The ties that bind and transform: knowledge remittances, relatedness and the direction of technical change

Valentina Di Iasio*[1,2] and Ernest Miguelez[2,3]

[1]Department of Economics, University of Southampton
[2]GREThA UMR CNRS 5113 - Université de Bordeaux
[3]AQR-IREA - University of Barcelona

Accepted for publication at the *Journal of Economic Geography*, 2021, lbab044,
https://doi.org/10.1093/jeg/lbab044

Abstract

This study investigates whether high-skilled migration in a sample of OECD countries fosters technological diversification in the migrants' countries of origin. We focus on migrant inventors and study their role as vectors of knowledge remittances. Further, we particularly analyze whether migrants spark related or unrelated diversification back home. To account for the uneven distribution of knowledge and migrants within the host countries, we break down the analysis at the metropolitan area level. Our results suggest that migrant inventors have a positive effect on the home countries' technological diversification, particularly for developing countries and technologies with less related activities around - thus fostering unrelated diversification.

Key words: migration, inventors, diversification, technical change
JEL Classification Numbers: O31, O33, F22

*Corresponding author: Department of Economics, University of Southampton. Email: valentinadiiasio@gmail.com

We would like to thank Mathieu Clément, Eric Rougier, Julien Seux, and Georgios Tsiachtsiras for useful remarks. We also received valuable suggestions from participants at the 9th Annual Conference on "Immigration in OECD Countries" (Paris, 2019), the CEMIR Junior Economist Workshop on Migration Research (Munich, 2019), PhD workshop (Insubria, 2019) and the 5th Geography of Innovation conference (Stavanger, 2020), as well as seminars at Beijing Normal University, the University of Bordeaux, the University of Sheffield and the University of Valencia. We also acknowledge financial support from the French National Research Agency (TKC project – reference: ANR-17-CE26-0016)

1 Introduction

Innovation and technical change are well-known drivers of economic growth of countries (Romer, 1994). Yet, technical change relies heavily on the countries' past technological trajectories, which tend to be path-dependent (Dosi, 1997). When countries manage to diversify into different activities, they tend to do it to technologically adjacent domains, as shown by the principle of relatedness (Hidalgo et al., 2007; Petralia et al., 2017; Hidalgo et al., 2018; Kogler et al., 2013; Boschma et al., 2015). However, in order to avoid technological lock-in, they must move into technological paths located far away from their current knowledge base (unrelated diversification) (Saviotti and Frenken, 2008). Unrelated diversification might be more difficult to create and more likely to fail, but if achieved, it can potentially foster structural change (Neffke et al., 2018), making countries less vulnerable to technology shocks and more prone to economic growth in the long run (Pinheiro et al., 2018). This might be especially relevant for developing countries, as they rely on a relatively low number of actual activities from which they can diversify into new technologies (Hidalgo et al., 2018). The question remains, however, on who are the agents able to spark (related and unrelated) technological change.

This paper investigates the relationship between skilled migration (proxied by inventors) in a sample of OECD countries and technological diversification in the migrants' country of origin.[1] Using the framework of the branching literature (Hidalgo et al., 2007; Hausmann et al., 2007; Essletzbichler, 2015; Rigby, 2015; Boschma, 2017), we test the hypothesis that migrant inventors abroad (inventor diasporas) stimulate new patent applications in their countries of origin in technologies in which the destination area is relatively specialized - while the country of origin is not, and therefore foster technical change at home. While this literature has generally focused on the internal factors driving technological diversifica-

[1] We use inventors as a proxy for high-skilled workers. Although we are aware that they are not exactly the same, the former are a critical component of the latter, and a good proxy for knowledge or STEM workers. Even though in parts of the text we refer to high-skilled migration, our empirical analysis is focused on inventors migration only.

tion, external factors have been mostly overlooked (Neffke et al., 2018; Whittle et al., 2020). This includes the potential role of international migrants (Bahar et al., 2020). Further, the literature has generally focused on the process of related diversification, and unrelated diversification has received less attention (Boschma, 2017). Thus, building upon the concept of relatedness (Hidalgo et al., 2018), which refers to the similarity between activities (products, industries, research areas) in terms of scientific knowledge, technical principles, heuristics, and common needs (Petralia et al., 2017), we test whether migration-induced diversification tends to be related or unrelated to the current knowledge base.

Migration, especially of the highly-skilled, is nowadays a widespread phenomenon. The third wave of globalization opened new opportunities for human capital to reallocate, generating an increase in international migration of college educated workers (Kerr et al., 2016). This has given raise to an increasing number of studies showing the influence of high-skilled migration on innovation in host countries (Stephan and Levin, 2001; Chellaraj et al., 2008; Hunt and Gauthier-Loiselle, 2010; Kerr and Lincoln, 2010; Moser et al., 2014; Bosetti et al., 2015; Ganguli, 2015; Akcigit et al., 2017; Choudhury and Kim, 2019). The relationship between high-skilled diasporas and home countries' access to foreign technology - knowledge remittances - has been also studied (Kerr, 2008; Agrawal et al., 2011; Breschi et al., 2017; Bahar et al., 2020; Fackler et al., 2020; Miguelez and Temgoua, 2020).[2] However, the role of diasporas in fostering technical, structural change at home is less known.

To understand the role of inventor diasporas on home country technological diversification, we rely on an original database with information on worldwide patent families (Miguelez et al., 2019; WIPO, 2019). We focus on the five most common destination countries for migrant inventors: United States (US), Germany, Switzerland, United Kingdom, and France. They are also among the most technology advanced countries in the world, and sources of international knowledge spillovers (Coe and Helpman, 1995; Keller, 2004; Coe et al., 2009).

[2] For an exhaustive review on skilled migration and knowledge diffusion, see Lissoni (2018); for a review on the effects of diasporas on home countries' development, see Bahar (2020).

As sending countries of these inventors, we work with a sample of 137 economies (both high-income and developing ones). We classify patent families in 636 technologies according to the first four digits of the International Patent Classification (IPC). As measures of diversification and technical change, we look at the growth of patents per country and technology, as well as entry into new specializations. For that we calculate a Revealed Technological Advantage (RTA) index (Soete, 1987) to measure relative specialization, and the time evolution of this index to look at countries that became specialized in classes in which they were not specialized in the past.

As in Bahar et al. (2020), we build our migration proxy exclusively based on inventors' migration data using the database by Miguelez and Fink (2017). Focusing on inventor migration as captured in patent applications can overcome many of the limitations associated with census-based data. It captures one specific class of high-skilled workers, more homogeneous than the group of tertiary-educated workers as a whole, often behind the creation and diffusion of ideas.

All in all, we introduce three main novelties with respect to the existing literature. First and foremost, we study the capacity of inventor diasporas to foster technological change in their home countries. Differently from Bahar et al. (2020), we account for the fact that geographic areas within host countries tend to specialize in very different technologies and skills (Kogler et al., 2013). Moreover, migrants do not evenly distribute within a country, but tend to agglomerate in highly innovative, urban areas (Kerr, 2010; Verginer and Riccaboni, 2021). Further, they tend to settle where previous co-nationals migrated (Munshi, 2003; Beine et al., 2011), thus leading to within country specialization in specific foreign nationalities. For instance, in the US, the MSA of San José is highly specialized in technologies such as telecommunications, computer technology, or semiconductors, while Detroit specializes in engines, turbines, or mechanical tools. Meanwhile, San José largely welcomes inventors from India, followed by inventors from China and, to a lesser extent, Germany. Detroit is home of mainly German inventors, followed at a distance by Indians and Chinese.

Second, as our most novel contribution, we qualify the direction of technical change by investigating whether inventor diasporas are more prone to foster unrelated technological change - whose development would have been more difficult had they relied upon the actual knowledge base of the country, which potentially may lead to structural change (Neffke et al., 2018).

Finally, we investigate heterogeneous effects based on the level of economic development of sending countries. Externally-driven technological change might be particularly important for developing countries, as they lack the preconditions necessary for diversifying into new technologies (Petralia et al., 2017). While, as argued above, related diversification is not a negative process per se, the risk is to become locked in the development of a certain group of technologies, narrowing down diversification opportunities and complicating the catching-up process with high-income countries (Hidalgo et al., 2007).

To anticipate the results to come, we find a positive and significant coefficient associated to migration, suggesting a positive relationship with technological diversification, in line with the literature earlier mentioned. We also find a negative and significant coefficient for the interaction between migration and relatedness density, supporting the hypothesis that external knowledge flows aid countries to diversify into new, unrelated technologies and break path dependency. Moreover, when analyzing the heterogeneity between high income and developing countries, we find that our core results specifically hold for the latter. This supports the hypothesis that having a diaspora abroad does not necessarily imply a brain drain for developing countries. By bringing new ideas, skilled diasporas may help compensate for the lack of domestic knowledge and foster technological development and diversification. Contrarily, high-income economies seem not to benefit that much from their skilled nationals abroad.

