Immigration and Inventor Productivity

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

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

[b]University of Neuchâtel, Neuchâtel, Switzerland

[c]WIPO(World Intellectual Property Organization), Genève, Switzerland

February 13, 2019

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 uniquely identify inventors across patents, we are able to track, for the first time, the worldwide migration patterns of nearly one million individual inventors. The econometric analysis seeks to explain the recurring finding in the literature that migrant inventors are more productive than non-migrant inventors. We investigate whether the effect of migration is the selection of inventors with high ability or whether migration per se boosts productivity. We find that migrant inventors were indeed intrinsically more productive than non-migrant inventors before the move (selection effect) but they also became all the more productive after migration (boost effect). We estimate that about 85% of the productivity difference is explained by post-move productivity increase and the remaining 15% is explained by selection. We explore the policy implications of this finding. The disambiguated inventor data are openly available to encourage follow-on research.

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. 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 ethical rationale behind these policies is that migration allows a better allocation of human capital, boosts the overall productivity at world level and hence lead to a win-win situation for origin and destination countries on the long term. The counter argument is that departures of intrinsically more productive individuals lead to a detrimental brain-drain for origin countries.

Despite the policy emphasis placed on high-skilled immigrants, academic research on the topic is thin. As recently emphasized in the extensive literature review of Kerr and Turner (2015, p. 1), "the limited evidence regarding the economic and distributional consequences of high-skilled immigration has become all the more apparent and worrisome given the policy focus on the issue." 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 (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 Docquier and Rapoport (2009), Moser et al. (2014), Breschi et al. (2017) and Fink and Miguélez (2017).

This paper studies the productivity patterns of migrants inventors. It builds on a recent body of literature in economic geography 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. Are migrant workers the most productive workers, and that is the reason why they were able to move (selection effect)? Or do they become more productive as a consequence of the move (boost effect)? If most existing studies confirm that migrants are more productive than non-migrants, it remains difficult to disentangle the selection and the boost effects. Franzoni et al. (2014) use cross-sectional survey data on scientists active in different fields and residing in 16 different countries. Applying an instrumental variable approach, 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 before migrating.

To study the origin of the productivity differential, we track 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. Specifically, we apply a machine learning algorithm in order to uniquely identify all inventors recorded in the database and identify migrant inventors thank to differences in country of residence vs. country of citizenship. These data will be openly available in order to encourage follow-on research.[1]

Results of inventor-level panel regression models suggest that migrant inventors are significantly more productive than natives. In our preferred specification, about 85% of this difference is explained by post-move productivity increase and the remaining 15% is explained by the fact that migrant inventors are intrinsically more productive than non migrants. An initial analysis of the disambiguated data further reveals that migrant inventors account for more than 10% of the population of inventors since the early 2000s. However, the majority of migrant inventors were already in the host country prior to their first patent application—suggesting that they may have moved for the purpose of education.

The rest of the paper is organized as follows. Section 2 provides a brief overview of the migration patterns, while Section 3 focuses on productivity patterns. Section 4 presents the

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

econometric model and the regression results. Section 5 concludes.

2 Migration Patterns

2.1 PCT data

We observe inventors listed in patent applications filed under the Patent Cooperation Treaty (PCT) route. The PCT is an international treaty that facilitates international patenting. It is administrated by the Geneva-based World Intellectual Property Organization (WIPO) and has 158 signatory member states as of July 1st, 2018.[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 1978–2012.

A limitation of the PCT data relates to the fact that inventors 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% and recall of about 80%.

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 slightly below one million (962,351

[2]See Fink and Miguélez (2013) for a detailed description of the main characteristics and the functioni ng 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.

observations). 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 1978–2012) 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)	65,422	5)Non migrant	871,003
Residence varies over time, while nationality is constant	2) Migrant (move recorded)	698	6) Migrant (move recorded)	8,316
Nationality varies over time, while residence is constant	3) Double nationality migrant (move not recorded)	370	7) Double nationality migrant (move not recorded)	11,460
Both residence and nationality vary over time	4) Double nationality migrant (move recorded)	207	8) Double nationality migrant (move recorded)	4,875

