I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12382

Migrant Inventors and the Technological Advantage of Nations

Dany Bahar
Prithwiraj Choudhury
Hillel Rapoport

MAY 2019

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12382

Migrant Inventors and the Technological Advantage of Nations

Dany Bahar
The Brookings Institution, Harvard CID, CESifo and IZA

Prithwiraj Choudhury
Harvard Business School

Hillel Rapoport
Paris School of Economics University Paris 1, CEPII, CESifo and IZA

MAY 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Migrant Inventors and the Technological Advantage of Nations*

We investigate the relationship between the presence of migrant inventors and the dynamics of innovation in the migrants' receiving countries. We find that countries are 25 to 50 percent more likely to gain advantage in patenting in certain technologies given a twofold increase in the number of foreign inventors from other nations that specialize in those same technologies. For the average country in our sample this number corresponds to only 25 inventors and a standard deviation of 135. We deal with endogeneity concerns by using historical migration networks to instrument for stocks of migrant inventors. Our results generalize the evidence of previous studies that show how migrant inventors "import" knowledge from their home countries which translate into higher patenting. We complement our results with micro-evidence showing that migrant inventors are more prevalent in the first bulk of patents of a country in a given technology, as compared to patents filed at later stages. We interpret these results as tangible evidence of migrants facilitating the technology-specific diffusion of knowledge across nations.

JEL Classification: O31, O33, F22
Keywords: innovation, migration, patent, technology, knowledge

Corresponding author:
Dany Bahar
The Brookings Institution
1775 Massachusetts Ave NW.
Washington DC, 20001
USA
E-mail: dbahar@brookings.edu

* The authors are thankful to Ina Ganguli, Francesco Lissoni, Ernest Miguelez, James Sappenfield and to four anonymous referees for comments and suggestions. We thank Luis Da Silva and Aaron Acero for excellent research assistance. All errors are our own.

"Through the ages, the main channel for the diffusion of innovations has been the migration of people."

(Cipolla, 1976, p. 121)

1 Introduction

It is a known fact that German and Austrian Jewish scientists and inventors who fled from Nazi Germany during the mid 1930s played a crucial role in boosting the innovation capabilities of the countries that received them, and in particular, of the United States. Moreover, this boost in innovation–which was a result of higher combined patenting activity for both immigrants and natives–was in research fields where the German scientists were active inventors in their home countries prior to the war, such as Chemistry (Moser et al., 2014). While there is plenty and growing evidence of the impact of migration on innovation (e.g., Kerr, 2008; Agrawal et al., 2008; Hunt and Gauthier-Loiselle, 2010; Kerr and Lincoln, 2010; Freeman and Huang, 2015; Ganguli, 2015; Bosetti et al., 2015; Choudhury, 2016; Akcigit et al., 2017; Breschi et al., 2017; Bernstein et al., 2018; Miguélez, 2018; Choudhury and Kim, 2018; Doran and Yoon, 2019)[1], there is less systematic evidence –based on a larger number of countries examined over a period of time– documenting the role migrant inventors and scientists play in their receiving countries, in boosting innovation of the same technologies for which their home countries have rich patenting activity in. This paper comes to fill this gap in the literature.

Why focus on technology-specific knowledge diffusion across geographic borders, from the home to the host country of the migrant? Here, we build on Bahar and Rapoport (2018) who document that migrants explain part of the process through which countries gain comparative advantage in certain export goods, and interpret this as migrants inducing industry-specific productivity shifts. Yet, the underlying mechanisms through which those productivity shifts occur are, for the most part, still unexplored. In this

[1] See Lissoni (2018) for a comprehensive survey of this literature.

paper we focus on one of those possible mechanisms: the diffusion of technologies from the home country to the host country via migrant inventors and scientists.

In particular, we ask: do migrants boost patent production in their countries of destination (alt. origin) in the same technology classes that their home (alt. receiving) countries specialize in? We find that, for any given country c and technology p, a twofold increase in the number of migrant inventors from other nations that specialize in patenting in technology p is associated with a 25% to 50% increase in the probability that c gains global technological advantage in p within a decade. In our exercise, gaining technological advantage implies achieving a number of patent applications in technology p that is proportionally larger than the global average. This twofold increase, for the average country in our sample, corresponds to only about 25 inventors (and a standard deviation of 135).

Our paper attempts to integrate two previously disconnected, yet important strands of the innovation and patenting literature: the literature on comparative patenting across countries and the literature documenting the role of migrant inventors in facilitating knowledge production across borders.

Regarding the first strand, there is a rich tradition of using patents to study comparative invention practices across countries. Glismann and Horn (1988) present an analysis of invention performance, as measured by patenting activities of six countries (France, Italy, Japan, United Kingdom, USSR, West Germany) relative to the United States for 41 SIC industries over 1963-1983, suggesting the existence of "catching-up" processes in terms of patenting activity. In a more recent paper, De Noni et al. (2018) assert that less innovative European regions (referred to as 'lagging-behind regions' in their paper) must actively work to reduce the gap between them and knowledge-intensive regions. The authors employ a seven-year panel dataset (2002–2008) using patent data at a regional level to validate the hypothesis that collaborations, and specifically with highly innovative regions, positively affect the innovation performances of lagging-behind regions.

This literature, arguably, has not been explicitly integrated with the literature on migrant inventors and knowledge production across borders.

The importance of geographic and political borders for knowledge transfer and knowledge production has been long studied in the patenting and innovation literature. Building on the rich literature of geographic localization of knowledge spillovers (Thompson and Fox-Kean, 2005; Henderson et al., 2005), Singh and Marx (2013) find a strong role of political borders in knowledge diffusion: the authors find both country and state borders to have independent effects on knowledge diffusion beyond what just geographic proximity in the form of metropolitan collocation or shorter within-region distances can explain. In this literature, Foley and Kerr (2013) find that that increases in the share of a firm's innovation performed by inventors of a particular ethnicity are associated with increases in the share of that firm's affiliate activity in countries related to that ethnicity. The authors also report that ethnic innovators appear to facilitate the disintegration of innovative activity across borders and to allow U.S. multinationals to form new affiliates abroad without the support of local joint venture partners. Almeida et al. (2014) study patent data, and find that the utility of ethnic knowledge and collaborators depends on the level of inventor embeddedness in the community; most inventors benefit by sourcing knowledge from, or collaborating with, other ethnic scientists and hence enhance innovation quality . In a recent paper, Kerr and Kerr (2018) connect collaborative patents to the ethnic composition of the U.S. inventors and cross-border mobility of inventors within the firm . In another recent paper, Berry (2018) studies global patent production within multinational firms and finds that "knowledge network embeddedness" with the headquarters, host country and other countries (measured as ratio of backward citations of patents within a context to total backward citations), increases future patent production for MNEs . Choudhury and Kim (2018) exploit a natural experiment and supply shock of Chinese and Indian migrant inventors in the U.S. to find that that ethnic migrant inventors are instrumental in transferring contextual knowledge (i.e. knowledge locked in geographic regions), such as the knowledge of herbal medicine, across borders.

In this context, our paper is a first attempt to link these two literatures, by finding important and economically significant effects of inventor migrants in boosting innovation in their receiving countries on specific technologies prevalent in their home countries, thus impacting cross-country innovation dynamics. Our paper finds a robust pattern of migrant inventors shaping innovation dynamics in their receiving countries, for particular technologies with rich patenting activity in their home countries prior to their move. In that sense, our findings generalize some of the important findings by Moser et al. (2014) on the spike of innovation in chemistry-related fields due to the inflow of Jewish scientists and inventors to the US in the early 1930s (summarized above); as well as findings by Bernstein et al. (2018) who study patenting behavior of immigrant inventors to the United States in recent decades and find that these inventors tend to "import" foreign technologies into the US (which they measure by the higher propensity of these migrant inventors to cite foreign patents and to work with foreign inventors). Beyond studies that focus on particular countries or historical episodes, our paper –to the best of our knowledge– is the first one to use contemporaneous data to establish at a global scale that migrant inventors do shape technology-specific innovation dynamics. This is what we consider the main contribution of our study.

We arrived at these findings by linking and analyzing several sources of data for 95 countries around the globe. First, we use data from the OECD on patenting activity reported by the United States Patenting Office (USPTO) for 651 technology subclasses as defined by the International Patenting Classification (IPC). Our focus on particular technologies is aligned with a rich prior literature in innovation that has used classification of patents according to technologies to study knowledge relatedness and technological distance between countries (e.g., Jaffe, 1986, 1989; Breschi et al., 2003).[2] To our measure of innovation based on patent classification, we incorporate data on bilateral stocks of migrant inventors compiled by Miguelez and Fink (2017), that measures the presence of foreign inventors in every host country. As

[2]Specifically, Breschi et al. (2003) employ a measure of knowledge-relatedness, using co-classification codes contained in patent documents, and examine the patterns of technological diversification of the whole population of firms from the United States, Italy, France, UK, Germany, and Japan patenting to the European Patent Office from 1982 to 1993.

will be described in detail below, we use these data to study the relationship between the international mobility of inventors and the spike in patenting activity in their receiving countries, in particular technologies that their home countries have a technological advantage in. To measure this we employ the Revealed Technological Advantage (RTA) measure, based on Soete (1987). For each country, technology and year in our sample we quantify its RTA, and use it to measure the yearly intensity with which a country specializes in a given technology. For any technology, an RTA above 1 implies that the inventors in a country in a given year filed proportionally more patents than the world as a whole.

Our exercise looks at two different outcomes to measure the dynamics of specialization of a country in a given technology using two decade-long periods, 1990-2000 and 2000-2010. First, we construct a binary variable that takes the unit value if a country-technology pair achieved a RTA of 1 or more in a period of ten years conditional on that country having started off the decade with zero patent applications in that same technology. We refer to this phenomenon as a technological "take-off". Second, in order to study accelerations, we calculate the decade-long growth rate in the number of patent applications for each country-technology pair (which naturally is defined only for country-technology pairs with some patent activity in the baseline period).[3] We then proceed to explore the extent to which the presence of migrant inventors from (alt. to) countries that have a technological advantage in a specific technological class explain the take-off and acceleration of that same technology in their receiving (alt. sending) countries over the course of the following decade.

In order to deal with endogeneity concerns arising from migrant inventors choosing their destination based on private information, the existence of previous trends on technology-specific patent production, or the presence of any other omitted unobservable variable that could bias our estimates, we adopt an instrumental variable approach. In particular, we use historic

[3] These two dependent variables are consistent with some of our previous work focused on measuring dynamics of comparative advantage based on international trade data (e.g., Bahar et al., 2014; Bahar and Rapoport, 2018; Bahar et al., 2019).

migrant networks to instrument for the presence of inventor migrants from the same nationalities. We do so based on data on bilateral migrant stocks compiled by Ozden et al. (2011) that include migrant figures for years 1970 and 1980.

Our baseline estimates also control for trade and for foreign investment to and from countries of destination of the migrant inventors, given that these flows also could explain patterns in the diffusion of knowledge.[4] Finally, in order to rule out the possibility that our results are being driven by prior trends not related to the actual presence of migrant inventors, we perform a number of falsification tests which make our main results disappear. We perform a large number of additional robustness tests to deal with possible alternative explanations to our results, which we explain in detail below.

Finally, we complement our results with suggestive evidence based on patent-level data that supports our main findings, using the PatentsViews dataset (USPTO, 2018). This dataset allows us to track patenting activity for each inventor across time and space, based on USPTO records. Using these data we find that, on average, the very first few patents within every country-technology pair are twice as likely of being invented by a group of inventors (or an inventor) that includes at least one (or is a) migrant with prior patenting experience in that same technology; this as compared to patenting activity in that same country-technology pair occurring at a later stage.

The rest of the paper is structured as follows: Section 2 outlines our empirical strategy; Sections 3 summarizes our main results; Section 4 conducts sub-sample analysis and summarizes results from several robustness checks; Section 5 presents suggestive evidence consistent with our overall findings analyzing patent-level data; Section 6 concludes. There is also an Online Appendix that accompanies the paper.

[4]For a thorough review on the different drivers of international knowledge diffusion, see Keller (2004).

2 Empirical strategy

2.1 Research question

We investigate the relationship between international migration flows and the dynamics of the innovation in migrants' receiving and sending countries. This question follows a research agenda exemplified in Bahar and Rapoport (2018), where we explored the role of migration –generally defined– on comparative advantage dynamics using exports data. The main conclusion from that study is that migrants serve as drivers of knowledge diffusion, which is reflected in the ability of countries to become significant exporters of the same goods that the migrants' countries of origin specialize in.

The results by Bahar and Rapoport (2018), suggestive of migrant-driven knowledge diffusion, lack specificity in terms of the underlying channels through which this process occurs. In this study we shift our focus to innovation dynamics and the role that migrant inventors –a particular subset of high-skilled migrants– play in it. We are interested in whether countries' ability to innovate in specific technologies (without prior patenting activity) is influenced by the presence of migrant inventors. Specifically, we ask the following question: can migrants induce patenting activity in their receiving (alt. sending) countries in the same technologies that their home (alt. destination) countries have an advantage in? For the sake of better understanding, let us use a simple example. Suppose there are two countries in the world: Israel (a country that specializes in patenting water technologies) and Chile (a country that specializes in patenting mining technologies). The analogous question then becomes whether the presence of more Israelis in Chile can explain its specialization in water technologies, and whether this same presence is also associated with the ability of Israel to specialize in mining technologies, as measured by patents applications.

2.2 Main data sources and sample construction

Data on patent applications (which we refer to as patents production as well throughout the paper) come from the OECD Stat database[5]. It counts all patents applications registered by the US Patent and Trademark Office (USPTO) by country of the inventor(s). The count disaggregates the number of patents for each technology subclass based on the International Patent Classification (IPC). An IPC subclass is defined by four characters, letters and numbers; in the paper whenever we refer to a technology, we are referring to an IPC subclass (which we often refer to as IPC code, too). The original dataset covers patenting of 121 countries and it extends from years 1976 to 2011. The assignment to patents to countries is based on the declared residence of the inventor(s) of the patent.[6] The dataset also includes figures for patent granted by the USPTO, also per country and IPC code; as well as all patent applications to and granted by the European Patent Office (EPO), which we also incorporate in our analysis for robustness checks. Our baseline specification, however, uses patent applications to the USPTO unless otherwise noted, given its more ample coverage of patenting activity. The reason we use patent applications as opposed to granted patents is because the latter typically appears in the dataset a few years after the invention actually happened. Hence, patent applications better fit our purposes of measuring production of innovation in a given year.