We partially deal with endogeneity adopting different strategies, including an instrumental variables approach. Following Frankel and Romer (1999); Ortega and Peri (2014); Bahar and Rapoport (2018), we use the prediction of a gravity model of migration to build a suitable instrument for our focal explanatory variables. Our results are robust to our IV strategy, the inclusion of control variables, and a large number of fixed effects.

The remaining of the paper is organized as follows: Section 2 surveys the literature on technological diversification, and the contribution of migration to knowledge diffusion and innovation; Section 3 describes the data with some more detail, and explains our methodology and empirical strategy; Section 4 presents the step by step results; finally, Section 5 concludes.

2 Related literature

While the production of new technologies is a widely unquestioned track to growth and development, less is known on the factors moving technological change one way or another. As diversification usually follows a path-dependent process (Dosi, 1997), it is assumed that the actual set of capabilities conditions which new activities will countries be able to develop (Boschma, 2017), in accordance with the concept of relatedness (Hidalgo et al., 2018). Several empirical studies show that the diversification possibilities at the country (Hidalgo et al., 2007; Petralia et al., 2017), region (Boschma, 2017; Rigby, 2015; Balland et al., 2019) and firm (Jaffe, 1986; Breschi et al., 2003) levels are affected by the related capabilities present in the country, region and firm. For instance, at the country level, Hidalgo et al. (2007) show that countries have a higher probability to add to their basket of export products that are related to the ones they already produce/export. An important implication is that developing countries are usually located in the periphery of the product space, with consequently fewer opportunities for diversification. Petralia et al. (2017) confirm the role of relatedness in binding countries' technological diversification patterns, particularly of countries at early stages of development, concluding that developing countries tend to be more exposed to the risk of technological lock-in. Developing countries seem, therefore, the places with more potential to benefit from the introduction of diversification from abroad.

This literature is particularly rich at the regional level. Neffke et al. (2011), looking at the evolution of Swedish regions, show that these tend to enter new industries when related sectors are already present locally, way more than if the new industry is unrelated to the current industrial base. Similar results are found for the US using technologies and patent data by Rigby (2015) and Boschma et al. (2015), among many others.

In general, this literature shows that related diversification in countries and regions reigns, while unrelated changes are more difficult to occur (Pinheiro et al., 2018). Yet, unrelated diversification is also possible, and has been shown to be beneficial for countries and regions - especially in the long run. Saviotti and Frenken (2008) stress the particular role of unrelated export diversification in ensuring long-term economic growth and development, for a sample of countries. Pinheiro et al. (2018) analyze the export diversification paths of countries over the long run, to show that unrelated diversification tends to occur in only 7.2% of the cases. However, countries entering more unrelated products tend to growth faster than those only entering related products, evidencing the importance of export diversity for development, which has been associated to higher resilient economic systems.

Here it is important to appreciate the role of external actors able to break lock-in and path dependency. Bahar et al. (2014) investigate the role of distance on the evolution of comparative advantages in trade, finding that countries are more likely to add in their basket of export products already exported by neighbor countries, even if they have different factors' endowments. These findings confirm that knowledge tends to be localized, therefore contributing to fuel the debate on the importance of human interactions for knowledge diffusion. Neffke et al. (2018) look at emerging economic activities in Swedish regions, and found that newcomer firms are more likely to introduce new, unrelated activities into regions, especially if they arrive reallocated from other regions. They are therefore the agents able to foster structural change in the economy. In a similar vein, Multinational Corporations (MNC) have been regarded to be key agents of structural change in regions (Elekes et al., 2019; Crescenzi et al., 2020). To our knowledge, the role of international skilled migrants has received less attention.

In the literature of the early 1970s, the emigration of high-skilled individuals was widely seen as a potential threat for developing countries, relatively less endowed with human capital and more vulnerable to its loss (Bhagwati and Hamada, 1974; Bhagwati, 1976). Yet, an increasing number of studies in recent years have reported that migrants may create transnational communities keeping connections with their home countries and establishing links with migrants living in other places (McAuliffe and Ruhs, 2017; Saxenian, 2007). Thus, the existence of a high-skilled diaspora exposes their home countries to foreign technological knowledge and may constitute an important resource or, borrowing the expression from Agrawal et al. (2011), a brain bank.

Knowledge remittances may travel through different, non-mutually exclusive, forms. One channel of technology transfer is the transmission of knowledge and skills from high-skilled migrants to their social contacts back home (referred to as *ethnically driven* knowledge flows (Breschi et al., 2017)), on a friendly or contractual basis. Knowledge transfers to home countries may occur also when high-skilled workers decide to return on a permanent or temporary basis, equipped with new skills and social networks (Baruffaldi and Landoni, 2012; Choudhury, 2016).[3]

Kerr (2008), by combining patents with industry-level manufacturing data, shows that the industry output of the sending countries increases as the respective ethnic communities develop knowledge in the US. Breschi et al. (2017) define a brain gain effect when a foreign patent receives a higher number of citations in the home country of the inventor. The authors highlight a positive effect of high-skilled migration on brain gain for all the emerging countries

[3]Sending countries can also benefit from their diasporas abroad through the action of MNCs, by means of multi-establishment, international teams, or through internal mobility of skilled labor (Branstetter et al., 2015; Choudhury and Kim, 2019).

except for India and underline the importance of absorptive capacity in the country of origin.

Kerr and Kerr (2018) scrutinize global collaborative patents, defined as patents where at least one inventor is located within the US and at least one resides in a foreign country, of US public firms. According to the authors, global collaborative patents are more impactful than those where all team is located either in the US or abroad. Moreover, US-based firms employing foreign inventors are more likely to engage in these collaborative patents. In a similar vein, Marino et al. (2019) analyze the citation patterns of global collaborative patents. The authors find that US-based inventors, whose foreign ethnicity matches the foreign region in which the other members of the team are located, act as bridges between the multinationals' headquarters and their home countries, facilitating the access to foreign knowledge for the latter. Miguelez (2018) explores the impact of high-skilled diaspora on cross-country patent collaborations between developed and developing countries, finding a positive and robust effect. Choudhury (2016) investigates the role of return migrant managers on the patent activity of 50 US multinationals' R&D centers based in India. The study finds that returnee migrant managers facilitate greater innovation among their local employees, as they connect them with ideas and resources of the US headquarters.

The studies mentioned so far focus on whether migration allows countries of origin to access foreign knowledge, yet they do not analyze whether these knowledge flows transform the home countries' economies. Moreover, the use of citations as a proxy for knowledge flows has been recently criticized as flawed (Jaffe and De Rassenfosse, 2019; Arora et al., 2018). A recent strand of literature focuses on the impact of migration, as a channel of knowledge diffusion, on the evolution of comparative advantages. Kerr (2018) finds that migrants networks contribute to technology transfers from the US, and that those transfers are sufficiently strong to promote exports from migrants' homelands to other countries. Bahar and Rapoport (2018) examine the impact of migration on the extensive (whether a country starts to export a new product from scratch) and the intensive margin (whether a country increases the exports for a given product) of trade of both sending and receiving countries. A follow-up study by Bahar et al. (2020), using data on patents and migrant inventors, shows a positive and robust impact of inventor migration on their host countries and non-significant results for migrants' home countries patenting.

The last described strand of literature, although providing interesting results on the relationship between migration and diversification, is essentially silent on how external knowledge flows interact with countries' endogenous productive and technological capabilities. It does not speak therefore on the qualitative aspects of diversification (related or unrelated to the current knowledge base). In this vein, our paper analyzes how relatedness and knowledge remittances interplay, and sheds light on whether or not inventors' migration creates social bridges between lagging-behind countries and developed areas where advanced knowledge is present, helping to break the technological path dependency of countries and promoting structural change.

3 Data and methods

3.1 Data and sample construction

To build the dependent variables (growth of number of patents and entry into new technology), we use an original database that gathers information on 34 million of worldwide patent families (Miguelez et al., 2019). The data cover all patent documents worldwide, filed in any patent office - provided that they are available in the European Patent Office's (EPO) Worldwide Patent Statistical Database (PATSTAT). We collapse all patents of the same family to the first filing of a given set of patent documents filed in one or more countries and claiming the same invention. Each set containing one first and, potentially, several subsequent filings is defined as a patent family.[4] Worldwide patents can be further split into internationally-oriented and domestically-oriented ones. Internationally-oriented patent fam-

[4] For a more extensive definition of patent families, see Martinez (2010).

ilies refer to patents filed by applicants seeking patent protection in at least one jurisdiction other than their country of residence. Domestic patent families refer only to filings in a home country. While our analysis is based on the use of both types of patents together, robustness checks in the online appendix repeats all main regressions using internationally-oriented patents only, which we use as an indicator of minimum quality of the patent, allowing us to reduce noise related to the idiosyncrasy of each national patent system. Miguelez et al. (2019) database provides geocoding information of patent documents, based on the inventors' addresses (when possible).[5] We then attribute all geocoded patent data into Metropolitan Statistical Areas in the US and metropolitan regions for the case of European countries.[6] Data are available from 1976 to 2017, and we use the period from 1996 to 2015 to build four non-overlapping 5-year time windows.[7] We then classify the patents in technological classes according to the 4 digits IPC codes and focus the analysis on the classes that appear in all the time windows.