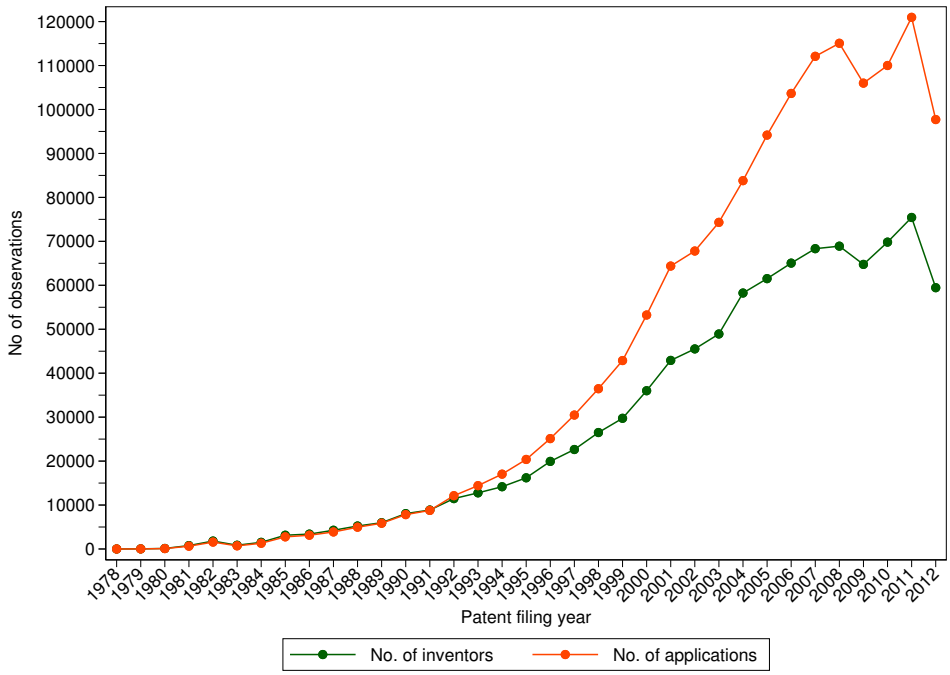
Notes: The total number of immigrant inventors is 91,348, of which 16,912 have double nationality. The dataset records one (or more than one) moves for 13,899 migrant inventors (cases 2, 4, 6 and 8). Time period: 1978-2012

Not surprisingly, the vast majority of inventors in the sample are not migrants. Indeed, 871,003 inventors, representing more than 90.5% of the total sample, have always resided in their country of nationality over the observed period (case 5). The remaining 91,348 inventors (9.5% 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 18% of the total migrant inventors (cases 3, 4, 7 and 8; 16,912 observations). Second, migrant inventors with single nationality, amounting to 74,438 observations (cases 1, 2 and 6). As shown in Table 1, the dataset records one (or more than one) move for 13,889 migrant inventors (cases 2, 4, 6 and 8; around 15% of the migrant group). These correspond to cases where we observe the same inventors in two patent applications with different residence/nationality status. We do not observe the change in the country of residence (*i.e.*, international move) for the majority of migrant inventors (cases 1 , 3 and 7, amounting to 85% 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 1978–2012. Through the end of the 1980s, both the number of inventors and applications is rather small. Starting in the early 1990s there is a notable increase in both the number of inventors and the number of applications. The growth reflects an uptake of the PCT system, rather than a burst in inventiveness (Danguy et al., 2013). Growth halts in 2008, likely as a result of the Global Financial Crisis. The sudden drop in 2011 is due to the legislative change in the United States noted above, which reduced the availability of information on nationality and residence.

