by inventors

`results_2020/NoInv-eps-converted-to.pdf`

This difference in patenting performance could be driven by differences in the technology area of specialization between natives and migrants (*i.e.*, migrant inventors could be attracted to more patent-intensive fields). To investigate potential differences across fields, Figure 8 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.

Figure 8: Technology fields of U.S inventors alongside those whose country of origin is one of the top-4 'sending' to the United States

results_2020/tech_countries-eps-converted-to.pdf

Note: The percentages sum to 100 for each origin country.

Following Schmoch (2008) we identify four main areas of technology: Chemistry, Electrical engineering, Instruments, and Mechanical engineering.[5] British, German and Chinese

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

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.

The analysis presented thus far shows that migrant inventors appear to more 'productive' than non-migrant inventors. However, they also work in larger teams and tend to specialize in fields that are more patent-intensive than natives. The econometric analysis that follows seeks to understand more formally the relationship between productivity and migration.

4 Migration and productivity

4.1 Empirical strategy

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

$$y_{i,t} = \beta_1 \mathrm{AfterMove}_{i,t} + \delta_i + \delta_t + \beta_2 \mathrm{X_{i,t}} + \epsilon_{i,t}, \tag{1}$$

where the dependent variable $y_{i,t}$ identifies the number of patents filed by inventor i in year t.

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), takes into consideration the fact that it is very unlikely that an inventor produces a patentable invention right after her move. It is also consistent with survey evidence presented in de Rassenfosse and Jaffe (2018, Appendix E) on the time between initial expenditure on R&D and the first patent filing.

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.

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. 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.[6]

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 first 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,

[6]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.

we control for the main technology class associated with patent applications by inventor i in year t, by including the set of nine technological categories based on the Cooperative Patent Classification (CPC). Finally, we will also interact the year dummies with the country of residence dummies and CPC dummies to account for unobserved country- and field-specific year effects.

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 2). 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 2 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. It also considerably strengthens our identification strategy, which is based on 'within' variations of productivity. 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.

4.2 Results

Baseline findings

Table 3 presents the results of the baseline OLS estimates. 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 0.25 additional patent application filed each year. The gradual inclusion of the other control variables in columns (2)–(6) reduces the magnitude of this point estimate to about 0.09 in column (6), our preferred specification. Compared with an average of 0.30 patents per inventor per year between 1990–2011, this figure implies a thirty-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 and technology field fixed effects in column (4), technology field × year fixed effects in column (5), and country of residence × technology field fixed effects in column (6).

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 set-up. 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 (7). We estimate the inventor fixed effects using only pre-move information. The predicted fixed effects are then included in an OLS regression model that covers both the pre-move and the post-move period. The coefficient of interest reaches 0.168, which provides a upper-bound estimate for the effect of migration.[7]

Table 3: Ordinary Least Squares regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
After Move	0.252***	0.269***	0.267***	0.089***	0.090***	0.089***	0.168***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)
Inv. fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	No
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	Yes	Yes	Yes	Yes
CPC * Year fix. eff.	No	No	No	No	Yes	Yes	Yes
Countr. of Res. * CPC fix. eff.	No	No	No	No	No	Yes	Yes
Inv. fix. eff. (before move)	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	5,999,400	5,994,582
Number of inventors	272,700	272,700	272,700	272,700	272,700	272,700	272,493
R^2	0.20	0.20	0.20	0.21	0.22	0.22	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. Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

Alternative specifications

The results reported in Table 3 are robust to a broad range of different specifications. First, in order to address the count data nature of our dependent variable (number of patents), we

[7] To run the model in column (7) of Table 3, we exclude all inventors for which we do not have information on the year in which they move.

estimate the main specifications as Poisson regressions with conditional fixed effects. The results, reported in Table B1 in Appendix B, largely confirm the main findings.[8]

Second, 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. Accordingly, we re-run the full specification by considering two distinct time periods: from 1990 to 2003 and from 2004 to 2011. The estimated coefficients of these two models are reported in Table 4. For comparative purposes, we also report the coefficient of Table 3, column (6). The sign and significance of the coefficient of interest are in line with the baseline results.

Third, 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 5, 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 (0.091). In the second case, reported in Table 6, 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 0.112.