For the migration variables, we use data from Miguelez and Fink (2017), who collect Patent Cooperation Treaty (PCT) applications containing information on inventors' nationality. This has to do with the requirement under the PCT that only nationals or residents of a PCT contracting state can file PCT applications. To verify that applicants meet at least one of the two eligibility criteria, the PCT application form asks for both nationality and residence. A limitation of this data is that we automatically exclude from the sample naturalized inventors. However, they still give a more precise measure of inventors migration than census data, that are generally available only every 10 years, and provide a skills breakdown according to only three schooling levels. The database covers the period 1980-2010. Using the period 1991-2010, we build four 5-years non-overlapping time windows. In the regressions

[5]Geocoded data originally collected from Bergquist et al. (2017); Yin and Motohashi (2018); Morrison et al. (2017); de Rassenfosse et al. (2019), and PatentsView.org, among others.

[6]Metropolitan regions in Europe are defined as NUTS3 regions or a combination of NUTS3 regions which represent all agglomerations of at least 250,000 inhabitants (see https://ec.europa.eu/eurostat/web/metropolitan-regions/background, accessed January 2020).

[7]The 1991-1995 period will be occasionally used to build some explanatory variables.

we introduce this variable with a 1-time window lag, in order to minimize issues of reverse causality.[8] Patent and inventor data from Miguelez and Fink (2017) are not provided at the metropolitan area level. In order to get that, we combine it with the OECD REGPAT database, where PCT patents are available at the NUTS3 and county levels (using inventors' addresses) - that we then group into, respectively, European metro regions and MSAs. We match both datasets using the available application number and the names of the inventors.

As inventors are often associated with more than one technological class (because their patents are, too), we prefer to group them into five technological areas (electrical engineering, instruments, chemistry, process engineering and mechanical engineering) according to the classification of Schmoch (2008). We do this in order to avoid fractionalizing head counts of inventors, or duplicating them across technologies (patents commonly belong to more than one technological class, IPC4, but are unlikely to belong to more than one technological area).[9] We then calculate how many inventors of a given nationality are working in a given technological area in the metropolitan area of destination.[10]

As mentioned in the introduction, we restrict the analysis to the most common destination countries for migrant inventors, that is, the US, Germany, Switzerland, United Kingdom, and France. Figure 1 shows the percentage of migrant inventors hosted by these five countries, showing that the US hosts 54 percent of the total. Focusing on these countries, we take into account the 76 percent of the total inventor migration.

[Figure 1 about here.]

As it is shown in the next section, our main explanatory variables combine information on the RTA index (built using information from all patent families) in destination cities, as well as the uneven settlement of migrants in space. We exploit the metro region desegregation level

[8] The choice of the 5-year time windows to compute our variables is customary in the related literature. Results using slightly different time windows do not alter our results - provided upon request.

[9] Table A19 in online Appendix 15 confirms our main results when we calculate migration at the IPC4 level.

[10] The correspondence between technological classes and technological areas is unique (Schmoch, 2008).

for European countries and MSAs for the US. This is possible since our database geocodes 80 percent of the total patent families at a fine geographical detail (Miguelez et al., 2019). Our final sample consists of 137 countries of origin (24 high-income and 113 developing countries), 636 technological classes, 4 time windows and 447 metropolitan areas of destination.

Table 1 shows the main migration corridors for the US MSAs and European metropolitan areas.[11] For the US the main corridors are from China and India to San Diego, San José, and Boston. In Europe corridors are dominated by intra-European flows, and the main ones are from Germany to Zürich and Basel, from the Netherlands to London and Paris, and from France to London and Lausanne. In Table 2 we remove China and India as possible countries of origin for the US, and other high-income countries for European metropolitan areas as possible origins. For the US the table shows that the main sources of migrant inventors are from Canada and the United Kingdom and the most attractive MSAs remain San Diego, San José, and Boston. For Europe, the main origin countries are India, China, and Russia, and the most attractive metropolitan areas are London and Paris - refer to the online Appendix 2 for a detailed descriptive analysis.

[Table 1 about here.]

[Table 2 about here.]

3.2 Empirical approach and variable construction

In order to explore the role of skilled diaspora on the technological diversification of migrants' home countries, we estimate the following regression, at the country-technology level, that accounts for heterogeneity in destination countries by building our variables of interest (migrants and relative specialization) exploiting information at the metropolitan level:

[11] Cross-country migration corridors are depicted in figure A3 in online Appendix 2.

$$
\begin{aligned}
Y_{c,t,tw} = &\alpha + \beta_1 Migration_{c,t,tw-1} + \beta_2 Rel\_dens_{c,t,tw-1} + \beta_3 Migration_{c,t,tw-1} * \\
&Rel\_dens_{c,t,tw-1} + \beta_4 Controls_{c,t,tw-1} + \gamma_{c,tw} + \delta_{t,tw} + \epsilon_{c,t,tw}
\end{aligned}
\tag{1}
$$

where we denote with c the inventors' home countries, t the technological class (belonging to one single technological area, ta), tw the time window.

β_1 is the first coefficient of interest, that is associated with our main explanatory variable - labelled *Migration* for simplicity. This is calculated at the origin country level as the sum of the interactions between the number of migrants from country c working in technological area ta in time window tw, resident in metropolitan area met, and a dummy R that takes the value 1 if the metropolitan area of destination has a comparative advantage in technology t (part of the technological area ta):

$$
Migration_{c,t,tw-1} = \sum_{met} MIG_{c,met,ta,tw-1} * R_{met,t \in ta,tw-1} \tag{2}
$$

We consider that a metropolitan area met has a comparative advantage in technology t if its relative specialization index is equal or greater than 1. The RTA of metropolitan areas is calculated as follows:

$$
RTA_{met,t,tw} = \frac{pat_{met,t,tw} / \sum_{t} pat_{met,tw}}{\sum_{c} pat_{t,tw} / \sum_{c} \sum_{t} pat_{tw}} \tag{3}
$$

where c refers to all countries.

The dependent variable is, for each specification, either the growth of number of patents or the entry of a new technology. As measure of growth we use the compound average growth rate (CAGR) in technology t for country c between the five years separating time window *tw* and *tw-1*, conditional on $pat_{tw-1} > 0$, that is:

$$Growth_{c,t,tw} = (\frac{pat_{c,t,tw}}{pat_{c,t,tw-1}})^{\frac{1}{5}} - 1 \ \ if \ \ pat_{tw-1} > 0 \tag{4}$$

Entry of a new technology in a given country is computed as follows: first, we measure the relative technological specialization for each country of origin, using the RTA:

$$RTA_{c,t,tw} = \frac{pat_{c,t,tw} / \sum_{t} pat_{c,tw}}{\sum_{c} pat_{t,tw} / \sum_{c} \sum_{t} pat_{tw}} \tag{5}$$

where $pat_{c,t,tw}$ is the number of patents that country c produced in technology t in time window *tw*. The *Entry* proxy measures whether country c starts to develop a comparative advantage in a new technology. The variable is a dummy that takes the value 1 if the RTA of country c is smaller than 1 in technology t in time window *tw-1* and equal or greater than 1 in time window *t*.[12] When using *Growth* as dependent variable we introduce a control for the total number of patents lagged one time window ($Tot\_pat_{c,t,tw-1} = \sum_{t} pat_{c,t,tw-1}$), while when using *Entry* we control for the continuous value of the actual RTA, always at $tw - 1$.

Next, the main goal of this paper is to understand how knowledge remittances and relatedness interplay in shaping the path of technological diversification of the countries of origin. We compute relatedness density between technologies following Rigby (2015), Boschma et al. (2015) and Balland et al. (2019), among others. First, we measure technological relatedness counting the frequency with which technologies i and j appear on the same patent and normalizing this count by total number of patents that record claims for i and j, in order to avoid the influence of size effects - technological relatedness is recomputed from scratch for

[12]In online Appendix 6 we present a robustness check in which we define *Entry* as a dummy that takes the value 1 if the RTA of country c in technology t is equal or smaller than 0.5 in time window *tw-1*, and equal or greater than 1 in time window *tw*. Table A6 shows that our main conclusions on developing countries hold even when defining *Entry* in this stricter, alternative way.

every time window.[13] The outcome is a $t*t$ network where the nodes are the technologies and the links their degree of relatedness. We then generate a dummy variable that takes the value 1 if the degree of relatedness of two technologies is ≥ 1. We then calculate the relatedness density, that measures the relatedness of the technology of interest to the set of technologies in which the country is already specialized. This measure is derived from the technological relatedness ($\phi_{i,j}$) of technology i to all the technologies j in which the country has relative specialization index > 1 (Equation 5), divided by the sum of technological relatedness of technology i to all the other technologies j:

$$Rel\_dens_{c,t,tw-1} = \frac{\sum_{j \in c, j \neq i, tw-1} \phi_{i,j,tw-1}}{\sum_{j \neq i} \phi_{i,j,tw-1}} * 100 \tag{6}$$

We then introduce in equation 1 an interaction variable between migration and relatedness density. A positive coefficient associated with this variable would suggest that relatedness reinforces the effect of knowledge remittances, confirming that knowledge brought in from abroad requires absorptive capacity to be understood (Cohen and Levinthal, 1990). On the other hand, a negative coefficient would imply that knowledge remittances act as substitute for relatedness, helping to diversify beyond the set of countries' technological capabilities and preventing the risk of lock-in.