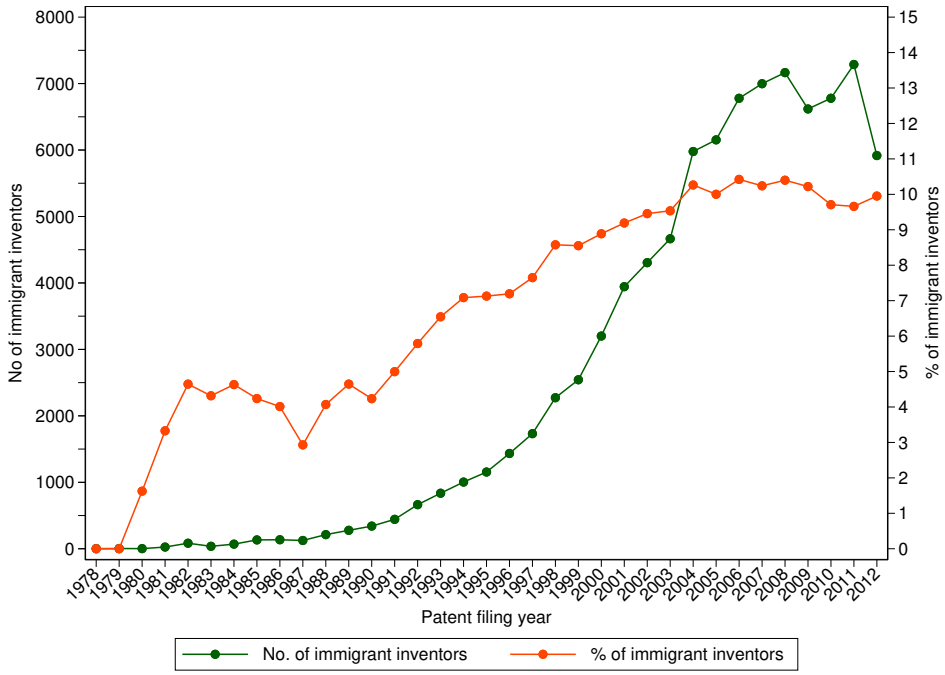
Figure 2 focuses on immigrant inventors. It depicts the number (left axis) and percentage (over the total sample of inventors, right axis) of immigrant inventors by patent filing year.

Figure 1: Number of PCT applications and 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 10% of all inventors listed in PCT applications since the mid-2000s. It seems to have reached a plateau since then.

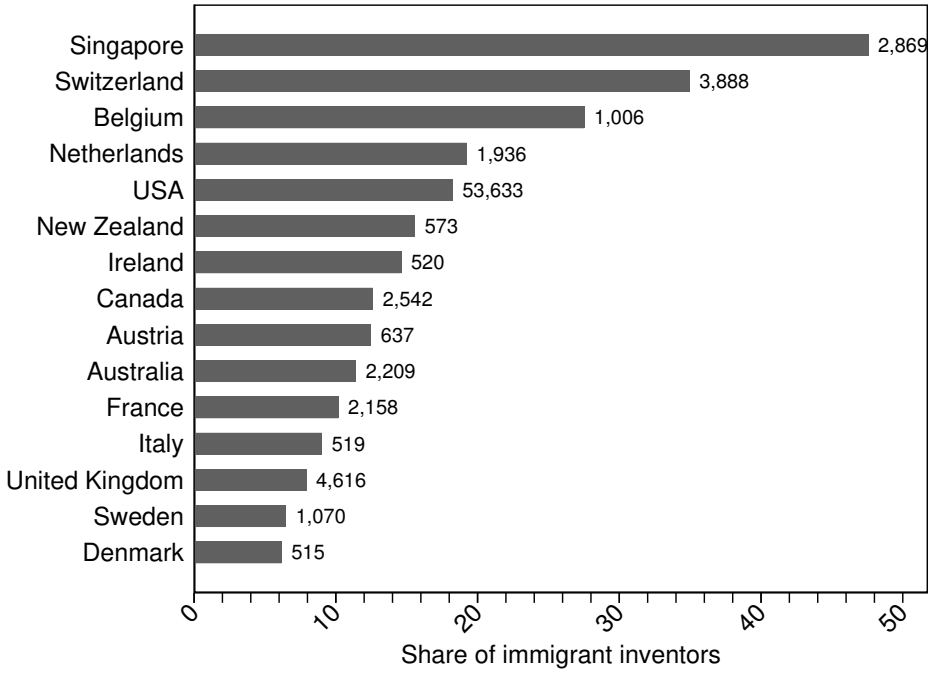
Figure 2: Number and percentage of immigrant inventors by patent filing year



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 share 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%, 35% and 27% of their resident inventors being a foreign national. The remaining countries show significantly lower shares, ranging from 19% for the Netherlands to 6% for Denmark. 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 53,000 foreign nationals that have filed at least one PCT patent application during the period 1978–2012.