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 B2 in Appendix B, confirm the positive impact of migration on productivity (coefficient about 0.05 in column 6).

Taken together, the results confirm the presence of a boost to productivity following a

[8] The coefficients estimated using the conditional fixed effect poisson model cannot be compared with those reported in Table 3. Indeed, convergence issues prevent us from estimating the model including some of the controls.

Table 4: Ordinary Least Squares regressions: Different time periods

	(1990-2011)	(1990-2003)	(2004-2011)
	(1)	(2)	(3)
After Move	0.089***	0.083***	0.150***
	(0.006)	(0.011)	(0.009)
Inventors fixed effects	Yes	Yes	Yes
Av. no of inventors	Yes	Yes	Yes
Time since last patent	Yes	Yes	Yes
Country of Residence fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
CPC fixed effects	Yes	Yes	Yes
CPC * Year fixed effects	Yes	Yes	Yes
Countr. of Res. * CPC fixed eff.	Yes	Yes	Yes
Number of observations	5,999,400	2,181,600	3,817,800
Number of inventors	272,700	272,700	272,700
R^2	0.22	0.20	0.16

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. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

move. Inventors seem to experience a thirty-percent increase in productivity.

5 Additional analyses

5.1 Exploring country-level heterogeneity

As mentioned in Section 2, one of the distinctive features of the PCT database is its world-wide coverage. It provides precise information about both the *nationality* and *residence* for each inventor in more than 120 countries. 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 7 presents the estimation results separately for each country. In all the models, the coefficients are positive and statistically

Table 5: Ordinary Least Squares regressions: Excluding immigrant inventors with double nationality

	(1)	(2)	(3)	(4)	(5)	(6)
After Move	0.252***	0.270***	0.269***	0.094***	0.092***	0.091***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Inventors fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Av. no of inventors	Yes	Yes	Yes	Yes	Yes	Yes
Time since last patent	No	Yes	Yes	Yes	Yes	Yes
Country of Residence fixed effects	No	No	Yes	Yes	Yes	Yes
Year fixed effects	No	No	No	Yes	Yes	Yes
CPC fixed effects	No	No	No	Yes	Yes	Yes
CPC * Year fixed effects	No	No	No	No	Yes	Yes
Countr. of Res. * CPC fixed eff.	No	No	No	No	No	Yes
Number of observations	5,751,152	5,751,152	5,751,152	5,751,152	5,751,152	5,751,152
Number of inventors	261,416	261,416	261,416	261,416	261,416	261,416
R^2	0.20	0.20	0.20	0.22	0.22	0.22

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. Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

significant. They range between 0.055 for Switzerland and 0.196 for China. It would be erroneous to interpret the differences across countries as saying that some countries offer a more productive environment for migrants 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.

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 8, are fairly consistent across countries. The coefficient of interest ranges from 0.026 for British inventors to 0.170 for Indian inventors. Again, it would be erroneous to interpret Table 8 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.

Table 6: Ordinary Least Squares regressions: Excluding immigrant inventors for which we do not observe the actual move

	(1)	(2)	(3)	(4)	(5)	(6)
After Move	0.284***	0.298***	0.290***	0.115***	0.117***	0.112***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Inventors fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Av. no of inventors	Yes	Yes	Yes	Yes	Yes	Yes
Time since last patent	No	Yes	Yes	Yes	Yes	Yes
Country of Residence fixed effects	No	No	Yes	Yes	Yes	Yes
Year fixed effects	No	No	No	Yes	Yes	Yes
CPC fixed effects	No	No	No	Yes	Yes	Yes
CPC * Year fixed effects	No	No	No	No	Yes	Yes
Countr. of Res. * CPC fixed eff.	No	No	No	No	No	Yes
Number of observations	5,478,462	5,478,462	5,478,462	5,478,462	5,478,462	5,478,462
Number of inventors	249,021	249,021	249,021	249,021	249,021	249,021
R^2	0.20	0.20	0.20	0.21	0.22	0.22

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. Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

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 suggest that migrants 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 (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.