This specification can incur in endogeneity issues due to omitted variables, reverse causality, and measurement error. We partially address the omitted variables issue including country per time ($\delta_{c,tw}$) and technology per time ($\delta_{t,tw}$) fixed effects, that allow us to control for time-variant characteristics that may correlate with both migration and diversification (such as the relative size of a technological class or country income). Yet, the choice of destinations of foreign inventors might be correlated with dynamics of specialization at both origin and destination. Bahar and Rapoport (2018) and Bahar et al. (2020) address this issue introducing a control for bilateral trade and FDI. A weakness of these measures is that they are not

[13] Using the association measure presented in Eck and Waltman (2009).

technology-specific and consider the overall bilateral flows. In our specification we introduce an alternative control, that is the total number of collaborative patents between the migrants' country of origin c and the metropolitan area of destination met in technology t in which the city of destination is specialized ($RTA_{met,t,tw} \geq 1$):

$$Copatents_{c,t,tw-1} = \sum_{met} Pat_{c,met,t,tw-1} * R_{met,t,tw-1} \tag{7}$$

In this way, we control for innovative collaborative activities between origin and destination that may drive inventors relocation.[14]

Next, lagging the variables of interest by one time window to minimize reverse causality does not completely resolve the issue, as both migration at time *tw-1* and diversification at time *tw* could be affected by long-term human capital investments in sending countries. We address these concerns by implementing an IV strategy. Following Frankel and Romer (1999), Ortega and Peri (2014) and Bahar and Rapoport (2018) we estimate a gravity model to compute predicted bilateral migration flows as follows:

$$\begin{aligned} Migrants_{c,met,tw} =& \alpha + \beta_1 Mig\_less\_skilled_{c,dc,tw} * share\_pop_{met \in dc,1980} + \beta_2 Distance_{met \in dc,c} \\ & + \beta_3 Contiguity_{met \in dc,c} + \beta_4 Colony_{dc,c} + \beta_5 Common\_language_{dc,c} \\ & + \beta_6 Common\_religion_{dc,c} + \gamma_{met} + \omega_c + \delta_{tw} + \epsilon_{c,met,tw} \end{aligned} \tag{8}$$

where the left hand side is the actual stock of migrant inventors from country c in metropolitan area met in time window tw. On the right-hand side we introduce three dummy variables at the country level: $Colony_{dc,c}$ indicating whether the two countries ever had a colony-coloniser relationship, $Common\_language_{dc,c}$ whether the two countries share the

[14] We are aware that our *Copatents* variable is not a substitute for bilateral trade and FDIs. Table A4 in online Appendix 4 presents the results when we add trade and FDI instead of *Copatents*. The results are robust to this alternative estimation.

same language, and $Common\_religion_{dc,c}$ whether they share the same religion. The data come from the CEPII *Gravity* dataset. To introduce variability at the metropolitan area level, we introduce a dummy indicating whether the metropolitan area of destination and the country of origin share a border $Contiguity_{met \in dc,c}$. We also introduce the straight line distance between the metropolitan areas of destination and the countries of origin ($Distance_{met \in dc,c}$). Finally, we proxy pre-existent diasporas at destination multiplying the stocks of less skilled migrants $Mig\_less\_skilled$ by origin country c, in destination country dc and time window tw with the population shares $share\_pop$ of metropolitan area met in a destination country dc. Data for unskilled migrants come from the Institute for Employment Research (IAB) and population data for metropolitan areas come from the History Database of the Global Environment (HYDE). We acknowledge that these variables may affect technology diffusion in different ways, above and beyond skilled migration (e.g., via trade or FDI), thus not meeting the exclusion restriction. Note, however, that our dependent variables are technology-specific. It is therefore more likely that our gravity variables would affect diversification into a new technology only by influencing skilled migrants working in that specific technology, rather than through other channels. Moreover, note that our instruments are not these variables per se, but the predicted influence of them on inventor migration flows.

Due to the high number of zeros in our dependent variable and its count nature, we estimate the equation by means of Pseudo-Poisson Maximum Likelihood (PPML) (Silva and Tenreyro, 2006). Once we estimate the predicted migration flows, we multiply them by a fixed value of R, based on the RTAs of metropolitan areas in the pre-sample period 1981-1985, that takes the value 1 if the metropolitan area of destination had a relative advantage in the technology under consideration, and finally we sum them up at the country level:

$$IV_{c,t,tw} = \sum_{met} \widehat{Migrants}_{c,met,tw} * R_{met,t,1981\_1985} \tag{9}$$

To account for the potential endogeneity of the interaction between *Migration* and

Rel_dens, we interact the IV with Rel_dens and use it as an additional exclusion restriction.

Moreover, since our main variable of interest is the sum of the product between specialization at destination and the number of migrants, we provide two falsification tests to rule out the possibility that our results are driven by only one of these dynamics. First, to verify that our results are not only driven by specialization at destination, we substitute the actual migration variable by randomizing the number of migrants. Second, to confirm that specialization at destination matters, we change the meaning of the dummy R that this time takes the value 1 if the metropolitan area of destination met does not have a comparative advantage in technology t.

To compute $Growth$ and the various $RTAs$ we use fractional counting, meaning that if a patent belongs to a number x of technologies (locations), it will be counted proportionally per technology (location): $1/x$. We transform $Migration$ and $Copatents$ using the inverse hyperbolic sine, a linear monotonic transformation similar to a logarithmic one, except that it is defined at zero (MacKinnon and Magee, 1990).

Table 3 presents basic descriptive statistics for the dependent variables, the variables of interest, and the control variables, firstly all together and then by type of country of origin. Not surprisingly, the average values in Table 3 witness a gap between developing countries and high-income economies in patenting activity. Also, the average number of collaborative patents is higher for high-income countries. Concerning migration, we notice that developing countries present a lower mean value but a higher maximum value, suggesting that migrants from developing countries may be more concentrated. Table A5 in online appendix presents the correlation matrix which shows that no concerns on collinearity are present. We add all these variables parsimoniously (in unreported results), to be sure that collinearity does not drive our results.

[Table 3 about here.]

4 Results

4.1 Baseline results

Table 4 estimates our main regressions on the pooled sample. As discussed earlier, our focal variable, *Migration*, is computed as the sum of the interactions between the number of migrants from country c working in technological area ta in time window tw and a dummy R that takes the value 1 if the metropolitan area of destination has a comparative advantage in technology t, part of the technological area ta.

In columns 1-3 we use the growth in the number of patents per technological class as the dependent variable. The coefficients of $Migration$ and *Relatedness density* are positive and statistically significant in all the specifications, while the interaction between these two variables presents a negative and significant coefficient. The coefficient for the number of collaborative patents is negative and significant at the one percent level. This result is somewhat counterintuitive. A potential explanation is that, as international collaborations mainly happen among a subset of countries, the variable is highly skewed and the coefficient reflects a spurious relationship due to the collinearity with the country-time fixed effects.

These results suggest that the presence of a high-skilled diaspora working in a destination specialized in a given technology has a positive impact on the number of patents that the country of origin files in that technology in the next five years. More specifically, doubling the number of migrant inventors working in a metropolitan area specialized in technology t increases the total number of patents filed in the country of origin in that technology by 3.7 percent. Although the coefficient may seem small, it is worth noting that a twofold increase implies a moderate number of inventors, as the average number of migrant inventors for the *Growth* sample is 158.86 (Table 3). Moreover, the negative coefficient associated with the interaction between migration and relatedness density ($Mig*rel$) suggests that the effect is stronger for technologies with lower degrees of relatedness, implying that knowledge remittances may act as substitute for relatedness. When Rel_dens is equal to the mean, doubling the stock of migrants increases the total number of patents filed by 1.1%. The effect of $Migration$ is either positive or non-significant for 90% of the observations, when sorted by the level of relatedness density.[15]

Results are similar in columns 4-6 for the case of *Entry*, our proxy of technological change, where we find a positive and statistically significant coefficient for migration and relatedness density, and a negative and significant coefficient for the interaction between the two. In this case, doubling the number of migrant inventors working in a metropolitan area specialized in technology t increases the probability that the country of origin starts to specialize in that technology by 6.5 percent. Note that in this sample the average number of migrant inventors is 44.44, suggesting that a relatively small increase in the number of inventors working abroad has a positive and highly significant effect on the probability of entry of a new technology in the country of origin. Here again, the negative coefficient associated to the interaction between migration and relatedness density implies that this effect specifically holds for technologies with a low degree of relatedness density, suggesting that having a high skilled diaspora helps the country of origin to diversify in technologies that will, otherwise, be unlikely to appear. Comparing our results to Bahar et al. (2020), we see that their equivalent coefficient is not always significant, which we attribute to the territorial breakdown by metro areas in destination countries we do.[16]

Doubling the stock of migrants when Rel_dens is equal to the mean increases the probability of entering in a new technology by 4%. The effect of $Migration$ is positive and significant for the 75%, and either positive or non-significant for the 90% of the observations, again when sorted by the level of relatedness density.[17]

[15] Additional computations are available in online Appendix 3.