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

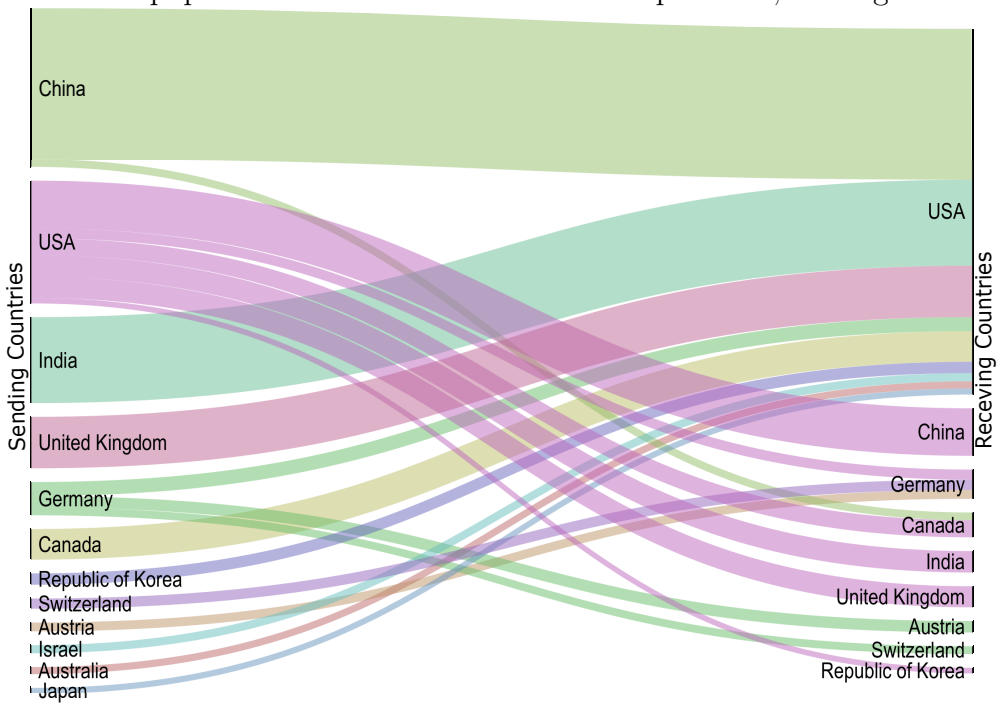


Note: Absolute numbers showed next to each bin. The same inventor can appear in more that one bin if s/he moves multiple times.

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 Switzerland, United Kingdom and Singapore. On the other hand, outward migration is more fragmented. China and India represent the two most important sending countries, followed by the United Kingdom, Germany and Canada. Note that Figure 4 is computed using the total sample of 91,348

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.

Figure 4: Most populated corridors for the total sample of 91,348 migrant inventors

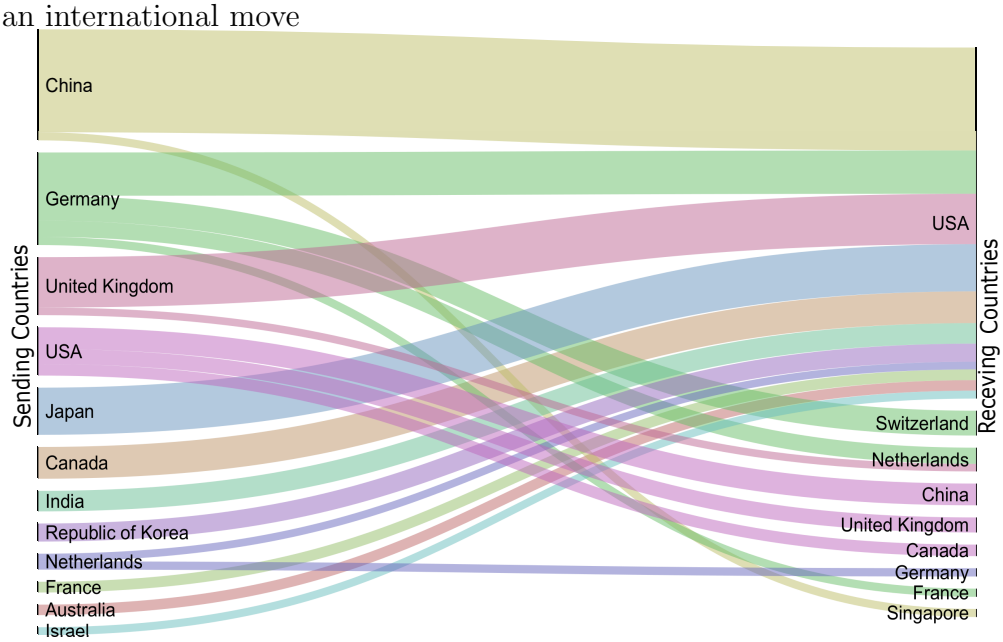


To provide some evidence on the mobility events included in our dataset, Figure 5 reports the most frequent immigration corridors of the 13,768 migrant inventors for which we observe the actual move. China and the Unites States remain the first sending and receiving countries, respectively. However, 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 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. Our data are silent on these issues.