The estimated coefficients are reported in Table 9. Column (4) presents the same regression model than that in column (1) of Table 7; except that sample size is now smaller because of data availability on inventor regional location. The coefficient of interest has a similar magnitude (0.138 vs. 0.144) such that we are confident that the difference in sample sizes does not drive our results. Controlling for regional dummies, from column (4) to column (5), leaves the coefficient essentially unchanged.

Table 7: Ordinary Least Squares regressions: Selected receiving countries

	US	DE	GB	CN	CH	Other countries
	(1)	(2)	(3)	(4)	(5)	(6)
After Move	0.138***	0.123***	0.068***	0.196***	0.055***	0.089***
	(0.007)	(0.024)	(0.017)	(0.062)	(0.020)	(0.006)
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
CPC * Year fix. eff.	Yes	Yes	Yes	Yes	Yes	Yes
Countr. of Res. * CPC fix. eff.	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	1,846,427	685,094	328,823	307,752	63,465	5,999,400
Number of inventors	86,956	32,146	15,887	15,156	3,361	272,700
R^2	0.26	0.25	0.37	0.06	0.36	0.22

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. Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

5.2 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 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

Table 8: Ordinary Least Squares regressions: Selected sending countries

	CN	US	DE	GB	IN	Other countries
	(1)	(2)	(3)	(4)	(5)	(6)
After Move	0.026**	0.094***	0.133***	0.060***	0.170***	0.133***
	(0.011)	(0.024)	(0.023)	(0.015)	(0.026)	(0.009)
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
CPC * Year fix. eff.	Yes	Yes	Yes	Yes	Yes	Yes
Countr. of Res. * CPC fix. eff.	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	5,431,553	5,352,105	5,324,769	5,332,236	5,311,453	5,474,488
Number of inventors	248,335	247,737	242,673	243,325	241,884	250,781
R^2	0.22	0.22	0.22	0.22	0.22	0.22

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. Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

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 digits) mode of her patent portfolio.

Interestingly, this rather stringent matching procedure only leads to a mild drop in the total number of inventors as shown in Panel C of Table 2.

Table 10 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.

6 Conclusions

The paper shows that migrant inventors *become* more productive after they have migrated. Overall, we find that migrants enjoy a thirty-percent gain in productivity following a move. This result is robust to a range of specifications and can be observed across destination and sending countries. We have arrived at this conclusion thanks to the careful disambiguation of the approximately 1 million inventors listed in PCT patent documents by way of 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

Table 9: Ordinary Least Squares regressions: Controlling for regional heterogeneity (US only)

	(1)	(2)	(3)	(4)	(5)
After Move	0.294***	0.318***	0.156***	0.144***	0.145***
	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
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
Year fixed effects	No	No	Yes	Yes	Yes
CPC fixed effects	No	No	Yes	Yes	Yes
CPC * Year fix. eff.	No	No	No	Yes	Yes
Regional fixed effects	No	No	No	No	Yes
Number of observations	1,206,018	1,206,018	1,206,018	1,206,018	1,206,018
Number of inventors	54,819	54,819	54,819	54,819	54,819
R^2	0.25	0.25	0.27	0.27	0.27

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. Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

Table 10: Ordinary Least Squares regressions: Matched sample

	(1)	(2)	(3)	(4)	(5)	(6)
After Move	0.254***	0.280***	0.276***	0.091***	0.087***	0.086***
	(0.010)	(0.010)	(0.010)	(0.008)	(0.008)	(0.008)
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	Yes	Yes	Yes
CPC * Year fix. eff.	No	No	No	No	Yes	Yes
Countr. of Res. * CPC fix. eff.	No	No	No	No	No	Yes
Number of observations	5,116,232	5,116,232	5,116,232	5,116,232	5,116,232	5,116,232
Number of inventors	232,556	232,556	232,556	232,556	232,556	232,556
R^2	0.20	0.20	0.20	0.21	0.21	0.21

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. Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

migrant inventors account for more than 10 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 the purpose of education.

We offer three 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 that we observe would tend to 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.[9] Second, migrant inventors may simply move to a more productive environment relative to their home country. This could arise, for example, because migrant inventors work in labs that have better equipment than at home. But we note that the econometric regression models broadly accounts for this explanation with the inclusion of country of residence fixed effects. Leaving a third possible explanation: a migrant's level of productivity is genuinely raised.