[16] We repeat our baseline results in table A7 without breaking down migration and specialization data by metropolitan areas in our 5 destination countries. Interestingly, results barely hold, suggesting how important is to account for territorial differences in specialization and migration patterns.

[17] Additional computations are available in online Appendix 3.

In Table 5, we split the sample into developing and high-income countries. We notice that when isolating the group of high-income countries, the coefficients of *Migration* is only significant at the 10 percent level in the *Growth* sample, while the interaction between migration and relatedness density is never significant, for both our dependent variables. On the other hand, the results on developing countries largely confirm the ones on the pooled sample. This suggests that, for this group of countries, a twofold increase in the number of migrant inventors working in a metropolitan area specialized in technology t increases the total number of patents filed in the country of origin by 5.2 percent and the probability that the country starts to specialize in that technology by 7 percent. This last result is particularly significant since, on average, as for the pooled sample, a twofold increase on the number of inventors working abroad implies a relative small number of people (22.93). The negative sign of the interaction between migration and relatedness density confirms that the effect specifically holds for technologies with a low degree of relatedness density and that knowledge remittances to developing countries act as a substitute for the presence of absorptive capacity. Appendix 8 digs deeper into the analysis of migration and relatedness across income levels by plotting the interaction of our focal variables with GDP per capita. The evidence presented there is coherent with the results shown when splitting the sample according to the income level.

[Table 4 about here.]

[Table 5 about here.]

4.2 Instrumental Variables

Tables 6 and 7 present the results for the IV strategy respectively on the pooled sample and separately on developing and high income countries.[18] We report the Kleibergen-Paap F statistics for all the estimations to test if our instrument is weak. As the values are always

[18]Table A8 in the online Appendix presents the results for the first stage.

larger than 10 - and in most of the cases larger than 100 (Lee et al., 2020), we are confident that there are no reasons for concern. The IV estimations largely confirm our baseline results, suggesting a positive and significant relationship between inventors migration and home countries technological diversification for the pooled sample and for developing countries. Next, as in the baseline results, we find a negative coefficient associated to the interaction between migration and relatedness, suggesting a substitution effect between external knowledge flows and developing related activities.

IV coefficients are very similar in magnitude compared to OLS when considering *Entry* in the pooled sample regression, and between 0.5 and two times larger for the rest of the estimations. As we hypothesize endogeneity to be driven by reverse causality, we would have expected our OLS coefficients to be biased upwards. We believe that there might be substantial reasons behind the downward bias in OLS estimates. First, as MNCs are important drivers of the international mobility of these skilled workers (international recruitment, cross-country transfers, etc.), they possibly internalize some of the gains and spillovers migrants produce (Ganguli, 2015). As our analysis aggregates the data by country-areas, we cannot break down the reinforcing effect of migration and MNCs, as found in Breschi et al. (2017). Second, skilled migration and proximity (geographical and others) tend to be substitutes (Oettl and Agrawal, 2008; Breschi et al., 2017). If this is the case, skilled people will tend to move to places where knowledge flows are more scarce, precisely because these flows cannot be accessed in any other way (e.g., between high-income and developing economies). In this scenario OLS estimates would underestimate the true relationship due to a negative correlation between migration and the errors.

[Table 6 about here.]

[Table 7 about here.]

4.3 Falsification tests

To rule out that our results are only driven by specializations in the metropolitan area of destination, we estimate two random models that randomize the stock of migrants 500 times (Bahar et al., 2020). In the first model we use a uniform distribution, while in the second model we shuffle the real number of migrants so to obtain a random variable with the same distribution of the original one. Figure 2 presents the results for the first model, which clearly shows that none of the estimated coefficients using the random variable is statistically significant, both for $Migration$ and the interaction between $Migration$ and Rel_dens. Figure 3 shows the results for the second model and also in this case the vast majority of coefficients estimated using the random variable are not significant (the number of significant coefficients ranges from 1 to 13).

Next, to rule out that the impact is only driven by migrant inventors stocks, we replicate the main specification reversing the sense of the dummy variable R that takes the value 1 if the metropolitan area of destination has 0 patents in technology t. Thus, our main variable of interest is the sum of the interactions between the actual number of migrants from country c working in technological area ta in time window tw and the dummy R that takes the value 1 if the RTA of the metropolitan area of destination is equal to 0 in technology t. Table 8 presents the results. We find that the coefficient of $Migration$ is not significant in all the specifications. On the other hand, some coefficients for the interaction between migration and relatedness density are significant, but positive, which we attribute to spillovers brought in by emigrants in different, but related technologies.

[Figure 2 about here.]

[Figure 3 about here.]

[Table 8 about here.]

4.4 International inventions

So far, we computed our measure of growth and technological change on all patent families regardless of the quality of inventions. To consider this latter element, we replicate the analysis restricting the sample only to internationally oriented patents - around 25 percent of the total, which we define as those whose families include applications to several countries' patent offices as well as those including applications in just one country, but filed by foreign firms (the patent applicant's country, as per its address, does not coincide with that of the patent office). The underlying hypothesis is that since the procedure to protect an idea internationally is particularly costly, these inventions represent the most valuable ones. Thus, we analyze the effect of knowledge flows on the development of high-quality inventions.

Table A10 in the online appendix provides the results for the pooled sample, confirming the results of Table 4. Concerning $Growth$, results confirm the magnitude of the $Migration$'s coefficient, that remains quite stable with a slight increase (0.1 percent), while for *Entry* we notice a slight decrease from 0.7 to 0.5. In Table A11 we split the sample into developing and high-income countries. Also in this case the results are similar to Table 5, with the exception that we find a significant coefficient of $Migration$ and the interaction on the sample of hig-income countries when using $Growth$ as dependent variable. The magnitude of the coefficient (2.1) is lower than for developing countries (5.2). Overall, we can conclude that the results are confirmed when we restrict the sample to international inventions only.

4.5 Further robustness checks

To assess the robustness of our results, we run a number of alternative estimations. The estimations in online Appendix 12 mitigate concerns on our results being driven by a group of outliers by replicating the analysis excluding the countries with the most sizeable diaspora, namely China (Table A13) and India (Table A14). In the same spirit, we exclude the US as destination country (Table A15). Following Bahar and Rapoport (2018) and Bahar et al.

(2020), we test the robustness of our results including additional controls for bilateral trade and FDI (Table A4). Online Appendix 13 is dedicated to the estimations with alternative dependent variables. In Table A16, to address concerns on possible reverse causality, we run the estimations by excluding international collaborations from the dependent variable. Following the same logic, in Table A17 we exclude PCT applications. In Table A18 presented in online Appendix 14, we transform our main variables of interest using a regular logarithmic transformation, which exclude zero cells from our estimations. Finally, in Table A19 of Appendix 15 we compute migrant inventors at the IPC4 level. Most of our results and conclusions hold for all the specifications.

5 Conclusions

In this paper we analyze the relationship between high-skilled migration and the technological diversification of the migrants' countries of origin. In particular, we investigate whether migrant inventors transfer productive knowledge back home and encourage development of new technologies in which the destination areas are specialized. One of the main novelty of this paper is that we take into account the uneven distribution of knowledge and the consequent migrants' concentration at destination (Carlino and Kerr, 2015) by breaking down at the metropolitan area level the way in which our focal explanatory variables are computed. Also, we aim at understanding whether the transfers of knowledge from abroad foster unrelated diversification, allowing the country of origin to extend the set of technological capabilities and preventing lock-in. As technological development is a strong predictor of economic and social development (Hidalgo et al., 2007; Hartmann et al., 2017), we specifically focus on developing countries and on the most common destinations, that is, the US, Germany, United Kingdom, Switzerland and France.

Our results suggest a positive and statistically significant effect of high-skilled migration on the direction of technical change back home. More importantly, we find that external knowledge from abroad is particularly beneficial for the development of technologies with a low degree of relatedness, thus fostering unrelated diversification in the home countries, and promoting technological structural change (Neffke et al., 2018). Our results are confirmed when we restrict the sample to international inventions only and are robust to several alternative specifications and our instrumental variables approach.