Our second main source of data is bilateral international stock of inventors compiled by Miguelez and Fink (2017). The dataset measures for every pair of countries i and j the number of patents by inventor from country i in country j and vice versa. It is based on patent applications filed under the Patent Cooperation Treaty (PCT), and has data for about 200 countries. The figures in this dataset are an imperfect measure of the stock of foreign inventors in each country by nationality of the inventor and year. They are imperfect because the number of inventors is contingent on their patenting

[5]https://stats.oecd.org

[6]For the (relatively few) cases of global collaborative patents (e.g., patents with inventors residing in different countries), the dataset assign a patent to each one of the countries of the inventors.

activity. In other words, it could count the same foreign inventor multiple times or, on at the other extreme, ignore her in a given year. For instance, if a foreign inventor living in country j files two patents in year t, then she would be double-counted. On the other hand, if a foreign inventor living in country j has patenting activity on years $t-1$ and $t+1$, but not in year t, then she would not be accounted for in the data in year t. To overcome this possible fluctuations, we compute the average stock between 1981 and 1990 of inventors living in each country i from each country j as our measure for 1990, and the average stock between 1991 and 2000 for our measure of inventor migrants in year 2000. While this is not a perfect solution, the average would not be heavily driven by particular outliers in the data. Despite these important caveats, we refer to these numbers as the stock of inventor migrants throughout the paper.

We include in our main dataset other bilateral measures to use as baseline controls: FDI stocks as well as data on bilateral trade. The FDI data comes from the OECD Stat database, and tracks FDI flows to or from OECD member countries (thus it also reports FDI for non-OECD as long as it is to or from an OECD partner).[7] Using these data we compute FDI stocks for the periods 1985 to 1990, and 1991 to 2000. We also use bilateral trade data that come from UN Comtrade with corrections implemented by Hausmann et al. (2014). With this dataset we compute stocks of bilateral trade for the periods 1985 to 1990, and 1991 to 2000, to be used as baseline controls. Both the FDI and trade flows are deflated using the US GDP deflator (base year 2000) from the World Development Indicators (WDI) by the World Bank before being transformed into stocks.[8]

We complement our dataset with overall bilateral migration from Ozden et al. (2011), which we use as part of our identification strategy. The migration dataset consists of total bilateral working age (25 to 65 years old)

[7]The FDI data by the OECD includes all financial flows that are cross-border transactions between affiliated parties (direct investors, direct investment enterprises and/or fellow enterprises) recorded during the reference period. The main financial instrument components of FDI are equity and debt instruments.

[8]We use 1985 as the lower limit for calculating these stocks given source data limitations.

foreign born individuals for years 1960, 1970, 1980, 1990 and 2000.

The final sample resulting from merging all the different datasets described above includes figures on patent applications and on migrant inventors for 95 countries across 651 different technology subclasses (i.e., 4-characters IPC codes). The list of countries with relevant statistics are presented in Online Appendix Section A. The final number of countries is a result of limiting the sample only to countries with some patenting activity in any technology subclass and any presence of migrant inventors. In order to measure decade-long changes in patenting activities for country-technology pairs, we define two decade-long periods for the analysis which are 1990-2000 and 2000-2010.

2.3 Empirical strategy

The aim of the paper is to study dynamics in patent production by a country in a particular well-defined technology subclass (measured by patent applications) as a function of the presence of foreign inventors from countries that specialize in that same technology. To do so, we need a measure that quantifies the extent to which a country specializes in a particular technology. Our choice is the Revealed Technological Advantage (RTA) index, based by Soete (1987) which in turn is analogous to the Revealed Comparative Advantage (RCA) index by Balassa (1965) used in international trade.

We compute the RTA for each country and technology subclass in a given year as follows:[9]

$$RTA_{c,p} \equiv \frac{patents_{c,p} / \sum\limits_{p} patents_{c,p}}{\sum\limits_{c} patents_{c,p} / \sum\limits_{c}\sum\limits_{p} patents_{c,p}}$$

Where $patent_{c,p}$ is the number of patent applications by inventors in country c in technology subclass p. This is an annual measure. For example, in the year 1990 about 3.25 percent of patents granted to Austrian inventors belonged to the technology A63C, which corresponds to "Skates, skis, water-

[9]This is analogous to how RTA is computed in the dataset by OECD (2013), though we construct the index ourselves.

shoes; roller skates; courts; and rinks". Overall, patents granted that year in that same technology to inventors all over the world represented 0.14 percent of all patents granted.

Hence, Austria's RTA in technology A63C in the year 1990 was $RTA_{AUT,A63C} = 3.25/0.14 \approx 23$, indicating that, relative to total patenting, inventors in Austria patent 23 times more in technology A63C than the world as a whole. Our choice of RTA is a proper one in this case because of the following reasons. First, we are interested in measuring the specialization of one country in a particular technology with respect to the rest of the world, and not with respect to another single country. Second, our measures of patent applications are all based on a single patent agency, and thus the numbers are comparable across countries and years. Third, similarly to Balassa's RCA, the benchmark value of 1 and above has an intuitive meaning, as can be understood through the example above.

To study the question at hand with our sample, we follow the empirical specification by Bahar and Rapoport (2018) and estimate:

$$\begin{aligned} Y_{c,p,t\rightarrow T} &= \beta_{im} \sum_{c'} inventors^{im}_{c,c',t} \times R_{c',p,t} + \beta_{em} \sum_{c'} inventors^{em}_{c,c't} \times R_{c',p,t} \\ &+ \beta_{FDI} \sum_{c'} FDI_{c,c',t} \times R_{c',p,t} + \beta_{trade} \sum_{c'} trade_{c,c',t} \times R_{c',p,t} \qquad (1) \\ &+ \gamma Controls_{c,p,t} + \alpha_{c,t} + \eta_{p,t} + \varepsilon_{c,p,t} \end{aligned}$$

The definition of the dependent –or left hand side (LHS) variable– $Y_{c,p,t\rightarrow T}$, changes with the estimation of different outcomes measuring changes in the intensity with which a country produces patents in a given technology. The first outcome we use is a binary variable which we denominate a technological "take-off", and it measures cases where a country with no patent applications whatsoever in a given technology at time t gains advantage in that same technology at time $t + 10$. In this case, we define $Y_{c,p,t\rightarrow T}$ as a binary variable, equal 1 if the number of patent applications in country c and technology p results in having a RTA of 1 or more in the period of time between t and T, conditional on having zero patent applications in that same technology at the beginning of the period ($patents_{c,p,t} = 0$). That is:

$$Y_{c,p,t \to T} = 1 \text{ if } patents_{c,p,t} = 0 \text{ and } RTA_{c,p,T} \geq 1$$

We impose two additional conditions on our take-off measure to avoid our results driven by noise: first, the country-technology pair under consideration must keep its RTA value above 1 for four years after the end of the year T (e.g., have a minimum RTA of 1 during the years $[T, T+5]$); and second, the country-technology pair under consideration must have had a RTA value equal to 0 during all four years before the beginning of year t (e.g., have a maximum RTA of 0 during the years $[t-5, t]$).

We alternate our LHS variable with a measure of growth in patent applications for every country and technology subclass between years t and T. In that case $Y_{c,p,t \to T}$ is simply the annual compound average growth rate (CAGR) in the number of patents in technology p granted to inventors in country c between years t and T, conditional on $patents_{c,p,t} > 0$. That is:

$$Y_{c,p,t \to T} = \left(\frac{patents_{c,p,T}}{patents_{c,p,t}} \right)^{1/T-t} - 1 \text{ if } patents_{c,p,t} > 0$$

Our main variables of interest are denoted by $\sum_{c'} inventors^{im}_{c,c',t} \times R_{c',p,t}$ and $\sum_{c'} inventors^{em}_{c,c',t} \times R_{c',p,t}$, where $R_{c',p,t} = 1[RTA \geq 1]$. These variables can be interpreted, respectively, as the stock at time t of immigrant inventors from and of emigrant inventors to other countries (denoted by c'), that specialize in the production of patents classified under technology subclass p, as indicated by the dummy variable $R_{c',p,t}$.

As controls, we also include the sum of the stock of FDI (inflows plus outflows) and the sum of the stock of trade (imports plus exports), using the same weighting structure as above. The rationale for including these controls is explained below in the identification section. In addition, we include country-year fixed effects, denoted as $\alpha_{c,t}$, to control for any country level time-variant characteristics that correlate with both national migration determinants and aggregate productivity levels such as income, size, institutions, etc. $\eta_{p,c}$ represent technology-year fixed effects, to allow for a different constant for each combination of year and IPC technology subclass.

We also include a vector of controls for baseline variables when measuring using CAGR on the LHS: the baseline (initial) level of patent applications for that same technology in that country, as well as the previous period CAGR of patent applications in the same technology to control for previous trends. We also add a binary variable indicating whether $patents_{c,p,t-10} = 0$ (at the beginning of the previous period, i.e., 1980 or 1990). We do this to control for possible distortions created by our own calculation of the *lagged CAGR* variable, which sums 1 to the number of patents in both the numerator and denominator, in cases with zero patents in the baseline year for the *lagged* CAGR.

All level variables are transformed using the inverse hyperbolic sine (MacKinnon and Magee, 1990). This linear monotonic transformation behaves similarly to a log-transformation, except for the fact that it is defined at zero. The interpretation of regression estimators in the form of the inverse hyperbolic sine is similar to the interpretation of a log-transformed variable.[10] Thus, since our LHS are not transformed using a logarithmic scale, then the interpretation of the estimators are linear-log.

2.4 Identification

Our main goal is to get unbiased estimators for β_{im} and β_{em}. This is challenging, as one might expect that the choice of country for foreign inventors might be correlated with dynamics of specialization in certain technologies. We try to overcome this through a number of ways.

First, we include as controls the overall trade and investment to and from the same countries where the inventors are coming from or going to, using the same weighting scheme $\sum_{c'} FDI_{c,c',t} \times R_{c',p,t}$ and $\sum trade_{c,c',t} \times R_{c',p,t}$. This controls for the possibility that it is trade and FDI with the same countries where the inventors come from or go to (denoted as c') explain innovation dynamics in country c. Using the same weighting scheme deals with being able to identify the total trade and FDI to those same countries

[10] The inverse hyperbolic sine ($asinh$) is defined as $log(y_i + \sqrt{(y_i^2 + 1)})$. Except for small values of y, $asinh(y_i) = log(2) + log(y_i)$.

c'. While ideally we would include trade and investment from those countries that relate to each particular technology p, that data is not available. Since our weighting procedure accounts for the total trade and FDI from those countries, they are inclusive of flows that relate to particular technologies. We believe these measures the way they are computed, even if imperfect, serve our purpose to control for these plausible channels.

Of course, there is a genuine discussion to have on whether these are, in fact, "bad controls". In other words, if the inflow of inventors from countries that specialize in a certain technology subclasses triggers more trade and investment that in turn boost innovation dynamics in the receiving country, then we would be underestimating our overall effect. However, we decide to keep them in the baseline specification as we are interested in estimating β_{im} and β_{em} which would measure the partial correlation (or marginal effect) regardless of trade and investment. However, in Online Appendix Section E we show that our results are robust to excluding these controls.

Second, in spite of adding those flows some important concern remains: there might be other country-technology characteristics, perhaps unobservables, that can explain both the inflow and outflow of inventors and dynamics of patent production. Our estimations, thus, might still suffer from endogeneity. We try overcoming this problem with the use of an instrumental variable both for immigrant and emigrant inventors. In particular, for each country-technology-year combination, we instrument the immigrant and emigrant inventors with the total stock, lagged by 20 years, of immigrants from and emigrants to those same countries (e.g., we apply the same weights defined by$R_{c',p,t}$). In that sense, the first stage of our 2SLS methodology involves estimating the following two equations:

$$
\begin{aligned}
\sum_{c'} inventors^{im}_{c,c',t} \times R_{c',p,t} = \gamma^1_{im} & \sum_{c'} immigrants_{c,c',t-20} \times R_{c',p,t} \\
& + \gamma^1_{em} \sum_{c'} emigrants_{c,c',t-20} \times R_{c',p,t} \\
& + \gamma^1_{FDI} \sum_{c'} FDI_{c,c',t} \times R_{c',p,t} \\
& + \gamma^1_{trade} \sum_{c'} trade_{c,c',t} \times R_{c',p,t} \\
& + \delta^1 Controls_{c,p,t} + \alpha_{c,t} + \eta_{p,t} + \varepsilon_{c,p,t}
\end{aligned}
$$

$$
\begin{aligned}
\sum_{c'} inventors^{em}_{c,c',t} \times R_{c',p,t} = \gamma^2_{im} & \sum_{c'} immigrants_{c,c',t-20} \times R_{c',p,t} \\
& + \gamma^2_{em} \sum_{c'} emigrants_{c,c',t-20} \times R_{c',p,t} \\
& + \gamma^2_{FDI} \sum_{c'} FDI_{c,c',t} \times R_{c',p,t} \\
& + \gamma^2_{trade} \sum_{c'} trade_{c,c',t} \times R_{c',p,t} \\
& + \delta^2 Controls_{c,p,t} + \alpha_{c,t} + \eta_{p,t} + \varepsilon_{c,p,t}
\end{aligned}
$$

where $immigrants_{c,c',t-20}$ is the 20-year-lagged stock of immigrants in country c from country c', and $emigrants_{c,c',t-20}$ is the equivalent figure for emigrants from country c in country c'. Note that the two variables use the same weighting scheme as before, thus counting only migrants from and in countries that specialize in technology subclass p. The second stage of this 2SLS estimation follows specification (1), using instead the predicted values for both $\sum_{c'} \widehat{inventors^{im}_{c,c',t}} \times R_{c',p,t}$ and $\sum_{c'} \widehat{inventors^{em}_{c,c',t}} \times R_{c',p,t}$ which we define as Z_{im} and Z_{em}, respectively. These are computed based on the first-stage estimation, in the following way:

$$Z_{im} = \hat{\gamma}^1_{im} \sum_{c'} immigrants_{c,c',t-20} \times R_{c',p,t} + \hat{\gamma}^1_{em} \sum_{c'} emigrants_{c,c',t-20} \times R_{c',p,t}$$

$$Z_{em} = \hat{\gamma}^2_{im} \sum_{c'} immigrants_{c,c',t-20} \times R_{c',p,t} + \hat{\gamma}^2_{em} \sum_{c'} emigrants_{c,c',t-20} \times R_{c',p,t}$$

For our instruments to be valid they should be able to explain enough variation of the endogenous variables. We expect this to be the case because historic and well-established migrant communities should work as a pull factor for the decision of inventors to migrate to particular countries. We find this to be the case in our sample based on the reported first-stage statistics (with some exceptions, which are discussed thoroughly).[11] Naturally, there could be concerns that the observed first-stage correlations between migrant inventors and overall migrants are artificially being driven by the weighting scheme, and not by the actual correlation between the bilateral number of inventors and the lagged bilateral number of migrants. However, this is not the case: lagged migrant stocks have a strong explanatory power on current stock of migrant inventors. We include evidence of this in Online Appendix Section B.