As shown in Table 1, a sizable proportion of migrant inventors in our sample (around 18%) 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.

Figure 5: Most populated corridors for the sample of 13,768 migrant inventors for which we observe an international move



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. 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 who were born and educated in the United States.

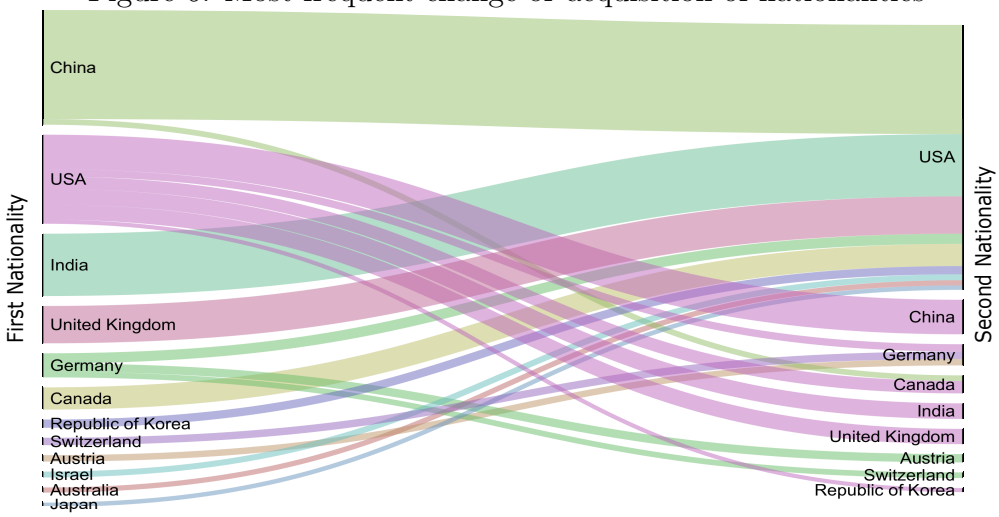
The data also show an interesting pattern of 'nationality exchange' among Germany and Austria on the one hand and Germany and Switzerland on the other hand. These countries share common border and language, and are characterized by a high degree of cultural similarities—and therefore also by intense bilateral migration flows.

3 Productivity patterns

This section presents a descriptive overview of the productivity patterns of inventors. It focuses on the sub-sample of inventors observed for the period 1990–2011. As is clear from Figure 1, the number of inventors before the year 1990 is rather small and following 2011 there is a sudden drop in observations.

There is in total 867,627 disambiguated inventors, who have filed a combined 1,310,850 unique PCT applications in the period 1990–2011, as indicated in Panel A of Table 2. About 9.6% of inventors are immigrants, and they account for 14.8% of total PCT appli-

Figure 6: Most frequent change or acquisition of nationalities



cations. These figures are a first hint that migrant inventors could be more "productive" than non-migrants—indeed, they have filed on average 2.34 PCT applications compared to 1.42 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: they invent less frequently alone (47.2% vs. 51%) and more often appear in teams of three or more inventors (25.5% vs. 24%).