Next, we also find that our results are critical for developing countries' innovation and diversification. That is, having a high-skilled diaspora helps developing economies to access foreign knowledge and catching-up with countries at the technological frontier. Moreover, fostering unrelated diversification, knowledge remittances help developing countries to prevent the risk of lock-in and promote long-term development (Saviotti and Frenken, 2008).

Our data provide detailed information on the localization of a great number of worldwide patent families. Yet, they do not allow us to identify the specific channel through which migrants transfer knowledge from destination areas to their home countries. We hypothesized that high-skilled migrants keep contacts with their countries of origin and transfer the knowledge acquired at destination to their social networks back in the country of origin. They may return back home, on a permanent or temporary basis, after some time abroad, with new skills and contacts. Future research, possibly at the micro-level, could investigate the specific mechanisms behind our results.

The focus of our analysis is on technological development. However, we did not investigate to which extent it translates into new production and export capacity for the migrants' countries of origin. This open question may guide further research aimed at understanding the role of knowledge flows in connecting technological and economic diversification.

Data availability statement: The data that support the findings of this study are openly available at 10.6084/m9.figshare.16867249.v3

References

Agrawal, A., Kapur, D., McHale, J., and Oettl, A. (2011). Brain drain or brain bank? the impact of skilled emigration on poor-country innovation. *Journal of Urban Economics*, 69(1):43–55.

Akcigit, U., Grigsby, J., and Nicholas, T. (2017). Immigration and the rise of american ingenuity. *American Economic Review*, 107(5):327–31.

Arora, A., Belenzon, S., and Lee, H. (2018). Reversed citations and the localization of knowledge spillovers. *Journal of Economic Geography*, 18(3):495–521.

Bahar, D. (2020). Diasporas and economic development: A review of the evidence and policy.

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

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

Bahar, D. and Rapoport, H. (2018). Migration, knowledge diffusion and the comparative advantage of nations. *The Economic Journal*, 128(612):F273–F305.

Balland, P.-A., Boschma, R., Crespo, J., and Rigby, D. L. (2019). Smart specialization policy in the european union: relatedness, knowledge complexity and regional diversification. *Regional Studies*, 53(9):1252–1268.

Baruffaldi, S. H. and Landoni, P. (2012). Return mobility and scientific productivity of researchers working abroad: The role of home country linkages. *Research policy*, 41(9):1655–1665.

Beine, M., Docquier, F., and Özden, Ç. (2011). Diasporas. *Journal of Development Economics*, 95(1):30–41.

Bergquist, K., Fink, C., and Raffo, J. (2017). Identifying and ranking the world's largest clusters of inventive activity'. *The Global Innovation Index 2017: Innovation Feeding the World*, pages 161–76.

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

Bhagwati, J. N. (1976). Taxing the brain drain. *Challenge*, 19(3):34–38.

Boschma, R. (2017). Relatedness as driver of regional diversification: A research agenda. *Regional Studies*, 51(3):351–364.

Boschma, R., Balland, P.-A., and Kogler, D. F. (2015). Relatedness and technological change in cities: the rise and fall of technological knowledge in us metropolitan areas from 1981 to 2010. *Industrial and corporate change*, 24(1):223–250.

Bosetti, V., Cattaneo, C., and Verdolini, E. (2015). Migration of skilled workers and innovation: A european perspective. *Journal of International Economics*, 96(2):311–322.

Branstetter, L., Li, G., and Veloso, F. (2015). The changing frontier: Rethinking science and innovation policy. In *The Rise of International Coinvention*, pages 135–168. University of Chicago Press.

Breschi, S., Lissoni, F., and Malerba, F. (2003). Knowledge-relatedness in firm technological diversification. *Research policy*, 32(1):69–87.

Breschi, S., Lissoni, F., and Miguelez, E. (2017). Foreign-origin inventors in the usa: testing for diaspora and brain gain effects. *Journal of Economic Geography*, 17(5):1009–1038.

Carlino, G. and Kerr, W. R. (2015). Agglomeration and innovation. In *Handbook of regional and urban economics*, volume 5, pages 349–404. Elsevier.

Chellaraj, G., Maskus, K. E., and Mattoo, A. (2008). The contribution of international graduate students to us innovation. *Review of International Economics*, 16(3):444–462.

Choudhury, P. (2016). Return migration and geography of innovation in mnes: a natural experiment of knowledge production by local workers reporting to return migrants. *Journal of Economic Geography*, 16(3):585–610.

Choudhury, P. and Kim, D. Y. (2019). The ethnic migrant inventor effect: Codification and recombination of knowledge across borders. *Strategic Management Journal*, 40(2):203–229.

Coe, D. T. and Helpman, E. (1995). International r&d spillovers. *European economic review*, 39(5):859–887.

Coe, D. T., Helpman, E., and Hoffmaister, A. W. (2009). International r&d spillovers and institutions. *European Economic Review*, 53(7):723–741.

Cohen, W. M. and Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, 35(1):128–152.

Crescenzi, R., Dyevre, A., and Neffke, F. (2020). Innovation catalysts: how multinationals reshape the global geography of innovation.

de Rassenfosse, G., Kozak, J., and Seliger, F. (2019). Geocoding of worldwide patent data. *Available at SSRN 3425764*.

Di Iasio, V. and Miguelez, E. (2021). Ties that bind data. doi: 10.6084/m9.figshare.16867249.v3.

Dosi, G. (1997). Opportunities, incentives and the collective patterns of technological change. *The economic journal*, 107(444):1530–1547.

Eck, N. J. v. and Waltman, L. (2009). How to normalize cooccurrence data? an analysis of some well-known similarity measures. *Journal of the American society for information science and technology*, 60(8):1635–1651.

Elekes, Z., Boschma, R., and Lengyel, B. (2019). Foreign-owned firms as agents of structural change in regions. *Regional Studies*, 53(11):1603–1613.

Essletzbichler, J. (2015). Relatedness, industrial branching and technological cohesion in us metropolitan areas. *Regional Studies*, 49(5):752–766.

Fackler, T. A., Giesing, Y., and Laurentsyeva, N. (2020). Knowledge remittances: Does emigration foster innovation? *Research Policy*, 49(9):103863.

Frankel, J. A. and Romer, D. H. (1999). Does trade cause growth? *American economic review*, 89(3):379–399.

Ganguli, I. (2015). Immigration and ideas: What did russian scientists "bring" to the united states? *Journal of Labor Economics*, 33(S1):S257–S288.

Hartmann, D., Guevara, M. R., Jara-Figueroa, C., Aristarán, M., and Hidalgo, C. A. (2017). Linking economic complexity, institutions, and income inequality. *World Development*, 93:75–93.

Hausmann, R., Klinger, B., et al. (2007). The structure of the product space and the evolution of comparative advantage. Technical report, Center for International Development at Harvard University.

Hidalgo, C. A., Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Glaeser, E., He, C., Kogler, D. F., Morrison, A., et al. (2018). The principle of relatedness. In *International conference on complex systems*, pages 451–457. Springer.

Hidalgo, C. A., Klinger, B., Barabási, A.-L., and Hausmann, R. (2007). The product space conditions the development of nations. *Science*, 317(5837):482–487.

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

Jaffe, A. B. (1986). Technological opportunity and spillovers of r&d: evidence from firms' patents, profits and market value.

Jaffe, A. B. and De Rassenfosse, G. (2019). Patent citation data in social science research: Overview and best practices. In *Research Handbook on the Economics of Intellectual Property Law*. Edward Elgar Publishing.

Keller, W. (2004). International technology diffusion. *Journal of economic literature*, 42(3):752–782.

Kerr, S. P., Kerr, W., Özden, Ç., and Parsons, C. (2016). Global talent flows. *Journal of Economic Perspectives*, 30(4):83–106.

Kerr, S. P. and Kerr, W. R. (2018). Global collaborative patents. *The Economic Journal*, 128(612):F235–F272.

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

Kerr, W. R. (2010). The agglomeration of us ethnic inventors. In *Agglomeration economics*, pages 237–276. University of Chicago Press.

Kerr, W. R. (2018). Heterogeneous technology diffusion and ricardian trade patterns. *The World Bank Economic Review*, 32(1):163–182.

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

Kogler, D. F., Rigby, D. L., and Tucker, I. (2013). Mapping knowledge space and technological relatedness in us cities. *European Planning Studies*, 21(9):1374–1391.

Lee, D. L., McCrary, J., Moreira, M. J., and Porter, J. (2020). Valid t-ratio inference for iv. *arXiv preprint arXiv:2010.05058*.

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

MacKinnon, J. G. and Magee, L. (1990). Transforming the dependent variable in regression models. *International Economic Review*, pages 315–339.

Marino, A., Mudambi, R., Perri, A., and Scalera, V. G. (2019). Ties that bind: The role of ethnic inventors in multinational enterprises' knowledge creation. In *Academy of Management Proceedings*, volume 2019, page 12100. Academy of Management Briarcliff Manor, NY 10510.