We see two channels by which this productivity increase could happen: human capital upgrading and better fit. By moving to a new country, migrants are exposed to new knowledge, new colleagues, new practices, which represent 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 fully exploit their creativity, for example due to ethnic or gender discrimination. 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 the productivity of inventors and, consequently, increases the overall patenting activity worldwide. This leads us back to the tension between the risk of brain drain associated to liberal migration policies and the overall benefits of free migration through better human capital allocation. Do emigration countries loose 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 rather small. This speaks in favor of the free movement of inventors to maximize global innovation outcomes.

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

7 Acknowledgements

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).

CRediT taxonomy: Conceptualization (Gabriele, Gaétan); Data curation (Gabriele, Orion); Formal analysis (Gabriele, Gaétan); Funding acquisition (Gabriele, Orion, Gaétan); Investigation (Gabriele, Orion); Methodology (Gabriele, Orion, Gaétan); Project administration (Gaétan); Resources (Gaétan); Software Programming (Gabriele, Orion); Supervision (Gaétan); Validation (Gabriele, Orion, Gaétan); Visualization (Gabriele); Writing – original draft (Gabriele, Gaétan); Writing – review & editing (all authors).

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 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.

B Additional robustness checks

Table B1: Poisson estimations

	(1)	(2)	(3)	(4)
After Move	0.307***	0.288***	0.196***	1.630***
	(0.008)	(0.007)	(0.008)	(0.023)
Inventor fixed effects	Yes	Yes	Yes	Yes
Country of residence fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes
CPC fixed effects	No	Yes	No	Yes
Number of observations	5,999,400	5,999,400	5,999,400	5,999,400
Number of inventors	272,700	272,700	272,700	272,700

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. Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

Table B2: Ordinary Least Squares regressions: Focus on chemistry

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
After Move	0.148***	0.165***	0.164***	0.052***	0.050***	0.050***	0.112***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
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	Yes	Yes	Yes	Yes
CPC * Year fix. eff.	No	No	No	No	Yes	Yes	Yes
Countr. of Res. * CPC fix. eff.	No	No	No	No	No	Yes	Yes
Inv. fix. eff. (before move)	No	No	No	No	No	No	Yes
Number of observations	1,382,348	1,382,348	1,382,348	1,382,348	1,382,348	1,382,348	1,380,874
Number of inventors	62,834	62,834	62,834	62,834	62,834	62,834	62,767
R^2	0.27	0.27	0.27	0.28	0.28	0.28	0.27

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. Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

References

ALESINA, A. AND E. L. FERRARA (2005): "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, 43, 762–800.

BALSMEIER, B., A. CHAVOSH, G.-C. LI, G. FIERRO, K. JOHNSON, A. KAULAGI, D. O'REAGAN, B. YEH, AND L. FLEMING (2015): "Automated disambiguation of us patent grants and applications," *Unpublished working paper, Fung Institute for Engineering Leadership.*

BERTOLI, S., H. BRÜCKER, G. FACCHINI, A. M. MAYDA, AND G. PERI (2012): "Understanding highly skilled migration in developed countries: The upcoming battle for brains," *Brain Drain and Brain Gain. The Global Competition to Attract High-Skilled Migrants, in T. Boeri, H. Bruecker, F. Docquier and H. Rapoport, Oxford University Press.*

BORJAS, G. J. (1994): "The Economics of Immigration," *Journal of Economic Literature*, 32, 1667–1717.

BORJAS, G. J. AND K. B. DORAN (2012): "The Collapse of the Soviet Union and the Productivity of American Mathematicians," *The Quarterly Journal of Economics*, 1143–1203.

BRESCHI, S., C. LENZI, F. LISSONI, AND A. VEZZULLI (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*, 353–369.

BRESCHI, S. AND F. LISSONI (2009): "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows," *Journal of Economic Geography*, 9, 439–468.

BRESCHI, S., F. LISSONI, AND E. MIGUELEZ (2017): "Foreign-Origin Inventors in the USA: Testing for Diaspora and Brain Gain Effects," *Journal of Economic Geography*, 17, 1009–1038.