In addition to the explanatory power of the first stage, for our instruments to be valid –and thus to be able to interpret our 2SLS estimators as causal– they needs to comply with the *exclusion restriction*. In our case, this exclusion restriction can be verbalized as follows: it must be *technology-specific* production (e.g., patent applications) in any given country are not correlated with historic presence of migrants other than through the presence of inventor migrants today.[12] In particular, our argument for the validity of

[11]In all of our 2SLS estimations we report the Kleibergen-Paap F statistic to be used to determine whether instruments are weak, which according to Stock and Yogo (2005), must be above 16.78 when using two endogenous variables and two instruments. We acknowledge that these critical values are not strictly usable in the case when we do not assume i.i.d., but for the most part, unless otherwise noted, our Kleibergen-Paap F statistics are high enough that there are no reasons for concern regarding weak instrumentation.

[12]Note that since we include trade and FDI in our 2SLS estimation (i.e., they are also

the instruments is that the existence of a historic migrant community from country c' in the destination country explains the flow of migrant inventors, a particular subset of high-skilled migrants; but –at the same time– that historic migrant community from country c' does not explain future dynamics of patent production other than through the inventor migrants that had arrived later on. Furthermore, to be able to interpret our 2SLS estimators as causal we also have to assume that countries do not engage in *technology-specific* innovation agreements based on their historic migrant networks that are not captured via FDI or trade (since we are controlling for those flows, too).

In summary, we believe these are reasonable assumptions to make, though we acknowledge there might still be weaknesses in our approach. Thus, to complement our efforts in establishing the relationship we also perform a number of falsification tests showing that our results are indeed driven by the flow of inventors, and do not respond to previous trends or other variables, observables or not, that are not accounted for in our main estimation.

2.5 Descriptive statistics

The OECD patents data includes information on the number of patent applications and patents granted by country on a yearly basis. We choose to present our main results using patent counts based on USPTO applications because, at least numerically, it includes much higher patenting activity than the EPO records, as well as a broader coverage in terms of countries and technologies.[13] Figure 1 presents total USPTO patents applications and grants by year in the dataset, for years 1990, 2000 and 2010. As expected, the number of applications surpasses the number of grants. Though, it is important to notice the vast majority of patents granted in any given year most certainly belong to a much earlier batch of patent applications. In our dataset we cannot identify what was the year of application for the patents

part of the first stage, though not used to compute Z_{im} and Z_{em}) we already control for the fact that the historical presence of migrants might affect future innovation through FDI and trade.

[13]However, in Online Appendix Section G we present results using EPO figures that are robust to our baseline estimations using USPTO data.

granted we see in any year. As explained above, since we are trying to get a measure of patent production in real time, we use patent applications to construct our variables used in the analysis.[14]

[Figure 1 about here.]

Table 1 present descriptive statistics for the sample we've put together. Panel A presents the summary statistics for the subsample that focuses on technology take-offs (i.e., for all observations of c, p and t for which $RTA = 0$), while Panel B does so for the the subsample focusing on growth of patent production (i.e., for all observations of c, p and t for which $patents_{c,p,t} > 0$).

[Table 1 about here.]

Panel A of Table 1 describes that the unconditional probability of a take-off for the average country and average technology subclass, pooling observations for two decades (1990 to 2000 and 2000 to 2010), is 2.2 percent. Note this is based on the sample limited to country-technology pairs with zero patent applications at the initial year of each decade. Panel B presents statistics based on the complementary sample; e.g., with at least one patent application at the beginning of each decade for every country-technology pair. This sample is used to measure the impact of migration on growth of technologies in terms of patent applications. The 10-year CAGR for the average country-technology pair –also pooling observations for two decades– is 0.7%, and varies between -30% to 80% in both extremes for some country-technology pairs. In this sample the baseline number of patent applications for the average country-technology pair is about 16.45. Notably, the number of observations that make up the "technology take-off" sample is almost six times as large as the sample described in Panel B. This is not surprising, since the vast majority of country-technology pairs have, in fact, no patent activity.

The tables also include figures for immigrant and emigrant inventors weighted using the scheme used on the right hand side of Specification

[14] See Online Appendix Section G2 for results using granted patents. The results are robust to our baseline estimations that use patent applications data.

(1). According to Panel A, which focuses on take-offs, the average country-technology pair in the sample has about 24 inventors that have immigrated from countries that specialize in that same technology, and about 80 inventors that have emigrated to countries specializing in that technology. Those same figures in the sample summarized in Panel B are about 710 and 610, respectively. The larger numbers of average inventor migrants in Panel B responds to the fact that such sample is composed mostly by developed nations, which host many more inventors (as those include only country-technology pairs with some patent applications).

The table also summarize our instrumental variables, which are the twenty year lagged stock of immigrants and of emigrants, weighted using the same weighting scheme as our right hand side variables of interest. These figures are about 71 thousand immigrants and 173 thousand emigrants in Panel A, and 556 thousand immigrants and 498 thousand emigrants in Panel B. As expected, these numbers are significantly larger than the number of inventor migrants, as inventors are only a very small subset of all migrants. Finally, the table also has statistics on total trade and FDI, in billions of dollars, for both sub-samples.

3 Main Results

The main question we aim to answer is whether a country can become a significant innovator of a particular technology –what we call a technology take off– if it has immigrant inventors from (or emigrant inventors in) other countries that specialize in patenting activity in that same technology. A simple look at raw data, represented in Figure 2, presents some evidence that such is the case. Average take-off rates of country-technology pairs are higher with increasing number of inventor immigrants from other countries that specialize in those same technologies. In particular, the figure shows the average take-off rate for each terciles of the stock of immigrant inventors. The probability of a technology take-off is below 1% for country-technology pairs for which the stock of inventor immigrants is in the first tercile, as opposed to above 4% for country-technology pairs for which the treatment variable is in the top tercile of the distribution. Note that for the sample as a whole, the unconditional probability of a take-off, as evidenced in Table 1, is 2.2 percent.

[Figure 2 about here.]

3.1 OLS and 2SLS estimations

Many confounding factors could explain Figure 2, of course. Therefore, we present results using more rigorous estimation techniques. The estimation of Specification (1) is presented on Table 2. The upper panel estimates the changes in the probability of technology subclass take-offs for the average country-technology pair as a function of migrant inventors. Technology take-offs are cases where a country achieves a RTA of one or more within a decade, starting off from no patenting activity (see Section 2.3 for formal definition). In the estimations, as mentioned above, all of the regressors have been transformed using the inverse hyperbolic sine, and therefore, the interpretation of the coefficients correspond to semi-elasticities (i.e., linear-log). The first three columns show results using OLS as the estimation technique, whereas Columns 4 to 6 use 2SLS estimations, using the instrumentation explained in the Section 2.3.

[Table 2 about here.]

The results of Panel A estimate the partial correlation of our variables of interest: immigrant inventors from and emigrant inventors in countries specializing in a given technology subclass at the beginning of the decade (separately in Columns 1 and 2, and jointly in Column 3) with respect to the take-off of the same technology by a country.[15] Results in Columns 1 and

[15]Our estimates are robust to using a maximum likelihood estimator given the binary distribution of our dependent variable in Panel A (specifically, the complementary log-log method which is more appropriate for our setting, following Singer and Willett, 2009) . It is also robust to using the methodology suggested by Horrace and Oaxaca (2006), to deal with the possibility of our results being driven by outliers. For more details, see discussion in Online Appendix Section C. In addition, Online Appendix Section D presents results using alternative left-hand side variables, including a more widespread –and less

3 show that a twofold larger stock of immigrant inventors from countries that specialize in technology p is associated with an increase in the probability of the receiving country specializing in patent applications in technology p by of 0.51 percentage points. Given that the unconditional probability of a take-off is 2.2 percent, this represent an increase in the probability of about 23%. The estimator for emigrant inventors is not statistically different from zero when jointly estimated with immigrant inventors (Column 3), and therefore we refrain from interpreting it.

The economic significance of this number is quite large. Given that the stock of immigrant inventors in the sample is of about 24 people for the average country-technology pair, and a standard deviation of 135 inventors, a twofold increase implies a relatively small number of inventors. Thus, according to our results, a small number of migrant inventors has significant and large explanatory power on the likelihood the receiving country will gain advantage in a new technology subclass, for which the country had no patent activity beforehand.

Columns 4 to 6 in the upper panel of Table 2 replicate the results using a 2SLS estimator. As explained above, our instrumental variable is the 20-year lagged total stock of immigrants from (and of emigrants in) the same countries as the migrant inventors. Given that we have two instruments, we can use them in estimations that include both variables of interest separately (Columns 4 and 5) as well as jointly (Column 6). The way of thinking about this instrumental variable is that historic migrant networks could serve as pull factors for innovators to migrate, which is reflected in our first stage. Note that the table reports the Kleibergen-Paap F statistics, which are large enough to eliminate any concerns of weak instrumentation.

The results are qualitatively similar to the OLS results, but higher in magnitude by a factor of 2 to 3. This is somewhat counterintuitive for the estimation of β^{IM}: if anything we would expect a positive bias in the OLS estimates, rather than a negative one, as unobserved forces that lead to more

restrictive– measure of take-off: a binary variable that takes the value of 1 whenever a country-technology pair goes from zero patent applications to any number higher than zero.

innovation might also pull immigrant inventors (the opposite would happen for emigration, but since the the OLS coefficient is indistinguishable from zero, statistically speaking, we don't focus on that result). But in fact, our 2SLS results are inconclusive when it comes to understanding whether the OLS bias is positive or negative: even though magnitude of the 2SLS estimates are larger than the OLS ones, the standard errors have also increased, and thus we cannot reject the hypothesis that both estimates for β^{IM} are statistically different.[16] Relying on the 2SLS estimates, the effect of a twofold increase in the number of immigrant inventors results in an increase of between 50% to 60% in the likelihood of the receiving country specializing in patent applications in the same technologies that the home countries of the migrants are specialized in.

Panel B of Table 2 estimates the partial correlation of our variables of interest on the growth rate of technology-specific patent applications for the average country, as a function of inventor immigrants from (and inventor emigrants in) other countries that specialize in that same technology. The sample used is limited to those country-technology pairs for which the initial value of patent applications is above zero. This for the simple reason that it is not possible to compute growth rates otherwise. But there is another, more fundamental, reason. In essence, this distinction allows us to focus on innovation dynamics for technologies that already are being patented in the country. In those cases, arguably, there is already a critical mass of inventors with knowledge on that specific technology subclass, as opposed to cases where there is no prior patenting activity whatsoever, such as country-technology pairs in the subsample used in Panel A. The distinction between

[16] In Bahar and Rapoport (2018), were we use a similar setting but a different instrument, we get similar results where the 2SLS estimates are inflated as compared to the OLS ones. That same issue is brought up by Frankel and Romer (1999) in their seminal paper aiming to estimate the causal effect of trade on economic growth. They write: *"we conclude that the most plausible explanation of the bulk of the gap between the IV and OLS estimates is simply sampling error. This implies that our most important finding is not that the IV estimates of trade's effects exceed the OLS estimates, but rather that there is no evidence that the IV estimates are lower. In addition, it implies that our IV estimates may be substantially affected by sampling error, and thus that the OLS estimates are likely to be more accurate estimates of trade's actual impact on income."*

an extensive margin (i.e., take-offs) and an intensive margin (i.e., growth) is often used in the international trade literature when studying the composition dynamics of countries' exports baskets.[17]

The results from Panel B present a qualitatively similar conclusion than those in Panel A. The OLS estimations (Columns 1-3) imply that, for the average country, a twofold increase in the number of immigrant inventors from other countries specializing in technology p explain a higher growth rate in patenting activity of the same technology p of 0.36 to 0.46 percentage points per year over the following decade. Similarly to before, based on Column 3, the results do not support an analogous finding for emigrant inventors (e.g., growth in patent activity for a given technology in a country cannot be significantly explained by the presence of its inventors in other countries specializing in that same technology). Since the decade-long unconditional growth rate for patent applications is 0.7% (see Panel B, Table 1), then the marginal effect for immigrant inventors according to Columns 1-3 corresponds to an increase of up to 65%. Note that in this sample, however, the average number of immigrant inventors is much larger, and corresponds to about 700.

Columns 3-6 estimate the same specification using 2SLS, with the instrumental variable strategy discussed above. However, in this case, the first stage statistics are not as large as the ones in Panel A, implying there could be a weak instrumentation, and therefore we refrain from concluding anything from those results. Note that the difference in the explanatory power of the instrument could be expected as the samples used in both panels are very different.

Note, too, that the estimators for our control regressors, namely FDI and trade, are volatile around the zero value –seldom statistically significant– across the different specifications in both panels. Their inclusion in our

[17]In Online Appendix Section D2 we use a measure of growth in patent applications that is defined for country-technology pairs with initial value of zero (i.e., a symmetric percentage change). With this growth measure as the dependent variable we are able to estimate our empirical specification on *all* the sample and find robust results; though the results are driven by the observations with no initial patenting activity. Our results using this aggregated measure, however, are statistically weaker.

model is part of our identification strategy to reduce concerns of biases in our regressors of interest β_{im} and β_{em}, and therefore their estimators are not thoroughly discussed.

Our main results support the idea that migrant inventors facilitate the spread of ideas reflected in significant patenting activity in technologies that the countries where the migrants inventors come from specialize. Given our specification, we rule out that this effect is driven by other international aggregate flows between those countries such as trade and investment. The economic significance of these effects are not small: a twofold increase in the stock of immigrant inventors from countries that are highly innovative in a given technology explains a 25% to 50% increase in the probability of the receiving country experiencing a take-off in patent activity in that same technology. This twofold increase for the average country in our sample, as noted above,, corresponds to about 25 inventors (and a standard deviation of about 135).

While in the main body of the paper we use patent applications as our main data source, Table G2 in the Online Appendix Section G shows that our main results are robust to using granted patents as opposed to patent applications. Naturally, there could be important gaps between the time of the application of the patent and the time of its acceptance by the USPTO, and that time gap could be problematic in our setting. However, the fact that the results are robust to using granted patents is complementary to our main estimation, as it provides suggestive evidence that the process through which immigrant inventors affect the dynamics of patent applications for a given technology is also reflected in the ability of the receiving country to convert some of those applications into granted patents.

3.2 Falsification tests

While our estimation methods in the previous section aim to deal with plausible endogeneity concerns between the flow of inventor migrants and patenting activity of countries, we acknowledge there could be violations to the exclusion restriction of our instrumental variable approach. The nature of our macro-level dataset, while allow us to make more general conclusions, also poses important challenges to our ability to perfectly identify the relationship we are studying.

However, in order to further deal with some remaining endogeneity concerns we propose two tests to explore whether our results are indeed consistent with the possibility of inventor migrants impacting patenting activity of their receiving countries, or alternatively, it is driven by idiosyncratic factors inherently in our data which we are not accounting for. For this purpose we perform two *falsification tests,* based on OLS estimates, by altering the right-hand-side variables of interest. While not a perfect approach for identification purposes, we consider it useful to show that our results do respond to actual variation in migrant inventors and not to existing pre-trends.

First, we replicate Specification (1), but this time using the weighting parameter $R_{c',p,t} = 1$ if $RTA_{c',p,t} = 0$. That is, we exploit variation in inventors migrating from and to countries c' that had zero patent applications in technology p at time t. Table 3 reports the results.