Table 2: Summary statistics of sample used for the econometric regression

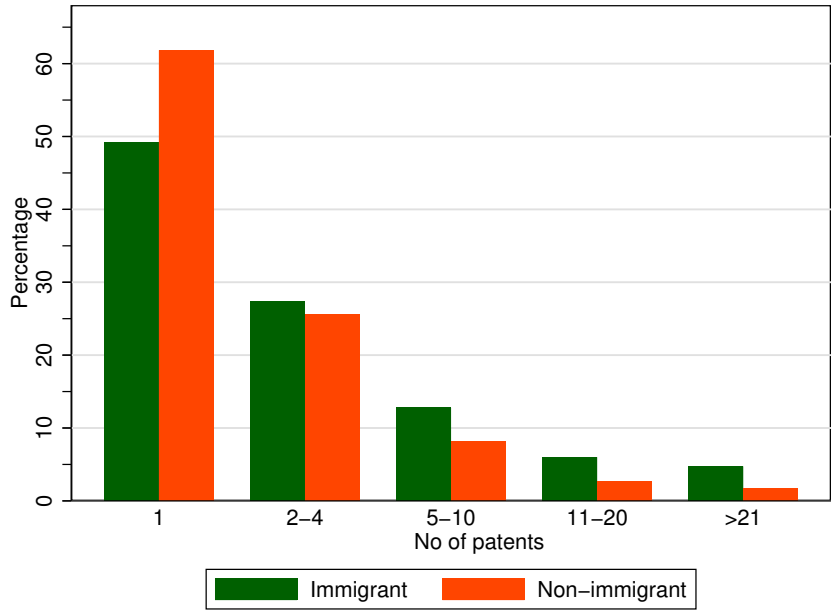
	Total	***Non Immigrant***	***Immigrant***
Panel A: Whole Sample			
No. of inventors	867,627	784,233	83,394
No. of applications	1,310,850	1,115,689	195,161
Average no. of applications per inventor	1.51	1.42	2.34
No. of applications with one inventor	661,271	569,011	92,260
No. of applications with at least three inventors	318,056	268,345	49,711
Panel B: Sample Used for the Estimations			
No. of Inventors	262,265	226,869	35,396
No. of applications	642,612	542,979	99,633
Average No. of applications per inventor	2.45	2.39	2.81
No. of applications with one inventor	457,743	384,924	72,819
No. of applications with at least three inventors	56,058	48,420	7,638

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.

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% of non-immigrant inventors have filed just one patent application compared with 50% in the case of immigrant inventors. Conversely, the proportion of immigrant inventors is always greater than the proportion of native ones in all following categories (2–4; 5–10; 11–20; and more than 21 patents).

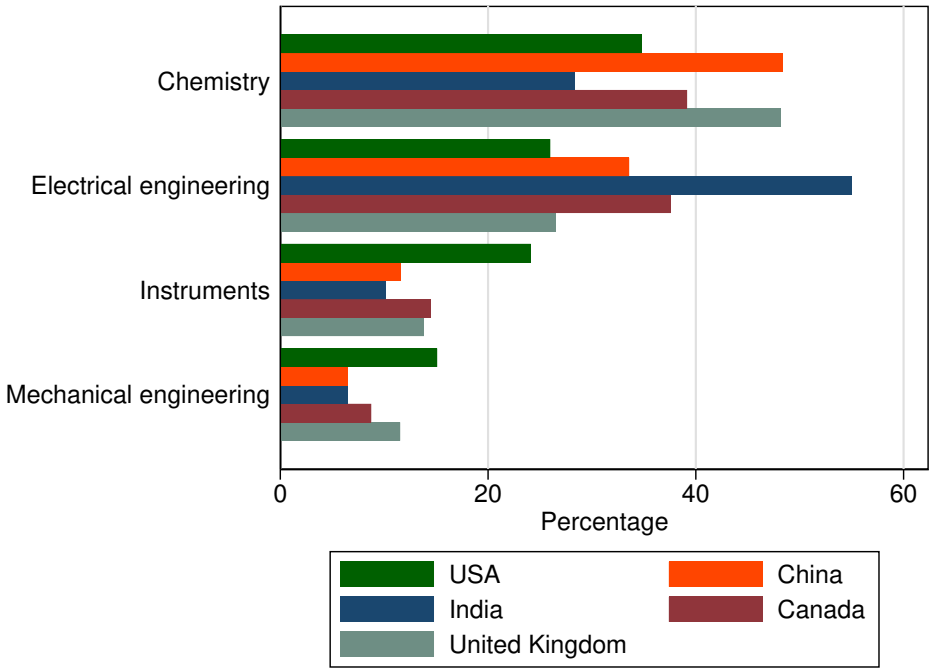
This difference in patenting performance could be driven by differences in the technology area of specialization between natives and migrants (*i.e.*, migrant inventors could be

Figure 7: Number of PCT applications produced by inventors



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, Canada 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 U.S.



Following Schmoch (2008) we identify four main areas of technology: Chemistry, Electrical engineering, Instruments, and Mechanical engineering.[4] British, Canadian and Chinese inventors migrating to the United States are more frequently found in chemistry compared to the baseline rate of about 35% for U.S. inventors. Along the same lines, Chinese, Canadian and in particular Indian inventors migrating to the United States are more frequently found in electrical engineering, compared to the baseline of about 25% for U.S. inventors. Electrical engineering is a field where inventors happen to file many patent applications. In our sample, each inventor filed 17.2 patents on average over the study period. On the other hand, U.S. inventors are relatively more numerous than immigrant inventors in the fields of instruments and mechanical engineering. The average number of patents per inventor over the study period was significantly lower in these fields, reaching 11.83 and 7.13, respectively.