Martinez, C. (2010). Insight into different types of patent families.

McAuliffe, M. and Ruhs, M. (2017). World migration report 2018. *Geneva: International Organization for Migration*.

Miguelez, E. (2018). *Inventor diasporas and the internationalization of technology*. The World Bank.

Miguelez, E. and Fink, C. (2017). Measuring the international mobility of inventors. In *The International Mobility of Talent and Innovation*, pages 114–161. Cambridge University Press.

Miguelez, E., Raffo, J., Chacua, C., Massimiliano Coda-Zabetta, M., Yin, D., Lissoni, F., and Tarasconi, G. (2019). Tied in: the global network of local innovation.

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

Morrison, G., Riccaboni, M., and Pammolli, F. (2017). Disambiguation of patent inventors and assignees using high-resolution geolocation data. *Scientific data*, 4:170064.

Moser, P., Voena, A., and Waldinger, F. (2014). German jewish émigrés and us invention. *American Economic Review*, 104(10):3222–55.

Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the us labor market. *The Quarterly Journal of Economics*, 118(2):549–599.

Neffke, F., Hartog, M., Boschma, R., and Henning, M. (2018). Agents of structural change: The role of firms and entrepreneurs in regional diversification. *Economic Geography*, 94(1):23–48.

Neffke, F., Henning, M., and Boschma, R. (2011). How do regions diversify over time? industry relatedness and the development of new growth paths in regions. *Economic geography*, 87(3):237–265.

Oettl, A. and Agrawal, A. (2008). International labor mobility and knowledge flow externalities. *Journal of international business studies*, 39(8):1242–1260.

Ortega, F. and Peri, G. (2014). The aggregate effects of trade and migration: Evidence from oecd countries. In *The Socio-Economic Impact of Migration Flows*, pages 19–51. Springer.

Petralia, S., Balland, P.-A., and Morrison, A. (2017). Climbing the ladder of technological development. *Research Policy*, 46(5):956–969.

Pinheiro, F. L., Alshamsi, A., Hartmann, D., Boschma, R., and Hidalgo, C. (2018). Shooting low or high: Do countries benefit from entering unrelated activities? *Papers in Evolutionary Economic Geography*, 18(07).

Rigby, D. L. (2015). Technological relatedness and knowledge space: entry and exit of us cities from patent classes. *Regional Studies*, 49(11):1922–1937.

Romer, P. M. (1994). The origins of endogenous growth. *Journal of Economic perspectives*, 8(1):3–22.

Saviotti, P. P. and Frenken, K. (2008). Export variety and the economic performance of countries. *Journal of Evolutionary Economics*, 18(2):201–218.

Saxenian, A. (2007). *The new argonauts: Regional advantage in a global economy*. Harvard University Press.

Schmoch, U. (2008). Concept of a technology classification for country comparisons. *Final report to the world intellectual property organisation, WIPO*.

Silva, J. S. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, 88(4):641–658.

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

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

Verginer, L. and Riccaboni, M. (2021). Talent goes to global cities: The world network of scientists' mobility. *Research policy*, 50(1):104127.

Whittle, A., Lengyel, B., Kogler, D. F., et al. (2020). Understanding regional branching knowledge diversification via inventor collaboration networks. Technical report.

WIPO (2019). The geography of innovation: Local hotspots, global networks. Technical report, World Intellectual Property Organization.

Yin, D. and Motohashi, K. (2018). Inventor name disambiguation with gradient boosting decision tree and inventor mobility in china (1985-2016). Technical report, Research Institute of Economy, Trade and Industry (RIETI).

Table 1: Top 20 migration corridors, 2000 - 2009

United States			Europe		
Origin	**Destination**	**Inventors**	**Origin**	**Destination**	**Inventors**
India	San Diego, CA	4736	Germany	Zürich, CH	1848
India	San José, CA	4439	Germany	Basel, CH	1615
China	San Diego, CA	4176	Netherlands	London, UK	1331
China	San José, CA	4153	France	London, UK	837
China	Boston-Worcester, MA	3123	France	Lausanne, CH	831
India	Boston-Worcester, MA	2099	Netherlands	Paris, FR	676
Canada	San Diego, CA	1845	Germany	London, UK	597
Canada	Boston-Worcester, MA	1818	United Kingdom	Basel, CH	595
China	Middlesex-Somerset, NJ	1678	United States	München, DE	589
China	Oakland, CA	1537	United States	London, UK	586
China	Chicago, IL	1507	United Kingdom	Paris, FR	484
Canada	San José, Ca	1436	Germany	Paris, FR	460
India	Chicago, IL	1358	Germany	Lausanne, CH	450
China	San Francisco, CA	1332	Italy	London, UK	441
China	Philadelphia, PA-NJ	1328	France	Basel, CH	351
United Kingdom	Boston-Worcester, MA	1319	France	Genéve, CH	338
India	Oakland, CA	1262	Italy	Paris, FR	314
India	Middlesex-Somerset, NJ	1191	United States	Paris, FR	305
United Kingdom	San Francisco, CA	1169	Greece	Mannheim-Lüdwigshafen, DE	278
Canada	San Francisco, CA	1144	United Kingdom	Frankfurt Am Main, DE	272

Source: Authors' calculations based on Miguelez and Fink (2017) data and OECD REGPAT database.

Table 2: Top 20 migration corridors, 2000 - 2009: no India and China for the US, only developing countries for Europe

United States			Europe		
Origin	**Destination**	**Inventors**	**Origin**	**Destination**	**Inventors**
Canada	San Diego, CA	1845	India	London, UK	222
Canada	Boston-Worcester, MA	1818	China	London, UK	197
Canada	San José, CA	1436	Russia	Mannheim-Lüdwigshafen, DE	141
United Kingdom	Boston-Worcester, MA	1319	China	München, DE	127
United Kingdom	San Francisco, CA	1169	China	Paris, FR	126
Canada	San Francisco, CA	1144	Tunisia	Paris, FR	100
United Kingdom	San Diego, CA	1071	Russia	Berlin, DE	87
United Kingdom	San José, CA	1049	China	Cambridge, UK	86
Germany	Boston-Worcester, MA	948	Algeria	Paris, FR	81
Korea	San Diego, CA	943	Russia	London, UK	79
Germany	San José, CA	898	India	Paris, FR	74
Korea	San José, CA	799	India	Mannheim-Lüdwigshafen, DE	73
Germany	San Diego, CA	799	Russia	Paris, FR	70
Israel	San José, CA	726	China	Stüttgart, DE	69
Germany	San Francisco, CA	711	South Africa	London, UK	69
Japan	San José, CA	646	Malaysia	London, UK	66
France	San Diego, CA	641	Ukraine	Rührgebiet, DE	66
Canada	Oakland, CA	632	Morocco	Paris, FR	65
France	San José, CA	602	Russia	Rührgebiet, DE	63
France	Boston-Worcester, MA	596	Romania	Paris, FR	63

Source: Authors' calculations based on Miguelez and Fink (2017) data and OECD REGPAT database.

Table 3: Descriptive statistics

	Growth sample											
	Pooled (Obs = 93510)				Developing (Obs = 42016)				High income (Obs = 51494)			
	mean	sd	min	max	mean	sd	min	max	mean	sd	min	max
Growth	-0.09	0.46	-1	6	-0.21	0.57	-1	6	0.01	0.32	-1	4
Migration	158.86	709.67	0	19561	132.51	866.11	0	19561	180.35	549.06	0	9412
Tot_pat	96.16	760.25	0	76351	38.09	409.14	0	31428	143.55	952.89	0	76351
Rel_dens	35.52	16.58	0	100	30.80	17.48	0	100	39.37	14.74	0	100
Copatents	5.03	50.26	0	3607	0.22	2.85	0	189	8.95	67.43	0	3607
	Entry sample											
	Pooled (Obs = 397500)				Developing (Obs = 318000)				High income (Obs = 79500)			
	mean	sd	min	max	mean	sd	min	max	mean	sd	min	max
Entry	0.06	0.23	0	1	0.05	0.22	0	1	0.08	0.28	0	1
Migration	44.44	359.61	0	19561	22.93	328.35	0	19561	130.49	454.00	0	9412
RTA	1.21	19.81	0	5036	1.19	22.01	0	5036	1.27	4.95	0	489
Rel_dens	16.92	17.94	0	100	11.87	14.78	0	100	37.13	14.99	0	100
Copatents	1.52	21.74	0	3581	0.29	9.62	0	2516	6.41	44.31	0	3581

Table 4: Pooled sample

	Growth			Entry		
	(1)	(2)	(3)	(4)	(5)	(6)
Migration	0.0097***	0.0368***	0.0367***	0.0017*	0.0066***	0.0065***
	(0.0027)	(0.0058)	(0.0057)	(0.0008)	(0.0011)	(0.0011)
Rel_dens	0.0057***	0.0089***	0.0087***	0.0011***	0.0017***	0.0015***
	(0.0004)	(0.0007)	(0.0007)	(0.0002)	(0.0002)	(0.0002)
Mig*rel		-0.0007***	-0.0007***		-0.0002***	-0.0001***
		(0.0001)	(0.0001)		(0.0000)	(0.0000)
Copatents			-0.0104**			-0.0217***
			(0.0035)			(0.0018)
Tot_pat	-0.1295***	-0.1292***	-0.1263***			
	(0.0064)	(0.0063)	(0.0065)			
RTA				-0.0002***	-0.0002***	-0.0002***
				(0.0001)	(0.0001)	(0.0001)
Country per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	93494	93494	93494	318000	318000	318000
R-squared	0.4427	0.4447	0.4450	0.0641	0.0645	0.0661

Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All RHS variables are lagged of one time window. Migration, Tot_pat, and Copatents are transformed using the inverse hyperbolic sine.