[Table 3 about here.]

When alternating the right-hand side variables of interest in this way, we find that the results are very different than the ones presented above in Table 2: the presence of migrant inventors from or in countries with no patent applications in technology p does not explain technology take offs (Columns 1-3) nor acceleration in patents application growth (Columns 4-6). If anything, we find negative results, implying that these countries are less likely to innovate in those technologies.

Second, we alternate our right-hand side variables of interests by randomizing the total actual number of inventor migrants from and in countries that specialize in a given technology. We simply assign a random number of immigrant and emigrant inventors using a uniform distribution before the weighted sum is calculated (i.e., the RTA values of the partner countries, c' in our notation thus far, are not modified).[18]

[18]In fact, we assigned values based on the 100th draw of random numbers.

The correlation between the variables of interest $\sum_{c'} inventors^{im}_{c,c',t} \times R_{c',p,t}$ and $\sum_{c'} inventors^{em}_{c,c',t} \times R_{c',p,t}$ with their corresponding constructed variables using random data are 0.17 and 0.22, respectively. Since both the real number of inventors and the constructed one using random numbers use the same weighting scheme, the positive correlation is not surprising. However, a deeper look into the data shows that the distribution of both the real and random-inventors variables are very different (see left panel of Figure 3, which plots the kernel density) and that the correlation between the two, while positive, is quite noisy (see right panel of Figure 3). Note that Figure 3 only compares the real and random-inventor variables for immigrant inventors. The figure limits the sample to year 2000 for visualization purposes. As expected, the plots using emigrant inventors are quite similar, and therefore we don't report them.

[Figure 3 about here.]

Table 4 presents the results of estimating our main specification but substituting the main variables of interest $\sum_{c'} inventors^{im}_{c,c',t} \times R_{c',p,t}$ and $\sum_{c'} inventors^{em}_{c,c',t} \times R_{c',p,t}$ with the ones that we constructed using random numbers for inventor immigrants and inventor emigrants. As can be seen, our main results disappear when reconstructing the variables that rely on random, not real, migrant inventors data.

[Table 4 about here.]

These two falsification tests are important as they should reduce any remaining concerns that our results are being driven by spurious correlations or previous trends. Also, they make the case that the results are not driven by any sort of scale effects. Thus, our main results present strong evidence that immigrant inventors play a role in explaining new patenting activities in their receiving countries, particularly in the same technology subclasses that their home countries specialize in.

4 Supplementary analysis

4.1 Heterogeneity of results

In order to study the relationships documented above in more detail, we re-estimate Specification (1) across different subgroups of our sample. We do this to understand whether there are differential trends across several dimensions, but also to explore whether a particular set of observations in the sample is driving the observed overall results. Table 5 summarizes this exercise.

[Table 5 about here.]

The left panel of Table 5 reports OLS estimates both for β_{im} and for β_{em}, while the right panel reports the 2SLS estimates for the same regressors. The table is based on the estimation that focuses on technology take offs (thus, country-technology pair observations are limited to having an initial number of granted patents equal to zero). The first row uses all observations (the same sample as presented in the upper panel of Table 2).

The rest of the rows present results for different cuts of the sample. Across the board, based on the 2SLS estimates, we find that our results typically hold only for immigrant inventors, and not for emigrants, consistently with our findings so far.

Additionally, our results are being driven by both OECD and non-OECD countries alike.[19] Also, the results are particularly driven by the period 2000-2010, where most of the patenting activity is concentrated for the decades we focus on.

Finally, we divide our sample into 8 IPC sections (which correspond to the first "character" of the 4-character IPC subclasses used throughout the paper). These are human necessities (A); performing operations and transporting (B); chemistry and metallurgy (C); textiles and paper (D); fixed

[19] We count countries in our sample as OECD members of only if they had been such prior to the first period studied (e.g., the classification does not count countries as OECD members if they became so during the 1990s or the 2000s). For a complete list of countries in the sample, including which ones we categorize as OECD members, see Online Appendix Section A.

constructions (E); mechanical engineering, lighting; heating; weapons, and blasting (F); physics (G), and electricity (H). While our OLS results do show some heterogeneity in the statistical significance of the results for β_{im}, when it comes to the 2SLS results, we do notfind any particular technology section driving the results. Thus, our results seem to be explanatory across all types of technologies using the broad classifications used in the estimations.

4.2 Further robustness tests

We perform a number of supplementary analysis for robustness purposes, which we briefly discuss in this subsection and refer the reader to the Online Appendix for more details when appropriate.

One possible concern regarding our estimations is that our right-hand side variables measuring migrant inventors, trade and FDI are highly multi-collinear, which could raise concerns our model is misspecified. However, note that all of our estimations include a large number of fixed effects, so our estimations correct for different scales. Additionally, we have computed the Variance Inflation Factor (VIF) for our main OLS estimation (Column 3 of Table 2, which include technology-by-year and country-by-year fixed effects) to assess the availability of enough independent variation among correlated variables. The mean VIF value was 1.31, which is within the acceptable range. Therefore, multi-collinearity seems a less relevant concern for our empirical strategy.

Also, we reestimate the specification estimating the impact of migrant inventors on technology take-offs based on patent applications using a non-linear estimation as it is often done for binary outcomes. In particular we implement the complementary log-log estimator, which is a better estimator than logit or probit if the probability of take off is small (e.g., there are many zeros) as is our case (see Singer and Willett, 2009). We also apply the methodology by Horrace and Oaxaca (2006) that corrects for predicted values of the dependent variable outside the zero to one range. For more details see Online Appendix Section C.

We also find that our results are robust to variations of our left-hand side variables. In particular, when using a binary variable that does not depend on RTA, as well as a growth rate that is defined when the initial value is zero (allowing us to estimate using the sample without splitting it into two subsamples). See Online Appendix Section D for more details and for the results.

Additionally, we find our results are robust to using patent application data based on the European Patent Office (as opposed to the USPTO), and find results that are qualitatively and quantitatively similar. This is robustness test is particularly important because it shows that our results are likely not driven by the "home advantage" effect, given that such bias would naturally be different for USPTO and EPO patents (Criscuolo, 2005).[20]

Another test we perform is to replicate our results using data on granted patents (as opposed to patent applications) according to the USPTO. While granted patents typically involve an important time gap, we still believe it is relevant to show our results are robust to using this measure, as a granted patent is –indeed– a confirmation that the innovation is novel enough. In the context of our results, this is crucial as it suggests our results are not only driven by an uptake on filling patents, but on actual innovation. For the results and more details see Online Appendix Section G.

We also explore the possibility that our results are driven by intellectual property theft practices (along the lines of the evidence on industrial espionage by Glitz and Meyersson, 2017). In other words, whether inventor migrants facilitate the spread of technologies through stealing intellectual property (IP) rather than through knowledge diffusion. To some extent, our results –particularly the ones using granted patents in Online Appendix Section G– address this possibility. This is because our data is based on innovations reported by a formal authority (e.g., the USPTO for our baseline results) that in essence should deal with IP thefts. In addition, the idea of IP theft should be less of a concern in our specification given that our

[20] An alternative approach could have been to limit our sample to OECD triadic patents, but this raises a number of other difficulties. First, as by definition it represents a subsample of all patents, it is unclear –at first sight– what are the biases driving the selection of patents into that group.

country-year fixed effects should control for IP protection intensity in each country at an aggregate level (e.g., though it does not control for variations of IP protection intensity within a country for different technologies). As an additional robustness test, however, we re-estimate our main specification excluding China from the sample. As known China is a country that i) is large in size and therefore it represents an important share of migrants, and ii) is known to have weaker intellectual property protection. Our results are robust to the exclusion of China and are documented in Online Appendix Section F1.

5 Migrant inventors and pioneering patents: using micro-evidence

This section complements our main findings by presenting descriptive evidence coming out from patent-level data. Specifically, we identify patterns consistent with our results, using the PatentsView dataset (USPTO, 2018). PatentsView is based on data files from the USPTO, and is particularly useful for tracking inventors across time and space, as it contains a unique identifier for each one of them.

In that sense, we use this information to explore patents with respect to the role of migrant inventors in patenting in technologies without prior history of patenting in a country. It is important to note that, as opposed to the micro data used by Miguelez and Fink (2017), PatentsView contains no information on the nationality of each inventor. Thus, we cannot identify whether the country of residence declared by the inventor when filing a patent is a country other than where she was born. In other words, using PatentsView we cannot identify whether an inventor in indeed a migrant, in the strict sense of the word. However, the data does allow us to follow each inventor across time and space contingent on patenting activity, and based on that we can identify whether an inventor has moved across borders since the first time she appears in the data.

The sample we have constructed based on the raw PatentsView data contains over 3.5 million patents, most of them with application year after 1974.[21] These patents are linked to over 5.3 million inventors. In the sample, on average, there are about 1.49 inventors per patent. We link each patent in the sample to one country, based on the country of residence reported by its inventor(s).[22] Our sample has patents that originate in 69 countries.[23] In addition, our sample has patents that belong to 640 different technologies as defined by 4-characters IPC codes, based on the first IPC subclass reported.[24]

Our goal in this section is to identify the role of migrant inventors in "pioneering" patents: e.g., the first few patents *ever* in a given country and in a given technology. To do so, we first sort all patents by date of application (day, month and year) within each country and IPC code, and divide these patents in 10 deciles based on chronological order by application date. For the purpose of computing deciles we had excluded from the sample all country-technology pairs with less than 10 patents throughout the whole period available to us in the dataset. The deciles simply identify the chronological order of patents in a given country-technology. Thus, the 1st decile corresponds to the first ever 10% of all patents filed by (inventors in) each country and each IPC code (the "pioneering" patents); while the tenth decile corresponds to the latest, or most recent, 10% of patents within that same cell.

In order to identify whether an inventor can be categorized as a migrant, given that PatentsView does not provide information on the nationality nationality of the inventor, we settled on a broader definition: an inventor that has moved across countries at least once throughout the years, according to her patenting record. While not perfect, we assume this should be a good

[21] We keep in our sample about 2000 patents with year of application before 1974. However, we do exclude a handful of patents with no data on the application date or that that report a year of application that is over 35 years apart from the granting year.

[22] For simplicity, in order to avoid double-counting when assigning patents to countries, we exclude from our sample less than 95,000 patents with inventors reporting residence in multiple countries (i.e., global collaborative patents).

[23] The number of countries is significantly reduced when we impose a restriction of at least 10 patents per country-technology cell in order to properly compute deciles, which we explain more in detail below.

[24] We have also eliminated small number of patents without information on their IPC class.

approximation for our purposes. Similarly, using the inventor unique identification, we also define whether she had worked on a patent classified under the same IPC subclass in a previous point in time *and* in a different country. For the sake of simplicity, we refer to these inventors as "experienced migrant inventors".

We proceed to estimate what is the share of patents with at least one experienced migrant inventor for each decile described above. Given that there are observables and unobservable variables that determine patent activity for each country, technology and year, we estimate the following specification:

$$anymiginventor_{i,p,c,t} = \sum_{k=1}^{10} \beta_k decile^k_{i,p,c} + \theta_{c,t} + \varphi_{p,t} + u_{i,p,c,t} \qquad (2)$$

where $anymiginventor_{i,p,c,t}$ is a binary outcome that takes the value 1 if at least one inventor in patent i (linked to technology p, country c and application year t) is an experienced migrant inventor. $decile^k_{p,c}$ is a dummy variable which takes the value 1 if patent i is in the k_{th} decile of all patents within each technology p and country c in terms of chronological order by application date. For example, $decile^1_{i,p,c} = 1$ if patent i is among the first 10% of all patents, ever, produced in country c and in technology p; while $decile^{10}_{i,p,c} = 1$ if patent i is among the latest –or most recent– 10% of patents in country c and technology p reported in our sample. In that sense, β_1 is an estimator of the probability that a "pioneering" patent (within any given country-technology pair) is linked to at least one experienced migrant inventor. In our estimation we exclude the 10th decile, due to perfect multicollinearity of the regressors. By doing this, our estimators β_1 to β_9 are always relative to the 10th decile (and therefore, $\beta_{10} = 0$). Finally, note that our specification includes a fixed effect for every country-year and every technology-year. This controls for every characteristic, observable and unobservable, as well as variant and invariant, for the country and technology dimensions. This would include, for instance, controlling for a demand shock for particular technologies in a given year or a spike in R&D investment in a given country and year; two things that could be correlated with the flow of migrant inventors.[25]

Our results are summarized in Figure 4 which plots the estimated values for all β_k regressors, with whiskers representing 95% confidence interval (the standard errors are clustered at the country, technology and year level). Note that, as mentioned above, the reference group in the estimation is the 10th decile, which include the latest or most recent 10% patents filed in each country-technology combination.

The graph reveals a striking pattern: patents in the first decile have a probability of around 0.88 percentage points higher of having been invented by at least one experienced migrant inventor than the latest, or most recent, 10% of patents filed in that same country and technology. Since in our sample the unconditional probability of seeing any experienced migrant inventor across all deciles is 0.86%, this implies that it is over two times more likely to see an experienced migrant as an inventor in first decile of patents in any given country-technology pair than in the latest batch.

Note in the figure that the probability (e.g., point estimate) keeps dropping and even becomes statistically insignificant (based on 95% confidence intervals) after the 4th decile. Interestingly enough, the fact that migrant inventors are essentially absent in patents that correspond to the last few deciles suggests that the ability to innovate in that technology, as time goes by, becomes fully embedded in local inventors.

This micro-evidence is consistent with our overall findings: migrant inventors play an important role in early stages of innovation for any given technology in any given country. Naturally, the deeper understanding of this processes is an essential part of our future research agenda.

[Figure 4 about here.]

[25] A more strict specification that includes country-technology-year fixed effects results in qualitatively similar results, though with a significant loss in precision in the estimation given the lack of sufficient within variation.

6 Conclusions

In this paper we study and provide robust econometric evidence of the role of immigrants inventors in shapinginnovation dynamics in their receiving countries. In particular, our analysis shows that –controlling for other means of exchange such as trade and FDI– countries receiving immigrant inventors from other nations that specialize in patenting in technology p are more likely to have an important increase in patent applications in that same technology.

Our estimates imply that a twofold increase in the number of inventor immigrants can explain an increase of 25 to 50 percent in the likelihood of gaining technological advantage in the same technology that the inventors' home countries are specialized in. In our sample this number can be as low as 25 inventors for the average country, with a standard deviation of about 135. Our econometric analysis includes the use of instrumental variables as well as a number of falsification tests to rule out our results being driven by spurious correlations or other alternative factors not accounted for.

Our findings using patent-level data are strong supporting evidence of our overall results in this paper: the likelihood of seeing an experienced migrant inventor patenting in very early innovation stages of technologies in any given country (e.g., the first few patents ever in that country-technology pair) is over two times larger than on a later stage. Furthermore, our findings support the idea that the ability to keep innovating in a given technology, as time goes by, becomes fully embedded in local inventors and is less reliant on experienced migrant inventors. We interpret this as evidence of knowledge diffusion.