The analysis presented thus far shows that migrant inventors appear to more "productive" than non-migrant inventors. However, they are also involved in larger inventive teams and tend to specialize in fields that are more patent-intensive than natives. The econometric analysis that follows seeks to understand better the relationship between productivity and migration. It exploits mobility events to understand whether inventors are intrinsically more productive than natives and/or whether they became more productive as a consequence of the move.

4 Migration and Productivity

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 one starting from the year t in which we observe the first move of inventor i. The sign and significance of the coefficient β_1 will therefore provide 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 captures inventor fixed effects. It accounts for unobservable, time invariant individual characteristics that may cause variation in patenting activity across different inventors. The mean of the individual fixed effects computed for the migrant and non-migrant inventor groups will inform us on whether there are systematic productivity differentials between these two groups. Next, δ_t includes a vector of year fixed effects that

[4] We exclude some residual technology areas accounting for less than 4% of the sample.

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, in order to control for country specific factors that may affect the inventor's patenting rate. We want to prevent our findings from being influenced by the fact that, say, an Indian inventor who moves to the United States is seen as more productive after his move simply because his company in India did not patent a lot. 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 elapsed between each patent application filed by inventor i. Fourth, 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 application). Finally, 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).

The sample is composed of all 867,627 inventors, who have filed a total of 1,310,850 PCT applications over the study period (see Table 2). However, because we estimate equation (1) with fixed effects, a total of 595,356 inventors that we observe for just one year are dropped from the sample. Moreover, we further exclude from the sample 10,006 inventors for which we do not observe the complete set of variables used for the estimations. Panel B of Table 2 reports basic descriptive statistics for this sample. 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.

Table 3 presents the results of the baseline OLS estimates, with robust standard errors. Column (1) shows a positive and highly 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 0.7 additional patent application filed each year. The inclusion of the other control variables reduces the magnitude of this point estimate to about 0.5. We add technology fields in column (2), year fixed effects in column (3), the variable "time since last patent" in column (4), and the average number of inventors in column (5).

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'. The results in column (6) account for this possibility. 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. This approach leaves the coefficient of interest essentially unchanged.[5]

Table 3: Ordinary Least Square regressions

	(1)	(2)	(3)	(4)	(5)	(6)
After Move	0.709***	0.709***	0.486***	0.483***	0.480***	0.514***
	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.041)
Inventors fixed effects	Yes	Yes	Yes	Yes	Yes	No
Country of Residence fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
CPC fixed effects	No	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes	Yes
Time since last patent	No	No	No	Yes	Yes	Yes
Av. no of inventors	No	No	No	No	Yes	Yes
Inventors fixed effects (before move)	No	No	No	No	No	Yes
Number of observations	915,385	915,385	915,385	915,385	915,385	827,326
Number of inventors	262,265	262,265	262,265	262,265	262,265	237,930
R^2	0.001	0.001	0.005	0.005	0.005	0.005

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 in which we observe the first international move of inventor i. Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

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 estimate the main specifications as Poisson regressions with conditional fixed effects. Results, reported in Table 4 largely confirm the main findings. Although much smaller in magnitude, the coefficient of the variable $AfterMove$ still shows a positive and highly significantly impact on the dependent variable of about 0.2 (marginal effect at mean reported).

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 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. In the second case, reported in Table 6, we use an even more stringent definition of immigrant inventors. Specifically, we exclude from the sample immigrant inventors for which we cannot observe any actual moves (*i.e.*, change in country of residence, cases 1, 3 and 7 in Table 1). Once again, the results confirm the main findings.

Finally, we have also purged the point estimate of the coefficient associated with the $AfterMove$ variable from the effect of the passing of time. We are concerned that the post-move productivity increase may result from a transitory effect that fades over time. We

[5]To run this model, we exclude all those inventors for which we do not have information on the year in which they move.

Table 4: Poisson regressions

	(1)	(2)	(3)	(4)	(5)
After Move	0.307***	0.288***	0.196***	0.198***	0.197***
	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)
Inventors fixed effects	Yes	Yes	Yes	Yes	Yes
Country of Residence fixed effects	Yes	Yes	Yes	Yes	Yes
CPC fixed effects	No	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes
Time since last patent	No	No	No	Yes	Yes
Av. no of inventors	No	No	No	No	Yes
Number of observations	915,385	915,385	915,385	915,385	915,385
Number of inventors	262,265	262,265	262,265	262,265	262,265

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 in which we observe the first move of inventor i. Coefficients are marginal effects calculated at the mean. Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

have re-estimated the model in column (5) of Table 6 by controlling for the number of years elapsed since the first move. The point estimates barely drops, reaching 0.395 (compared to 0.432).