Breschi, S., F. Lissoni, and G. Tarasconi (2014): *Inventor Data for Research on Migration and Innovation: A Survey and a Pilot*, WIPO Economic and Statistics Series, Economic Research Working Papers.

Choudhury, P. (2016): "Return migration and geography of innovation in MNEs: a natural experiment of knowledge production by local workers reporting to return migrants," *Journal of Economic Geography*, 16, 585–610.

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

Danguy, J., G. de Rassenfosse, and B. van Pottelsberghe de la Potterie (2013): "On the origins of the worldwide surge in patenting: an industry perspective on the R&D–patent relationship," *Industrial and corporate change*, 23, 535–572.

de Rassenfosse, G. and A. B. Jaffe (2018): "Econometric evidence on the depreciation of innovations," *European Economic Review*, 101, 625–642.

Docquier, F. and H. Rapoport (2009): "Documenting the Brain Drain of "La Crème de La Crème"," *Journal of Economics and Statistics*, 229, 679–705, 00000.

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

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

——— (2017): *The International Mobility of Talent and Innovation*, Cambridge University Press.

Franzoni, C., G. Scellato, and P. Stephan (2014): "The Mover's Advantage: The Superior Performance of Migrant Scientists," *Economics Letters*, 122, 89–93.

Friedberg, R. M. and J. Hunt (1995): "The Impact of Immigrants on Host Country Wages, Employment and Growth," *The Journal of Economic Perspectives*, 9, 23–44.

Gaston, N. and D. Nelson (2002): "The Employment and Wage Effects of Immigration: Trade and Labour Economics Perspectives," in *Trade, Investment, Migration and Labour Market Adjustment*, ed. by D. Greenaway, R. Upward, and K. Wakelin, Palgrave Macmillan UK, The International Economic Association, 201–235.

Gaulé, P. and M. Piacentini (2012): "Chinese Graduate Students and U.S. Scientific Productivity," *The Review of Economics and Statistics*, 95, 698–701.

Glaeser, E. L. (1999): "Learning in Cities," *Journal of urban Economics*, 46, 254–277.

Griliches, Z. (1990): "Patent statistics as economic indicators: A survey." *Journal of Economic Literature*, 28, 1661–1707.

Guellec, D. and B. van Pottelsberghe (2000): "Applications, grants and the value of patent," *Economics Letters*, 69, 109–114.

Harhoff, D., F. M. Scherer, and K. Vopel (2003): "Citations, family size, opposition and the value of patent rights," *Research policy*, 32, 1343–1363.

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," *Journal of Political Economy*, 87, 972–990.

Kerr, W., C. Ozden, and H. Rapoport (2018): "Editorial: Foreword by the guest editors," *Journal of Economic Geography*, 18, 691–693.

Kerr, W. and S. Turner (2015): "Introduction: US High-Skilled Immigration in the Global Economy," *Journal of Labor Economics*, 33, S1 – S4, 00003.

Milojević, S. (2013): "Accuracy of simple, initials-based methods for author name disambiguation," *Journal of Informetrics*, 7, 767–773.

Moser, P., A. Voena, and F. Waldinger (2014): "German Jewish Émigrés and US Invention," *American Economic Review*, 104, 3222–55.

Okkerse, L. (2008): “How to Measure Labour Market Effects of Immigration: a Review,” *Journal of Economic Surveys*, 22, 1–30.

Piguet, É. (2010): “Entrepreneurship among Immigrants in Switzerland,” *OECD*, 149–179.

Romer, P. M. (1994): “The Origins of Endogenous Growth,” *The Journal of Economic Perspectives*, 8, 3–22.

Schmoch, U. (2008): “Concept of a Technology Classification for Country Comparisons,” *Final Report to the World Intellectual Property Organisation.*

Singh, J. and A. Agrawal (2010): “Recruiting for Ideas: How Firms Exploit the Prior Inventions of New Hires,” *Management Science*, 57, 129–150.

Stephan, P. E. and S. G. Levin (2001): “Exceptional Contributions to US Science by the Foreign-Born and Foreign-Educated,” *Population Research and Policy Review*, 20, 59–79.

Zheng, Y. and O. Ejermo (2015): “How do the foreign-born perform in inventive activity? Evidence from Sweden,” *Journal of Population Economics*, 28, 659–695.