This paper fills a gap in the literature that explores some of our previous work on the role migrants play in facilitating the transfer of knowledge across borders (Bahar and Rapoport, 2018; Choudhury, 2016; Choudhury and Kim, 2018). Specifically, it explores a particular channel through which inventor migrants –a small and very particular subset of high-skilled migrants– can heavily influence innovation dynamics in their receiving countries.

By providing robust results of how migration affects transfer of specific technologies across borders, from the home country of the migrant to the host country,using a large number of countries, studied over two decade-long periods, our study contributes to the literature on migrants and innovation (e.g., Kerr, 2008; Agrawal et al., 2008; Hunt and Gauthier-Loiselle, 2010; Kerr and Lincoln, 2010; Freeman and Huang, 2015; Ganguli, 2015; Bosetti et al., 2015; Choudhury, 2016; Akcigit et al., 2017; Breschi et al., 2017; Bernstein et al., 2018; Miguélez, 2018; Choudhury and Kim, 2018; Doran and Yoon, 2019). More broadly, our findings indicate that migrant inventors can play an important role in shaping the patent production function in their host countries. Arguably, these dynamics driven by migrant inventors play an important role in improving other economic outcomes at large that follow patenting and innovation, such as productivity and, ultimately, economic growth. Hence, this study is another piece of evidence that the overall medium to long-term economic gains from migration are large and persistent over time.

References

Agrawal, Ajay, Devesh Kapur, and John McHale. "How do spatial and social proximity influence knowledge flows? Evidence from patent data." *Journal of Urban Economics* 64, 2: (2008) 258–269.

Akcigit, Ufuk, John Grigsby, and Tom Nicholas. "Immigration and the Rise of American Ingenuity." *American Economic Review* 107, 5: (2017) 327–331.

Almeida, Paul, Anupama Phene, and Sali Li. "The Influence of Ethnic Community Knowledge on Indian Inventor Innovativeness." *Organization Science* , May 2019: (2014) 141223041331,004.

Bahar, Dany, Ricardo Hausmann, and Cesar A. Hidalgo. "Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion?" *Journal of International Economics* 92, 1: (2014) 111–123.

Bahar, Dany, and Hillel Rapoport. "Migration, Knowledge Diffusion and the Comparative Advantage of Nations." *The Economic Journal* 128, 612: (2018) F273–F305.

Bahar, Dany, Samuel Rosenow, Ernesto Stein, and Rodrigo Wagner. "Export take-offs and acceleration: Unpacking cross-sector linkages in the evolution of comparative advantage." *World Development* 117: (2019) 48–60.

Balassa, B. "Trade Liberalisation and Revealed Comparative Advantage." *The Manchester School* 33, 2: (1965) 99–123.

Bernstein, Shai, Rebecca Diamond, Timothy James McQuade, and Beatriz Pousada. "The Contribution of High-Skilled Immigrants to Innovation in the United States." *Stanford University Graduate School of Business Working Paper* .

Berry, Heather. "The Influence of Multiple Knowledge Networks on Innovation in Foreign Operations." *Organization Science* 29, 5: (2018) 855–872.

Bosetti, Valentina, Cristina Cattaneo, and Elena Verdolini. "Migration of skilled workers and innovation: A European Perspective." *Journal of International Economics* 96, 2: (2015) 311–322. `http://dx.doi.org/10.1016/j.jinteco.2015.04.002`.

Breschi, Stefano, Francesco Lissoni, and Franco Malerba. "Knowledge-relatedness in firm technological diversification." *Research Policy* 32, 1: (2003) 69–87.

Breschi, Stefano, Francesco Lissoni, and Ernest Miguelez. "Foreign-origin inventors in the USA: Testing for diaspora and brain gain effects." *Journal of Economic Geography* 17, 5: (2017) 1009–1038.

Choudhury, Prithwiraj. "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, 3: (2016) 585–610.

Choudhury, Prithwiraj, and Do Yoon Kim. "The Ethnic Migrant Inventor Effect: Codification And Recombination of Knowledge Across Borders.", 2018.

Cipolla, Carlo M. *Before the Industrial Revolution: European Society and Economy 1000-1700*. London: Methuen, 1976, 1st edition.

Criscuolo, Paola. "The 'home advantage' effect and patent families. A comparison of OECD triadic patents, the USPTO and the EPO." *Scientometrics* 66, 1: (2005) 23–41.

De Noni, Ivan, Luigi Orsi, and Fiorenza Belussi. "The role of collaborative networks in supporting the innovation performances of lagging-behind European regions." *Research Policy* 47, 1: (2018) 1–13. https://doi.org/10.1016/j.respol.2017.09.006.

Doran, Kirk, and Chungeun Yoon. "Immigration and Invention: Evidence from the Quota Acts." *mimeo* .

Foley, C Fritz, and William R Kerr. "Ethnic Innovation and U.S. Multinational Firm Activity." *Management Science* 59, 7: (2013) 1529–1544.

Frankel, Jeffrey, and David Romer. "Does Trade Cause Growth?" *American Economic Review* 89, 3: (1999) 379–399.

Freeman, Richard B., and Wei Huang. "Collaborating with People Like Me: Ethnic Co-Authorship within the U.S." *Journal of Labor Economics* 33, S1: (2015) S289–S318.

Ganguli, Ina. "Immigration and Ideas: What Did Russian Scientists "Bring" to the United States?" *Journal of Labor Economics* 33, S1: (2015) S257–S288.

Glismann, Hans H., and Ernst-Jürgen Horn. "Comparative Invention Performance of Major Industrial Countries: Patterns and Explanations." *Management Science* 34, 10: (1988) 1169–1187.

Glitz, Albrecht, and Erik Meyersson. "Industrial Espionage and Productivity." *IZA Discussion Paper Series* , 10816: (2017) 1–50.

Hausmann, Ricardo, César A Hidalgo, Sebastián Bustos, Michele Coscia, Alexander Simoes, and Muhammed A. Yildirim. *The Atlas of Economic Complexity: Mapping Paths to Prosperity.* Cambridge, MA: MIT Press, 2014.

Henderson, Rebecca, Adam Jaffe, Manuel Trajtenberg, Peter Thompson, and Melanie Fox-Kean. "Patent citations and the geography of knowledge spillovers: A reassessment: Comment." *American Economic Review* 95, 1: (2005) 461–466.

Horrace, William C., and Ronald L. Oaxaca. "Results on the bias and inconsistency of ordinary least squares for the linear probability model." *Economics Letters* .

Hunt, Jennifer, and Marjolaine Gauthier-Loiselle. "How Much Does Immigration Boost Innovation?" *American Economic Journal: Macroeconomics* 2, 2: (2010) 31–56.

Jaffe, Adam B. "Technological Opportunity and Spillovers of R & D: Evidence from Firms' Patents, Profits, and Market Value." *The American Economic Review* 76, 5: (1986) 984–1001.

Jaffe, Adam B. "Characterizing the "technological position" of firms, with application to quantifying technological opportunity and research spillovers." *Research Policy* 18, 2: (1989) 87–97.

Keller, Wolfgang. "International Technology Diffusion." *Journal of Economic Literature* XLII, September: (2004) 752–782.

Kerr, Sari Pekkala, and William R. Kerr. "Global Collaborative Patents." *The Economic Journal* 128, 612: (2018) F235–F272.

Kerr, W R. "Ethnic Scientific Communities and International Technology Diffusion." *Review of Economics and Statistics* 90, 3: (2008) 518–537.

Kerr, W.R., and W. Lincoln. "The supply side of innovation: H-1B visa reforms and US ethnic invention." *Journal of Labor Economics* 28, 3.

Lissoni, Francesco. "International migration and innovation diffusion: an eclectic survey." *Regional Studies* 52, 5: (2018) 702–714.

MacKinnon, JG, and Lonnie Magee. "Transforming the Dependent Variable in Regression Models." *International Economic Review* 31, 2: (1990) 315–339.

Miguélez, Ernest. "Inventor diasporas and the internationalization of technology." *World Bank Economic Review* 32, 1: (2018) 41–63.

Miguelez, Ernest, and Carsten Fink. "Measuring the international mobility of inventors: A new database." *The International Mobility of Talent and Innovation: New Evidence and Policy Implications* , 8: (2017) 114–161.

Moser, Petra, Alessandra Voena, and Fabian Waldinger. "German Jewish Émigrés and US Invention." *American Economic Review* 104, 10: (2014) 3222–3255.

OECD. "Revealed technology advantage in selected fields." .

Ozden, C., C. R. Parsons, M. Schiff, and T. L. Walmsley. "Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960-2000." *The World Bank Economic Review* 25, 1: (2011) 12–56.

Singer, Judith D., and John B. Willett. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford: Oxford University Press, 2009.

Singh, Jasjit, and Matt Marx. "Geographic Constraints on Knowledge Spillovers: Political Borders vs. Spatial Proximity." *Management Science* 59, 9: (2013) 2056–2078.

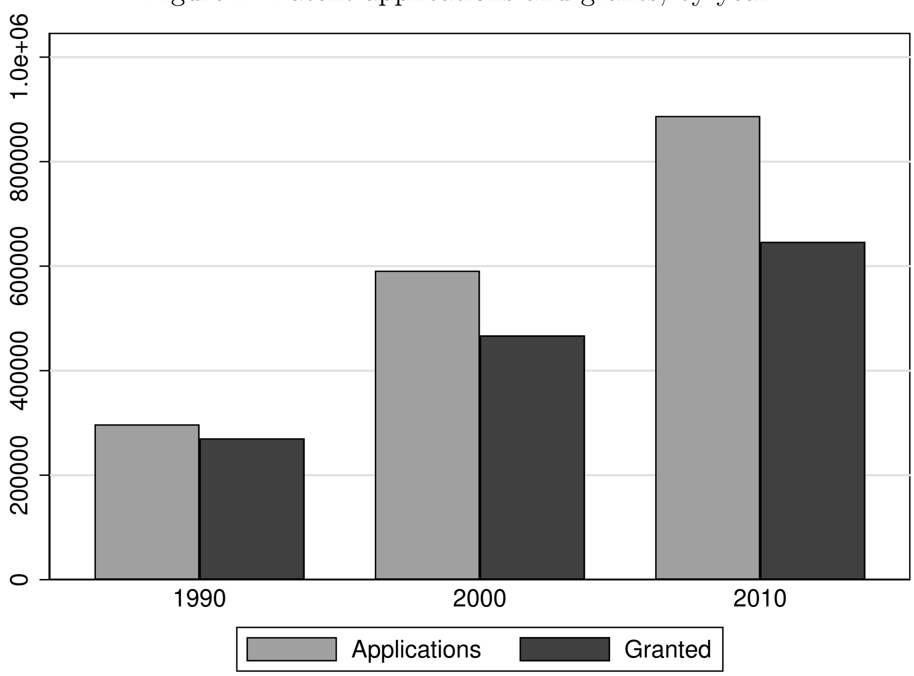
Soete, Luc. "The impact of technological innovation on international trade patterns: The evidence reconsidered." *Research Policy* 16, 2-4: (1987) 101–130.

Stock, JH, and M Yogo. "Testing for Weak Instruments in Linear IV Regression." In *Identification and Inference for Econometric Models*, edited by DWK Andrews, New York: Cambridge University Press, 2005, 80–108.

Thompson, Peter, and Melanie Fox-Kean. "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment." *The American Economic Review* 95, 1: (2005) 450–460.

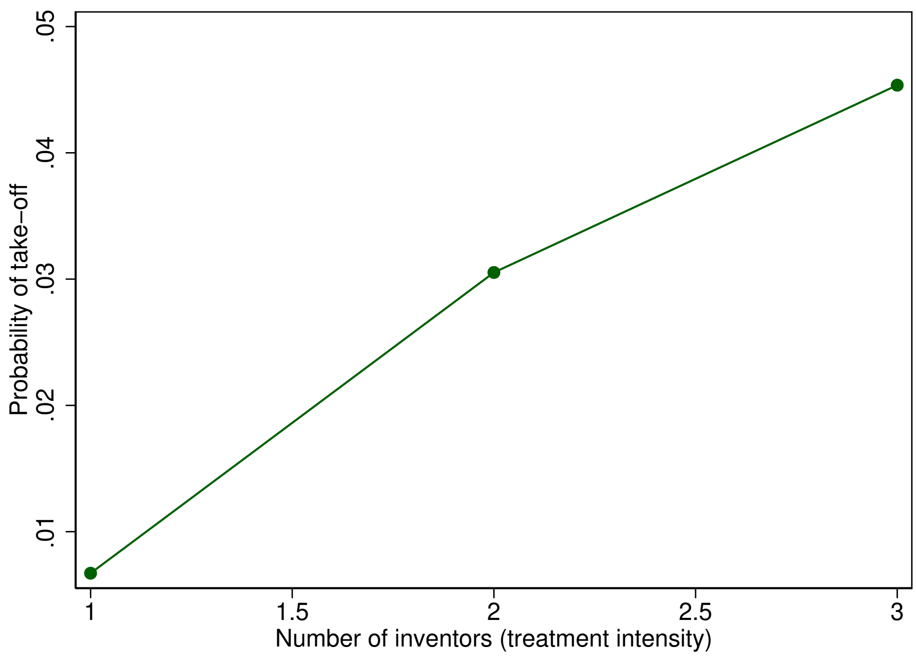
USPTO. "PatentsView Dataset.", 2018. `http://www.patentsview.org/`.

Figure 1: Patent applications and grants, by year



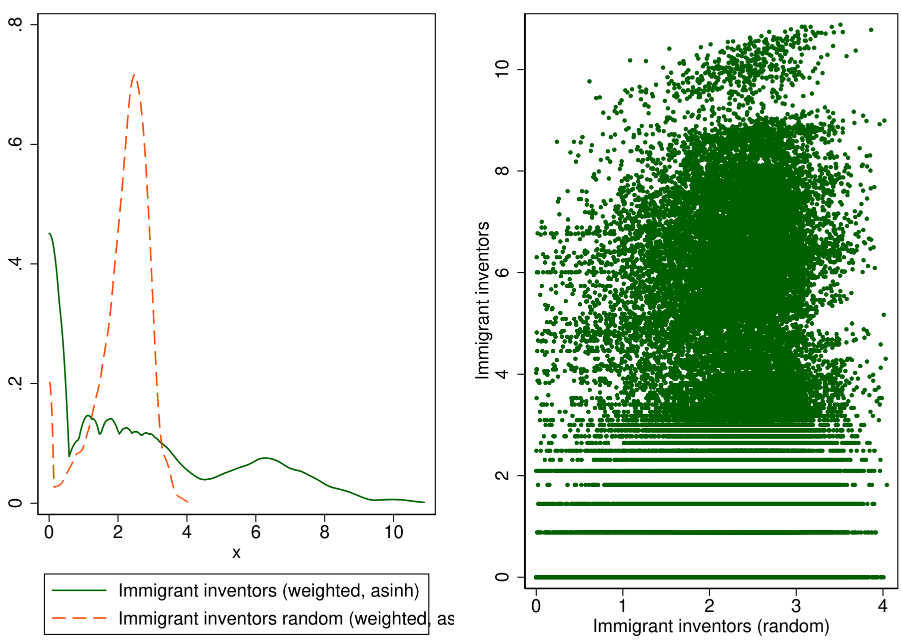
This figures presents the total number of patent applications and granted in years 1990, 2000 and 2010 around the globe, based on the records of the USPTO.