It is possible to use the estimated inventor fixed effects to study whether migrant inventors are intrinsically more productive than non-migrant inventors. We have computed the mean of the fixed effects for the group of non-migrant U.S. inventors and inventors who migrated to the United States. By construction, these estimates are purged from the influence of all variables in the model. We use the model in column (5) of Table 6 to make sure that we observe pre-move patenting activity for all migrant inventors. We find a positive and statistically significant difference in the group means of the fixed effects between migrants and natives of about 0.09 patent per year (p-value of 0.0001). This finding suggests that migrant inventors are actually intrinsically more productive than natives. Overall, migrant inventors produce about half a PCT patent more per year than natives (0.432+0.09=0.522). About 85% of this difference is explained by post-move productivity increase (0.432/0.512) and the remaining 15% is explained by the fact that migrant inventors exhibit intrinsically higher productivity levels.

5 Conclusions

The paper shows that migrant inventors *become* more productive after they have migrated. Furthermore, migrants inventors are also intrinsically more productive than non-migrant inventors. We estimate that about 85% of this difference is explained by post-move productivity increase and the remaining 15% is explained by the fact that migrant inventors exhibit intrinsically higher productivity levels. We have arrived at this conclusion thanks to the careful disambiguation of the approximately 1 million inventors listed in PCT patent

Table 5: Ordinary Least Square Regressions: excluding immigrant inventors with double nationality

	(1)	(2)	(3)	(4)	(5)
After Move	0.560***	0.560***	0.331***	0.328***	0.327***
	(0.060)	(0.060)	(0.060)	(0.060)	(0.060)
Inventors fixed effects	Yes	Yes	Yes	Yes	Yes
Country of Residence fixed effects	Yes	Yes	Yes	Yes	Yes
CPC fixed effects	No	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes
Time since last patent	No	No	No	Yes	Yes
Av. no of inventors	No	No	No	No	Yes
Number of observations	845,476	845,476	845,476	845,476	845,476
Number of inventors	247,315	247,315	247,315	247,315	247,315
R^2	0.001	0.001	0.004	0.004	0.004

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 in which we observe the first move of inventor i. Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

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 further reveals that migrant inventors account for more than 10% 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 answers as to *why* migrant inventors become more productive. 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. Second, the skills and potential of migrant inventors may be better exploited by their host country 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 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. By moving to a new country (often more technologically-advanced), 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 implement their creativity, for example due to ethnic or gender discrimination. Our data are silent on the

Table 6: Ordinary Least Square Regressions: excluding immigrant inventors for which we do not observed the actual move

	(1)	(2)	(3)	(4)	(5)
After Move	0.674***	0.674***	0.437***	0.434***	0.432***
	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)
Inventors fixed effects	Yes	Yes	Yes	Yes	Yes
Country of Residence fixed effects	Yes	Yes	Yes	Yes	Yes
CPC fixed effects	No	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes
Time since last patent	No	No	No	Yes	Yes
Av. no of inventors	No	No	No	No	Yes
Number of observations	834,144	834,144	834,144	834,144	834,144
Number of inventors	239,614	239,614	239,614	239,614	239,614
R^2	0.001	0.001	0.005	0.005	0.005

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 in which we observe the first move of inventor i. Robust standard errors in parentheses.
***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

individual motives for migration, and we leave it to further research to tease out these possible remaining mechanisms.

Regarding policy implications, the central finding is that migration boosts the productivity of inventors and thus 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, migrant inventors were only slightly more productive than non-migrants before migration and if, as we have also demonstrated, migration by itself is the booster of productivity, then the brain drain effect appears rather small. This speaks in favor of letting inventors free to choose the country in which they want to work and patent their discoveries.

6 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); 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). The disambiguation approach developed and applied herein is fully spelled out in a forthcoming manuscript (Penner et al., 2018). We outline the basic principles below.

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.

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.

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.

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.

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.

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.

Penner, O., G. Pellegrino, and G. de Rassenfosse (2018): "Machine learning based disambiguation of patent inventors," *mimeo*.

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*.

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.