Figure 2: Patent applications take offs, raw data



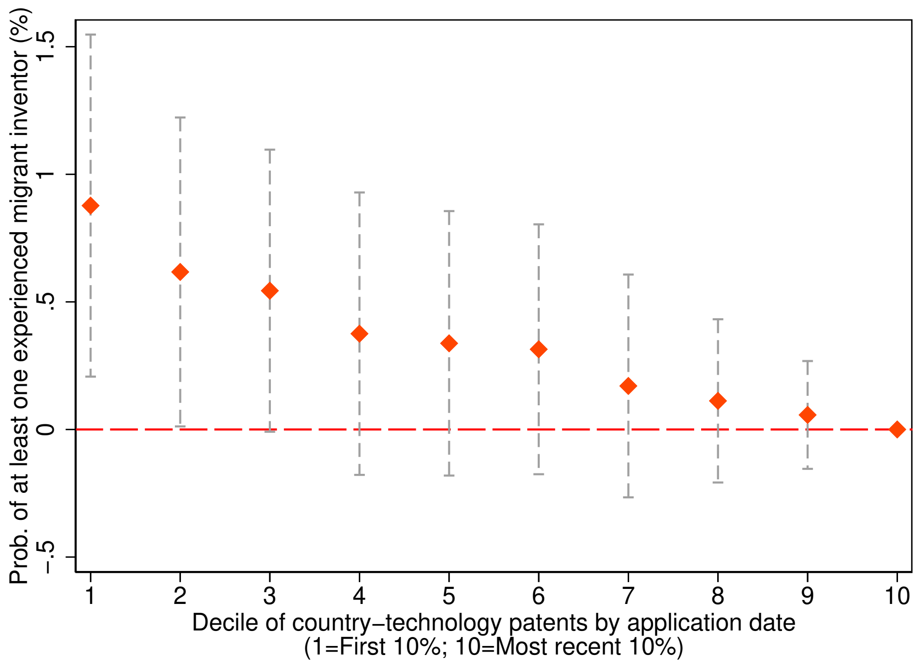
This figures presents the average probability of a patent technology take-off (using patent applications) per quartile of the treatment: the stock of immigrants from countries that excel in the technology under consideration (e.g., produces patents that technology with a RTA above 1). The figure plots average levels of technology take offs, with no controls whatsoever.

Figure 3: Comparing real and random-inventor immigrants



This left panel of the figure plots the kernel distribution of the presents the variable of interest $\sum_{c'} inventors^{im}_{c,c',t} \times R_{c',p,t}$ both using the real numbers of inventor immigrants and a constructed one using random numbers; the right panel plots the values of both of these variables. Both panels use data for year 2000 only, for visualization purposes.

Figure 4: Prob. of experienced migrant inventor in patent, by chronological deciles



This figures presents results for the estimation of Specification (2). It plots the estimated values of $\beta_k$with their corresponding 95% confidence intervals (based on standard errors clustered at the country, technology and year level). The estimator measures the probability of a patent having at least one migrant inventor (i.e., an inventor who had lived in another country before) and with previous experience in patenting in that same technology (i.e., IPC code) in patents divided by deciles of chronological order of appearance within each country and technology pair. The 1st decile includes the first ever 10% of all patents filed by inventors in each country and in each technology, while the 10th decile –our reference group for the estimation– includes the latest or most recent 10% of patents filed within that country and technology. The figure shows that the likelihood of innovations (as measured by patent applications) invented by at least one migrant inventor with prior experience is much higher during the earlier stage of patenting activity in any given country and in any given technology, as compared to later stages.

Table 1: Summary Statistics

Variable	**N**	**Mean**	**sd**	**Min**	**Max**
Panel A - Take-off sample ($RTA_{c,p,t_o} = 0$)					
Take-off technology (patent applications)	105,304	0.022	0.15	0.0	1.0
Immigrant inventors (weighted)	105,304	24.289	135.54	0.0	8,863.0
Emigrant inventors (weighted)	105,304	79.779	411.80	0.0	7,624.0
Total immigrants lagged (thous.,weighted)	105,304	71.412	327.76	0.0	8,181.8
Total emigrants lagged (thous.,weighted)	105,304	173.072	458.87	0.0	8,309.7
Total FDI (billion USD, weighted)	105,304	32.652	217.93	0.0	10,014.4
Total trade (billion USD, weighted)	105,304	107.308	259.55	0.0	4,858.3
Panel B - Growth sample ($patents_{c,p,t_o} > 0$)					
CAGR technology (patent applications)	18,386	0.007	0.08	-0.3	0.8
Baseline patent apps	18,386	16.447	119.13	1.0	9,730.0
Immigrant inventors (weighted)	18,386	709.743	2,116.57	0.0	26,642.0
Emigrant inventors (weighted)	18,386	610.616	1,138.12	0.0	8,229.0
Total immigrants lagged (thous.,weighted)	18,386	556.917	972.48	0.0	8,585.5
Total emigrants lagged (thous.,weighted)	18,386	498.713	744.52	0.0	8,356.7
Total FDI (billion USD, weighted)	18,386	840.103	1,574.50	0.0	14,556.1
Total trade (billion USD, weighted)	18,386	1,114.625	1,386.03	0.0	12,332.5

This table presents the sample summary statistics for the variables used in the paper. The upper panel presents the sample used in the estimations of technology take offs, where we limit the sample to those country-technology observations that have no patents granted in the beginning of the 1990-2000 and 2000-2010 periods). The lower panel presents results used in the estimations of granted patents growth regressions, where we limit our observations to those country-technology pairs with number of patents above zero at the beginning of the 1990-2000 and 2000-2010 periods.

Table 2: Migrant inventors and patent applications take offs and growth

Panel A - Dependent variable: Patent applications take-off (binary)

	OLS			2SLS		
	est1	est2	est3	est4	est5	est6
Immigrant inventors	0.0051		0.0051	0.0114		0.0137
	(0.001)***		(0.001)***	(0.005)**		(0.005)**
Emigrant inventors		0.0018	0.0001		0.0032	-0.0034
		(0.001)***	(0.001)		(0.002)	(0.003)
Total FDI	-0.0001	-0.0001	-0.0001	-0.0003	-0.0001	-0.0002
	(0.000)	(0.000)	(0.000)	(0.000)**	(0.000)	(0.000)
Total Trade	0.0001	0.0003	0.0001	-0.0004	0.0001	-0.0001
	(0.000)	(0.000)**	(0.000)	(0.000)	(0.000)	(0.000)
N	105304	105304	105304	105304	105304	105304
r2	0.09	0.09	0.09	0.09	0.09	0.09
KP F Stat				60.41	39.05	55.44

Panel B - Dependent variable: Patent applications growth (CAGR)

	OLS			2SLS		
	est1	est2	est3	est4	est5	est6
Immigrant inventors	0.0046		0.0036	0.0251		0.0159
	(0.001)***		(0.001)***	(0.010)**		(0.006)***
Emigrant inventors		0.0042	0.0029		0.0174	0.0101
		(0.001)***	(0.001)**		(0.010)*	(0.011)
Total FDI	0.0003	0.0003	0.0003	0.0002	0.0001	0.0001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Total Trade	0.0027	0.0021	0.0016	-0.0041	-0.0044	-0.0060
	(0.002)	(0.001)	(0.001)	(0.003)	(0.004)	(0.005)
Baseline patent apps, log	-0.0178	-0.0176	-0.0178	-0.0186	-0.0174	-0.0181
	(0.002)***	(0.002)***	(0.002)***	(0.002)***	(0.002)***	(0.002)***
Previous Exports Growth	-0.1422	-0.1436	-0.1424	-0.1349	-0.1425	-0.1374
	(0.039)***	(0.039)***	(0.039)***	(0.041)***	(0.039)***	(0.038)***
Zero Exports in t-1	-0.0015	-0.0013	-0.0014	-0.0015	-0.0008	-0.0011
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
N	18349	18349	18349	18349	18349	18349
r2	0.55	0.55	0.55	0.53	0.54	0.54
KP F Stat				8.61	11.02	5.83

This table presents results of the estimation of Specification (1). Columns 1-3 show OLS estimations while columns 4-6 show results for 2SLS regressions. Panel A presents results for take offs in patenting applications (limiting the sample to cases where the initial patent applications for that country-technology pair was zero), while Panel B estimates future CAGR in patent applications for country-technology pairs that already had some patenting recorded in the baseline year. All specifications include country-year and technology-year fixed effects. SE clustered at the country level are presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table 3: Migrant inventors and patent applications, falsification test 1

Dependent variables: Take-off (1-3) and growth (4-6)						
	Take-off			Growth		
	est1	est2	est3	est4	est5	est6
Immigrant inventors	-0.0058		-0.0053	-0.0084		-0.0068
	(0.002)***		(0.002)**	(0.002)***		(0.002)***
Emigrant inventors		-0.0028	-0.0022		-0.0064	-0.0055
		(0.001)**	(0.001)		(0.002)***	(0.002)***
Total FDI	0.0003	0.0004	0.0004	0.0001	0.0001	0.0001
	(0.000)**	(0.000)***	(0.000)**	(0.000)	(0.000)	(0.000)
Total Trade	-0.0004	-0.0008	0.0019	-0.0048	-0.0043	-0.0003
	(0.003)	(0.004)	(0.003)	(0.005)	(0.005)	(0.005)
Baseline patent apps, log				-0.0180	-0.0183	-0.0184
				(0.002)***	(0.002)***	(0.002)***
Previous Exports Growth				-0.1491	-0.1463	-0.1488
				(0.038)***	(0.040)***	(0.038)***
Zero Exports in t-1				-0.0008	-0.0012	-0.0008
				(0.004)	(0.004)	(0.004)
N	105304	105304	105304	18349	18349	18349
r2	0.09	0.09	0.09	0.55	0.55	0.55

This table presents results of the estimation of Specification (1) with a slight modification. Instead of including in the RHS the number of migrant inventors from and in countries that specialize in each technology in the baseline year, we use instead the number of migrant inventors from and in countries with *zero patent applications in each technology at the beginning of the period. Columns 1-3 show OLS estimations while columns 4-6 show results for 2SLS regressions. Panel A presents results for take offs in patenting applications (limiting the sample to cases where the initial patent applications for that country-technology pair was zero), while Panel B estimates future CAGR in patent applications for country-technology pairs that already had some patenting recorded in the baseline year. All specifications include country-year and technology-year fixed effects. SE clustered at the country level are presented in parenthesis.*

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table 4: Migrant inventors and patent applications, falsification test 2

Dependent variables: Take-off (1-3) and growth (4-6)						
	Take-off			Growth		
	est1	est2	est3	est4	est5	est6
Immigrant inventors (random)	-0.0023		-0.0023	0.0042		0.0042
	(0.002)		(0.002)	(0.003)		(0.003)
Emigrant inventors (random)		0.0003	0.0003		0.0008	0.0007
		(0.002)	(0.002)		(0.002)	(0.002)
Total FDI	-0.0000	-0.0000	-0.0000	0.0004	0.0004	0.0004
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Total Trade	0.0006	0.0006	0.0006	0.0041	0.0041	0.0041
	(0.000)***	(0.000)***	(0.000)***	(0.002)**	(0.002)**	(0.002)**
Baseline patent apps, log				-0.0175	-0.0176	-0.0175
				(0.002)***	(0.002)***	(0.002)***
Previous Exports Growth				-0.1439	-0.1439	-0.1438
				(0.039)***	(0.039)***	(0.039)***
Zero Exports in t-1				-0.0015	-0.0015	-0.0015
				(0.004)	(0.004)	(0.004)
N	105304	105304	105304	18349	18349	18349
r2	0.09	0.09	0.09	0.55	0.55	0.55

This table presents results of the estimation of Specification (1) with a slight modification. Instead of including in the RHS the number of migrant inventors from and in countries that specialize in each technology in the baseline year, we use instead the number of migrant inventors from and in countries with *zero patent applications in each technology at the beginning of the period. Columns 1-3 show OLS estimations while columns 4-6 show results for 2SLS regressions. Panel A presents results for take offs in patenting applications (limiting the sample to cases where the initial patent applications for that country-technology pair was zero), while Panel B estimates future CAGR in patent applications for country-technology pairs that already had some patenting recorded in the baseline year. All specifications include country-year and technology-year fixed effects. SE clustered at the country level are presented in parenthesis.*

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table 5: Migrant inventors and patent applications, sub-sample analysis

		OLS		2SLS	
	N	β_{im}	β_{em}	β_{im}	β_{em}
All Observations	105304	0.005***	-0.000	0.013***	-0.004
Non OECD	76695	0.006***	0.000	0.016**	-0.004
OECD	28609	0.003	0.000	0.016**	-0.005
Period 1990-2000	54042	0.000	0.000	0.004	-0.001
Period 2000-2010	51262	0.007***	-0.000	0.015***	-0.005
Chemistry; Metallurgy	13564	0.009***	0.001	0.015	-0.004
Electricity	7571	0.004	0.002	0.010	-0.005
Fixed Constructions	4985	0.000	0.002	0.021	-0.013
Human Necessities	13211	0.009***	-0.000	0.016	-0.010
Mechanical Engineering; Lighting; Heating; Weapons; Blasting	15958	0.004*	-0.002	0.017	-0.006
Performing Operations; Transporting	27529	0.001	-0.003	0.005	-0.002
Physics	12788	0.007**	0.001	0.021	-0.003
Textiles; Paper	6663	0.003*	-0.001	0.016*	-0.011

This table summarizes OLS regressions for different cuts of the sample. The reported beta coefficients are standardized to have zero mean and unit standard deviation, for comparison purposes. The dependent variable in all specifications is technology take off. Significance levels reported based on SE clustered at the country level

Online Appendix for

Migration inventors and the technological advantage of nations

Dany Bahar, Raj Choudhury and Hillel Rapoport

May 28, 2019

A Extended summary statistics

Table A1 presents summary statistics for each of the 95 countries in the sample. The table presents the average probability of take-off for each country (presented in percentage format) based on observations of country-technology pairs with no patent applications in the baseline year for each period. Similarly, the growth rate (also presented in percentage format) is based on observations of country-technology pairs with more than zero patent applications in the baseline period. The * symbol next to a country name implies that the country is part of the OECD prior to the first period studied (e.g., it does not count countries as OECD members if they became so during the 1990s or the 2000s).

[Table A1 about here.]

Table A1: Summary statistics by country

	1990-2000		2000-2010			1990-2000		2000-2010	
Country	**Take off**	**CAGR**	**Take Off**	**CAGR**	**Country**	**Take off**	**CAGR**	**Take Off**	**CAGR**
United Arab Emirates	0.00	-3.99	5.54	-5.88	South Korea	1.39	10.24	5.58	6.77
Argentina	0.32	-3.46	7.03	-3.50	Kuwait	0.00	-6.70	6.45	-3.33
Armenia	0.00	.	2.01	-4.34	Lebanon	0.00	-6.70	1.71	-4.13
Australia *	0.82	1.10	6.92	0.92	Liechtenstein	0.32	-3.72	2.12	-5.34
Austria *	0.51	-0.04	7.90	0.41	Sri Lanka	0.00	-6.09	1.56	-5.86
Belgium *	1.16	0.82	7.84	0.54	Lithuania	0.00	.	2.79	-2.66
Bulgaria	0.00	-6.16	3.13	-2.47	Luxembourg *	0.49	-4.64	5.53	-3.67
Bosnia and Herzegovina	0.00	.	1.08	-6.70	Latvia	0.00	.	2.32	0.84
Belarus	0.15	.	2.02	-2.96	Morocco	0.00	-6.70	1.10	-5.58
Bermuda	0.00	-6.70	0.47	-7.11	Monaco	0.16	-4.69	1.60	-5.78
Brazil	0.84	-1.76	10.36	0.62	Moldova	0.00	-6.70	1.69	-6.70
Canada *	0.55	3.17	10.14	0.94	Mexico	0.85	-3.15	10.17	-2.80
Switzerland *	1.30	0.00	2.16	0.21	Macedonia, FYR	0.00	.	0.00	-6.70
Chile	0.31	-5.68	6.55	-2.70	Malta	0.00	0.00	1.54	-7.93
China	2.16	2.36	15.42	15.21	Mongolia	0.00	-6.70	0.31	.
Colombia	0.00	-3.15	4.91	-3.42	Malaysia	0.47	0.30	9.79	-0.13
Costa Rica	0.00	-7.31	1.40	-4.29	Nigeria	0.00	-6.70	0.15	-2.04
Cuba	0.31	.	0.31	-3.44	Netherlands *	1.29	0.57	8.37	1.08
Cayman Islands	0.00	-6.70	2.32	-5.02	Norway *	1.14	-0.94	7.18	-1.12
Cyprus	0.00	-6.70	3.56	-3.92	New Zealand *	0.52	-1.44	6.02	-1.23
Czech Republic	0.48	-4.77	8.96	-0.94	Pakistan	0.00	-6.70	3.10	-6.70
Germany *	1.02	3.44	1.25	0.40	Panama	0.00	-6.70	0.93	-6.70
Denmark *	1.30	0.06	6.07	-0.59	Peru	0.00	-7.59	2.79	-6.70
Algeria	0.00	-6.70	1.08	-6.70	Philippines	0.79	-5.35	5.56	-4.13
Ecuador	0.00	-6.70	0.62	-6.70	Poland	0.16	-3.75	10.07	-1.36
Egypt	0.00	-6.70	5.38	-4.09	North Korea	0.00	-6.70	0.46	-6.70
Spain *	1.76	-0.21	10.44	1.37	Portugal *	0.47	-4.36	6.48	-1.29
Estonia	0.00	.	5.15	-0.62	Russian Federation	2.39	-2.36	8.11	-1.76
Finland *	1.89	1.24	5.52	-1.37	Saudi Arabia	0.31	-1.99	12.97	3.97
France *	0.00	0.77	5.96	0.81	Singapore	1.91	4.87	7.64	0.15
United Kingdom *	0.00	1.51	3.40	0.63	El Salvador	0.00	.	0.00	-6.70
Georgia	0.00	.	1.54	-6.70	Slovakia	0.77	.	4.18	-4.48
Greece *	0.31	-5.60	6.08	-1.61	Slovenia	0.92	.	3.91	-2.37
Guatemala	0.00	-6.70	0.46	-7.93	Sweden *	1.35	2.73	4.74	-1.63
Hong Kong	1.36	1.60	7.35	-0.95	Seychelles	0.00	-6.70	0.77	.
Croatia	0.61	.	4.11	-3.53	Thailand	0.47	-1.90	6.81	-4.04
Hungary	0.00	-4.43	7.00	-1.97	Trinidad and Tobago	0.00	-6.70	2.48	-7.23
Indonesia	0.00	-4.02	3.50	-3.80	Tunisia	0.00	-6.70	1.55	-2.23
India	1.15	4.39	14.42	5.68	Turkey *	0.00	-5.21	8.32	-1.16
Ireland *	0.35	-0.74	6.69	0.18	Ukraine	0.77	.	5.52	-3.63
Iran	0.00	-6.70	8.83	-4.14	Uruguay	0.00	-6.70	1.39	-6.89
Iceland *	0.77	-4.46	3.06	-3.39	United States *	0.00	3.05	3.03	-0.04
Israel	1.62	3.15	3.10	0.72	Uzbekistan	0.00	.	0.15	-6.70
Italy *	0.79	0.98	7.01	0.21	Venezuela	0.00	-3.86	1.17	-5.12
Jamaica	0.00	-6.70	0.93	-6.70	South Africa	0.00	-3.26	5.27	-3.31
Jordan	0.00	-6.70	2.15	-6.70	Zimbabwe	0.00	.	0.15	-6.70
Japan *	0.00	2.72	4.48	0.37		.	.	.	.
Kazakhstan	0.00	.	1.24	-6.70		.	.	.	.
Kenya	0.00	-4.46	0.78	-3.26		.	.	.	.

The table presents values for the probability of take-off and the growth rate (CAGR), both in percentage terms, averaged across all technology subclasses for each country in each period in our sample. Consistently with our analysis, the table presents the average probability of take-off for each country based on observations of country-technology pairs with no patent applications in the baseline year for each period. Similarly, the growth rate is based on observations of country-technology pairs with more than zero patent applications in the baseline period. The * symbol indicates whether the country is an OECD member (based on the 1980 cutoff, prior to the first period of analysis in our exercise).

B Explanatory power of bilateral migration on future bilateral inventor stocks

As noted in the paper, one plausible concern of our identification strategy of instrumenting current stocks of migrant inventors with total migration stocks lagged by 20 years is that the first-stage correlation of the 2SLS estimation is artificially being driven up by the weighting scheme we use in our baseline specification. However, this is not the case, as we show in the following exercise.

We estimate the following specification using bilateral stocks of migration and of inventor migrants using data for years 1990 and 2000:

$$inventors_{c,c',t} = \beta_M migrants_{c,c',t-20} + \gamma_c + \gamma_{c'} + \theta_t + \varepsilon_{c,c',t}$$

where c and c' are countries, and t is year; $inventors_{c,c',t}$ is the stock of migrant inventors from country c' in country c at time t and $migrants_{c,c',t-20}$ is the stock of total migrants from country c' in country c at time t-20 (that is, for years 1970 and 1980). Both terms are transformed using the inverse hyperbolic sine, and thus β_M can be interpreted as an elasticity. γ_c and $\gamma_{c'}$ are receiving and sending country fixed effects, respectively; whereas θ_t represents year fixed effects. The last term represents the error.

The estimation for this specification is presented in Table B1. The table reports the estimation of β_M using three different estimations which vary on the inclusion of fixed effects, as well as standard errors in parenthesis and t-stats below them for the purpose of this exercise. The first estimation (Column 1) includes only year fixed effects; Column 2 includes year fixed effects as well as sending country and receiving country fixed effects; Column 3 includes sending country and year (combined) fixed effects as well as receiving country-by-year fixed effects. In all columns we see that there is a positive elasticity and perhaps more importantly, a strong explanatory power with t-stats of above 4.5. Looking at columns 2 and 3, which include the country-dimension fixed effects, our point estimates indicate that, on average, an immigrant stock that is 10% larger explains 0.3% larger inventors migrant stock from the same countries 20 years later.

[Table B1 about here.]

Table B1: Explanatory power of lagged migration on inventor migrants

Dependent variable: Inventors immigrant stock			
	(1)	(2)	(3)
Immigrant Stock (t-20)	0.0718	0.0339	0.0328
	(0.015)***	(0.007)***	(0.006)***
	4.658	5.168	5.152
N	44104	44104	44104
Adj. R2	0.16	0.34	0.41
θ_t	Y	Y	-
$\gamma_c, \gamma_{c'}$	N	Y	-
$\gamma_c \times \theta_t, \gamma_{c'} \times \theta_t$	-	-	Y

This table estimates the elasticity of the stock of migrants inventors to the stock of total migrants twenty years before, using bilateral figures. Column 1 includes only year fixed effects; column 2 includes receiving and sending country fixed effects, as well as year fixed effects; column 3 includes receiving country and year (combined) as well as sending country and year (combined) fixed effects. The estimation uses years 1990 and 2000 as baseline years. Standard errors clustered at the receiving country and sending country level and are presented in parenthesis. t-stats are presented below the standard errors

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

C Alternative methods for dependent binary variables

As noted, our dependent variable measuring technology take offs is a binary variable. Hence, our OLS estimation represents a linear probability model (LPM). We believe that given the computational difficulties posed by estimating non-linear models using high-dimensional fixed effects, the choice of a LPM is a reasonable one. However, for robustness purposes, we re-estimate our main specification using the complementary log-log estimator. This particular maximum likelihood estimator, used to estimate models where the dependent variable responds to a binary distribution, is a more advisable option than logit or probit if the probability of take off is small, as in our case (Singer and Willett, 2009). For computational reasons, to limit the number of simultaneous fixed effects, we estimate this clog-log model for the 2000 to 2010 decade only, focusing only on take offs, naturally. Table C1 present the results, which are robust to the ones presented in the main body of the paper.

[Table C1 about here.]

Alternatively, we also consider the approach of Horrace and Oaxaca (2006) for binary data. The approach suggests, after the first estimation of a linear probability model, dropping from the sample the observations for which the predicted value falls out of the unit interval, and use this sub-sample to re-estimate the linear probability model. Horrace and Oaxaca (2006) show this approach may reduce the potential biases of the linear probability models. Table C2 reports the results using this method. In fact, only very few observations in our sample have a predicted value outside of the 0 to 1 range (less than half a percent), so that the number of observations in this estimation is very close as those in Panel A of Table 2, therefore arriving to almost identical results. This suggests the potential bias from linear probability models is minimal, if at all.

[Table C2 about here.]

Table C1: Migrant inventors and patent applications take offs, c-log-log estimation

Dependent variable: Patent applications take-off (binary)			
	est1	est2	est3
Take-off USPTO			
Immigrant inventors	0.0915		0.1030
	(0.036)**		(0.036)***
Emigrant inventors		-0.0151	-0.0494
		(0.044)	(0.046)
Total FDI	0.0055	0.0065	0.0066
	(0.010)	(0.010)	(0.010)
Total Trade	0.0208	0.0443	0.0314
	(0.023)	(0.023)*	(0.024)
N	41221	41221	41221
r2_p			

This table presents results of the estimation of Specification (1), focusing on take offs (a binary dependent variable), using a (maximum likelihood) complementary log-log estimator. All specifications include country and technology fixed effects. SE clustered at the country level are presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table C2: Migrant inventors and patent applications take offs, HO (2006)

Dependent variable: Patent applications take-off (binary)			
	est1	est2	est3
Immigrant inventors	0.0051		0.0051
	(0.001)***		(0.001)***
Emigrant inventors		0.0018	0.0001
		(0.001)***	(0.001)
Total FDI	-0.0001	-0.0001	-0.0001
	(0.000)	(0.000)	(0.000)
Total Trade	0.0002	0.0003	0.0001
	(0.000)	(0.000)**	(0.000)
N	104792	104792	104776
r2	0.09	0.09	0.09

This table presents results of the estimation of Specification (1), focusing on take offs (a binary dependent variable), using the methodology suggested by Horrace and Oaxaca (2016). All specifications include country and technology fixed effects. SE clustered at the country level are presented in parenthesis.

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

D Alternative dependent variables

We extend our exercise, for robustness purposes, to include alternative take-off and growth measures to used as our dependent variable.

Table D1 replicates the estimation shown in Panel A of Table 2, this time using a modified version of take-offs that does not rely on RTA thresholds. The measure uses a much simpler way to account for new technology subclasses in a given country-technology pair based on zero or larger than zero patent applications. In particular, in this case the dependent variable is measured as:

$$Y_{c,p,t \to T} = 1 \text{ if } patents_{c,p,t} = 0 \text{ and } patents_{c,p,T} \geq 1$$

where $T - t = 10$.

[Table D1 about here.]

The results in Table D1 are qualitatively consistent with those in Panel A of Table 2, with larger point estimates, as is expected. Note that while the results hold, we are unable to conclude anything about gaining comparative advantage in certain technology subclasses due to inventor migrants based on them, as opposed to the RTA-based measures.

We also replicate our main specification using an alternative measure of growth, that includes those observations that start-off with zero patents. This measure is the symmetric percentage change, and is defined as:

$$Y_{c,p,t \to T} = \frac{patents_{c,p,T} - patents_{c,p,t}}{0.5 * (patents_{c,p,T} + patents_{c,p,t})} \times \frac{1}{T - t}$$

where $T - t = 10$. In a sense, this growth measure allows us to include all observations, as opposed to CAGR, since we can include those country-technology pairs for which $patents_{c,p,t} = 0$ as their symmetric percentage change (SPC) is defined. This measure also presents a symmetric growth rate, so that a change from $patents_{c,p,t} = x_1$ to $patents_{c,p,T} = x_2$ represents the same percentage change (in absolute value) than from $patents_{c,p,T} = x_2$ to $patents_{c,p,t} = x_1$. Table D2 presents the results using all the sample (i.e., including those country-technology pairs that start-off with zero patent applications).

[Table D2 about here.]

The OLS results of Table D2 show that using this alternative dependent variable, we find consistent results with our main estimations: immigrant inventors explain faster growth rates in patent applications in the same technologies where the countries they migrated from specialize in. The 2SLS results are somewhat weaker in terms of statistical significance, but do the point estimates do hold particularly for the sample of country-technology pairs that start-off with no patent applications whatsoever (i.e., the same sample used in Panel A of Table 2.

Table D1: Migrant inventors and patent applications new technology subclass

Dependent variable: $patents_{c,p,t+10} > 0 \vert patents_{c,p,t} = 0$ **(binary)**						
	OLS			2SLS		
	est1	est2	est3	est4	est5	est6
Immigrants Inventors	0.0352		0.0332	0.0375		0.0515
	(0.004)***		(0.004)***	(0.015)**		(0.020)***
Emigrants Inventors		0.0159	0.0052		0.0056	-0.0157
		(0.003)***	(0.002)***		(0.006)	(0.008)*
Total FDI	0.0001	0.0003	-0.0001	0.0000	0.0007	0.0004
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)*	(0.000)
Total Trade	0.0023	0.0029	0.0018	0.0021	0.0042	0.0030
	(0.001)***	(0.001)***	(0.001)**	(0.001)	(0.001)***	(0.001)**
N	115040	115040	115040	115040	115040	115040
r2	0.26	0.25	0.26	0.26	0.25	0.26
KP F Stat				52.00	121.74	15.55

This table presents results of the estimation of Specification (1). Columns 1-3 show OLS estimations while columns 4-6 show results for 2SLS regressions. The table reports estimations for an alternative measure for take offs in patenting applications. This measure takes the value 1 if the number of patent applications is above 0 in country-technology pairs that had none country-technology applications a decade before (the sample sample is limited to cases where the initial patent applications for that country-technology pair was zero). All specifications include country-year and technology-year fixed effects. SE clustered at the country level are presented in parenthesis.

$^{*}p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$

Table D2: Migrant inventors and patent applications growth (SPC)

Dependent variable: Patents application growth (SPC), yearly average				
	OLS		2SLS	
	All	$patents_{c,p,t_o} > 0$	All	$patents_{c,p,t_o} > 0$
Immigrant inventors	0.0048	0.0052	0.0009	0.0035
	(0.001)***	(0.001)***	(0.003)	(0.003)
Emigrant inventors	0.0007	0.0010	0.0000	-0.0022
	(0.000)*	(0.000)**	(0.002)	(0.001)
Total FDI	0.0000	0.0000	0.0001	0.0002
	(0.000)	(0.000)	(0.000)	(0.000)**
Total Trade	0.0006	0.0004	0.0011	0.0010
	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Baseline patent apps, log	0.0066	0.0000	0.0072	0.0000
	(0.002)***	(.)	(0.002)***	(.)
Previous Exports Log-Growth	-0.6118	0.0000	-0.6112	-0.1865
	(0.026)***	(.)	(0.026)***	(0.013)***
Zero Exports in t-1	0.0813	-0.0361	0.0801	0.0000
	(0.005)***	(0.002)***	(0.005)***	(.)
N	123690	105304	123690	105304
r2	0.30	0.27	0.30	0.27
KP F Stat			56.60	55.72

This table presents results of the estimation of Specification (1). Columns 1-2 show OLS estimations while columns 3-4 show results for 2SLS regressions. The table reports estimations for an alternative measure for decade-long growth in patenting applications: the symmetric percentage change. This measure of growth is defined for country-technology pairs that start-off with zero patent applications in the baseline year. Columns 1 and 3 use all sample, while columns 2 and 4 limits the sample to those observations with initial value 0 in their patent applications count, similarly to Panel A of Table 2. All specifications include country-year and technology-year fixed effects. SE clustered at the country level are presented in parenthesis.

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

E Excluding controls for trade and FDI

Table E1 presents results of our main specification excluding controls for FDI and trade, as some might regard these as "bad controls", as the presence of migrant inventors could also explain an increase flows in trade and FDI. Our main results are robust to this exercise, and the point estimates remain fairly similar.

[Table E1 about here.]

Table E1: Migrant inventors and patent applications take offs and growth, excl. controls

Panel A - Dependent variable: Patent applications take-off (binary)						
	OLS			2SLS		
	est1	est2	est3	est4	est5	est6
Immigrant inventors	0.0050		0.0050	0.0088		0.0132
	(0.001)***		(0.001)***	(0.003)***		(0.005)***
Emigrant inventors		0.0020	-0.0000		0.0039	-0.0041
		(0.001)***	(0.001)		(0.002)**	(0.003)
N	105304	105304	105304	105304	105304	105304
r2	0.09	0.09	0.09	0.09	0.09	0.09
KP F Stat				58.46	61.18	52.05
Panel B - Dependent variable: Patent applications growth (CAGR)						
	OLS			2SLS		
	est1	est2	est3	est4	est5	est6
Immigrant inventors	0.0060		0.0040	0.0204		0.0126
	(0.001)***		(0.001)***	(0.007)***		(0.006)**
Emigrant inventors		0.0056	0.0039		0.0135	0.0060
		(0.001)***	(0.002)**		(0.005)**	(0.008)
Baseline patent apps, log	-0.0179	-0.0176	-0.0178	-0.0184	-0.0175	-0.0181
	(0.002)***	(0.002)***	(0.002)***	(0.002)***	(0.002)***	(0.002)***
Previous Exports Growth	-0.1423	-0.1438	-0.1424	-0.1357	-0.1420	-0.1379
	(0.039)***	(0.039)***	(0.039)***	(0.040)***	(0.039)***	(0.038)***
Zero Exports in t-1	-0.0016	-0.0014	-0.0014	-0.0015	-0.0009	-0.0012
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
N	18349	18349	18349	18349	18349	18349
r2	0.55	0.55	0.55	0.54	0.54	0.54
KP F Stat				18.94	28.86	6.72

This table presents results of the estimation of Specification (1), excluding controls for trade and FDI. Columns 1-3 show OLS estimations while columns 4-6 show results for 2SLS regressions. Panel A presents results for take offs in patenting applications (limiting the sample to cases where the initial patent applications for that country-technology pair was zero), while Panel B estimates future CAGR in patent applications for country-technology pairs that already had some patenting recorded in the baseline year. All specifications include country-year and technology-year fixed effects. SE clustered at the country level are presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

F Excluding China from the sample

Table F1 presents results of our main specification excluding China, as an attempt to rule out a confounding story that attributes our results to intellectual property stealing facilitated by migrant networks. We choose to exclude China because i) it represents an important share of migrants due to its size, and ii) is known to have weaker intellectual property protection. Our main results are robust to this exercise.

[Table F1 about here.]

Table F1: Migrant inventors and patent applications take offs and growth, excl. China

Panel A - Dependent variable: Patent applications take-off (binary)

	OLS			2SLS		
	est1	est2	est3	est4	est5	est6
Immigrant inventors	0.0050		0.0049	0.0112		0.0131
	(0.001)***		(0.001)***	(0.005)**		(0.005)**
Emigrant inventors		0.0016	0.0000		0.0035	-0.0029
		(0.001)***	(0.001)		(0.002)	(0.003)
Total FDI	-0.0001	-0.0001	-0.0001	-0.0003	-0.0002	-0.0002
	(0.000)	(0.000)	(0.000)	(0.000)**	(0.000)	(0.000)
Total Trade	0.0001	0.0003	0.0001	-0.0004	0.0001	-0.0002
	(0.000)	(0.000)**	(0.000)	(0.000)	(0.000)	(0.000)
N	104274	104274	104274	104274	104274	104274
r2	0.09	0.09	0.09	0.09	0.09	0.09
KP F Stat				60.08	38.51	54.57

Panel B - Dependent variable: Patent applications growth (CAGR)

	OLS			2SLS		
	est1	est2	est3	est4	est5	est6
Immigrant inventors	0.0049		0.0040	0.0210		0.0145
	(0.001)***		(0.001)***	(0.009)**		(0.005)***
Emigrant inventors		0.0041	0.0026		0.0138	0.0070
		(0.001)***	(0.001)**		(0.009)	(0.011)
Total FDI	0.0003	0.0003	0.0003	0.0002	0.0001	0.0001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Total Trade	0.0021	0.0017	0.0011	-0.0033	-0.0030	-0.0045
	(0.001)	(0.001)	(0.001)	(0.003)	(0.004)	(0.004)
Baseline patent apps, log	-0.0170	-0.0167	-0.0169	-0.0176	-0.0166	-0.0172
	(0.002)***	(0.002)***	(0.002)***	(0.002)***	(0.002)***	(0.002)***
Previous Exports Growth	-0.1472	-0.1485	-0.1473	-0.1418	-0.1479	-0.1435
	(0.040)***	(0.040)***	(0.040)***	(0.041)***	(0.040)***	(0.039)***
Zero Exports in t-1	-0.0014	-0.0012	-0.0013	-0.0013	-0.0008	-0.0011
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
N	18077	18077	18077	18077	18077	18077
r2	0.53	0.53	0.53	0.52	0.53	0.52
KP F Stat				8.55	11.31	5.62

This table presents results of the estimation of Specification (1), excluding China from our sample. Columns 1-3 show OLS estimations while columns 4-5 show results for 2SLS regressions. Panel A presents results for technology take offs (limiting the sample to cases where the initial patents granted for that country-technology pair was zero), while Panel B estimates future CAGR of granted patents for country-technology pairs that already had some patenting recorded in the initial. All specifications include country-year and technology-year fixed effects. SE clustered at the country level are presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

G Using alternative patents data

Table G1 replicates the results of Tables 2, but using EPO patent applications as the main source of data, following the description in Section 2.3 (as opposed to USPTO patent applications, used in our main results). Our main results hold with these data.

[Table G1 about here.]

In addition, Table G2 replicates our main results using granted patents –as opposed to patent applications– to construct our dependent variables as well as those control variables that are based on patents data. In the table we document that our main results are robust to using granted patents as opposed to patent applications.

We consider these results of high relevance as, even when there could be a significant time gap between the time of patent application and the moment it is granted, the take-offs and accelerations we document in innovation are also present when focusing on patents that have gone through the process imposed by patent agencies that certify the innovation.

[Table G2 about here.]

Table G1: Migrant inventors and patent applications take-offs and growth (EPO)

Panel A - Dependent variable: Patent applications take-off (binary)						
	OLS			2SLS		
	est1	est2	est3	est4	est5	est6
Immigrant inventors	0.0031		0.0032	0.0032		0.0013
	(0.001)***		(0.001)***	(0.002)*		(0.002)
Emigrant inventors		0.0007	-0.0004		0.0037	0.0030
		(0.001)	(0.001)		(0.002)**	(0.002)
Total FDI	0.0001	0.0001	0.0001	0.0001	0.0000	0.0000
	(0.000)	(0.000)*	(0.000)	(0.000)	(0.000)	(0.000)
Total Trade	0.0000	0.0002	0.0000	-0.0000	-0.0002	-0.0002
	(0.000)	(0.000)*	(0.000)	(0.000)	(0.000)	(0.000)
N	105039	105039	105039	105039	105039	105039
r2	0.07	0.07	0.07	0.07	0.07	0.07
KP F Stat				59.62	39.14	55.56
Panel B - Dependent variable: Patent applications growth (CAGR)						
	OLS			2SLS		
	est1	est2	est3	est4	est5	est6
Immigrant inventors	0.0027		0.0022	0.0141		0.0120
	(0.001)**		(0.001)*	(0.007)**		(0.006)*
Emigrant inventors		0.0021	0.0013		0.0086	0.0019
		(0.002)	(0.002)		(0.006)	(0.008)
Total FDI	0.0001	0.0001	0.0001	0.0000	-0.0000	0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Total Trade	-0.0004	-0.0005	-0.0008	-0.0041	-0.0034	-0.0043
	(0.001)	(0.001)	(0.001)	(0.002)**	(0.003)	(0.003)
Baseline patent apps, log	-0.0207	-0.0206	-0.0207	-0.0213	-0.0205	-0.0212
	(0.002)***	(0.002)***	(0.002)***	(0.002)***	(0.002)***	(0.002)***
Previous Exports Growth	-0.1259	-0.1265	-0.1258	-0.1211	-0.1248	-0.1215
	(0.026)***	(0.026)***	(0.026)***	(0.028)***	(0.027)***	(0.027)***
Zero Exports in t-1	-0.0010	-0.0009	-0.0009	-0.0012	-0.0009	-0.0012
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
N	18623	18623	18623	18623	18623	18623
r2	0.46	0.46	0.46	0.46	0.46	0.46
KP F Stat				15.18	23.77	7.33

This table presents results of the estimation of Specification (1), using patent applications reported by the EPO to construct the dependent variable as well patent-relevant control variables. Columns 1-3 show OLS estimations while columns 4-6 show results for 2SLS regressions. Panel A presents results for take offs in patenting applications (limiting the sample to cases where the initial patent applications for that country-technology pair was zero), while Panel B estimates future CAGR in patent applications for country-technology pairs that already had some patenting recorded in the baseline year. All specifications include country-year and technology-year fixed effects. SE clustered at the country level are presented in parenthesis.

$^{*}p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$

Table G2: Migrant inventors and granted patent take offs and growth

Panel A - Dependent variable: Patent applications take-off (binary)

	OLS			2SLS		
	est1	est2	est3	est4	est5	est6
Immigrant inventors	0.0047		0.0044	0.0075		0.0079
	(0.001)***		(0.001)***	(0.004)*		(0.005)*
Emigrant inventors		0.0024	0.0010		0.0032	-0.0006
		(0.001)***	(0.001)		(0.002)	(0.002)
Total FDI	-0.0000	0.0000	-0.0000	-0.0001	-0.0000	-0.0001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Total Trade	0.0001	0.0002	0.0001	-0.0001	0.0001	-0.0000
	(0.000)	(0.000)*	(0.000)	(0.000)	(0.000)	(0.000)
N	106620	106620	106620	106620	106620	106620
r2	0.09	0.09	0.09	0.09	0.09	0.09
KP F Stat				59.93	39.29	55.80

Panel B - Dependent variable: Patent applications growth (CAGR)

	OLS			2SLS		
	est1	est2	est3	est4	est5	est6
Immigrant inventors	0.0026		0.0012	0.0171		0.0055
	(0.001)**		(0.001)	(0.010)*		(0.005)
Emigrant inventors		0.0038	0.0033		0.0137	0.0109
		(0.001)***	(0.002)**		(0.009)	(0.010)
Total FDI	0.0004	0.0004	0.0004	0.0003	0.0002	0.0002
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Total Trade	0.0002	-0.0005	-0.0007	-0.0039	-0.0043	-0.0047
	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)
Baseline patent apps, log	-0.0170	-0.0168	-0.0168	-0.0175	-0.0166	-0.0168
	(0.002)***	(0.002)***	(0.002)***	(0.002)***	(0.002)***	(0.002)***
Previous Exports Growth	-0.1291	-0.1300	-0.1298	-0.1275	-0.1315	-0.1305
	(0.044)***	(0.044)***	(0.044)***	(0.045)***	(0.042)***	(0.042)***
Zero Exports in t-1	-0.0025	-0.0022	-0.0023	-0.0027	-0.0017	-0.0019
	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)
N	17032	17032	17032	17032	17032	17032
r2	0.50	0.50	0.50	0.49	0.49	0.49
KP F Stat				11.42	17.25	5.96

This table presents results of the estimation of Specification (1), using granted patents to construct the dependent variable as well patent-relevant control variables. Columns 1-3 show OLS estimations while columns 4-6 show results for 2SLS regressions. Panel A presents results for take offs in patenting applications (limiting the sample to cases where the initial patent applications for that country-technology pair was zero), while Panel B estimates future CAGR in patent applications for country-technology pairs that already had some patenting recorded in the baseline year. All specifications include country-year and technology-year fixed effects. SE clustered at the country level are presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$