Table 5: Developing and high-income

	Growth					
		High-income			Developing	
	(1)	(2)	(3)	(4)	(5)	(6)
Migration	0.0093**	0.0187*	0.0187*	0.0141**	0.0516***	0.0515***
	(0.0031)	(0.0072)	(0.0073)	(0.0049)	(0.0070)	(0.0071)
Rel_dens	0.0044***	0.0056***	0.0056***	0.0066***	0.0106***	0.0105***
	(0.0005)	(0.0010)	(0.0010)	(0.0009)	(0.0008)	(0.0008)
Mig*rel		-0.0003	-0.0003		-0.0011***	-0.0011***
		(0.0001)	(0.0002)		(0.0002)	(0.0002)
Copatents			0.0005			-0.0212
			(0.0038)			(0.0121)
Tot_pat	-0.1181***	-0.1184***	-0.1185***	-0.1408***	-0.1391***	-0.1379***
	(0.0089)	(0.0090)	(0.0089)	(0.0086)	(0.0085)	(0.0087)
Country per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	51489	51489	51489	41967	41967	41967
R-squared	0.3250	0.3254	0.3254	0.4842	0.4872	0.4874

	Entry					
		High-income			Developing	
	(1)	(2)	(3)	(4)	(5)	(6)
Migration	0.0020	0.0050	0.0049	0.0033***	0.0070***	0.0070***
	(0.0023)	(0.0046)	(0.0045)	(0.0008)	(0.0012)	(0.0012)
Rel_dens	0.0009***	0.0013**	0.0013**	0.0014***	0.0019***	0.0018***
	(0.0002)	(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0003)
Mig*rel		-0.0001	-0.0001		-0.0002***	-0.0002***
		(0.0001)	(0.0001)		(0.0000)	(0.0000)
Copatents			-0.0036			-0.0281***
			(0.0026)			(0.0074)
RTA	-0.0044**	-0.0044**	-0.0044**	-0.0002***	-0.0002***	-0.0002***
	(0.0013)	(0.0013)	(0.0013)	(0.0000)	(0.0000)	(0.0000)
Country per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	63600	63600	63600	254400	254400	254400
R-squared	0.0707	0.0708	0.0709	0.0804	0.0807	0.0811

Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All RHS variables are lagged of one time window. Migration, Tot_pat, and Copatents are transformed using the inverse hyperbolic sine.

Table 6: IV estimations

	(OLS)		(IV)	
	(1)	(2)	(3)	(4)
	Growth	Entry	Growth	Entry
Migration	0.0367***	0.0065***	0.0618***	0.0068**
	(0.0057)	(0.0011)	(0.0103)	(0.0024)
Rel_dens	0.0087***	0.0015***	0.0105***	0.0018***
	(0.0007)	(0.0002)	(0.0008)	(0.0002)
Mig*rel	-0.0007***	-0.0001***	-0.0012***	-0.0002***
	(0.0001)	(0.0000)	(0.0002)	(0.0001)
Country per time FE	Yes	Yes	Yes	Yes
Technology per time FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	93494	318000	93494	318000
R-squared	0.4450	0.0661	0.4444	0.0660
Underidentification test			0.000	0.000
Kleibergen-Paap statistics			107.873	389.814

Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Estimations reported in Column 1 and 3 include Tot_pat and Copatents as controls, while estimations in Column 2 and 4 include RTA and Copatents. All RHS variables are lagged of one time window. Migration, Tot_pat, and Copatents are transformed using the inverse hyperbolic sine transformation.

Table 7: IV estimations - developing and high-income

	Developing			
	(OLS)		(IV)	
	(1)	(2)	(3)	(4)
	Growth	Entry	Growth	Entry
Migration	0.0515***	0.0070***	0.0712***	0.0116***
	(0.0071)	(0.0012)	(0.0113)	(0.0022)
Rel_dens	0.0105***	0.0018***	0.0117***	0.0022***
	(0.0008)	(0.0003)	(0.0009)	(0.0003)
Mig*rel	-0.0011***	-0.0002***	-0.0014***	-0.0003***
	(0.0002)	(0.0000)	(0.0002)	(0.0001)
Country per time FE	Yes	Yes	Yes	Yes
Technology per time FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	41967	254400	42016	254400
R-squared	0.4874	0.0811	0.4877	0.0809
Underidentification test			0.000	0.000
Kleibergen-Paap statistics			57.815	309.660
	High-income			
	(OLS)		(IV)	
	(1)	(2)	(3)	(4)
	Growth	Entry	Growth	Entry
Migration	0.0187*	0.0049	0.0558*	-0.0039
	(0.0073)	(0.0045)	(0.0246)	(0.0174)
Rel_dens	0.0056***	0.0013**	0.0076***	0.0015*
	(0.0010)	(0.0004)	(0.0016)	(0.0006)
Mig*rel	-0.0003	-0.0001	-0.0007*	-0.0001
	(0.0002)	(0.0001)	(0.0003)	(0.0002)
Country per time FE	Yes	Yes	Yes	Yes
Technology per time FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	51489	63600	51494	63600
R-squared	0.3254	0.0709	0.3239	0.0706
Underidentification test			0.001	0.000
Kleibergen-Paap statistics			52.8508	64.595

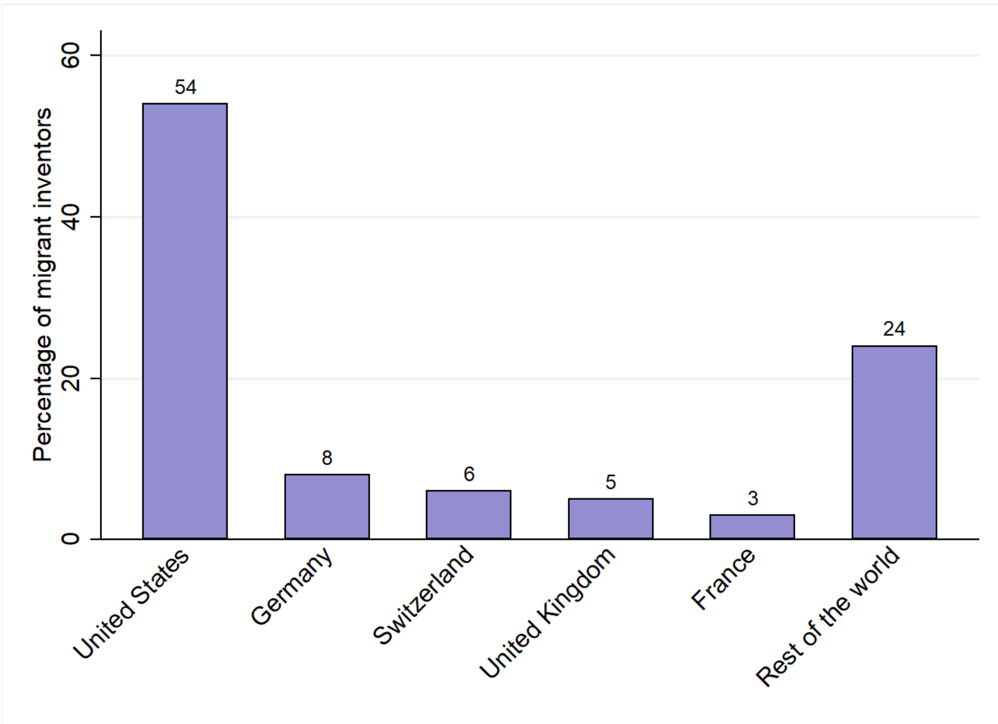
Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Estimations reported in Column 1 and 3 include Tot_pat and Copatents as controls, while estimations in Column 2 and 4 include RTA and Copatents. All RHS variables are lagged of one time window. Migration, Tot_pat, and Copatents are transformed using the inverse hyperbolic sine transformation.

Table 8: Falsification test 2: R=1 if RTA = 0

		(Growth)			(Entry)	
	Pooled	High-income	Developing	Pooled	High-income	Developing
	(1)	(2)	(3)	(4)	(5)	(6)
Migration (R=1 if RTA = 0)	0.00515	-0.00284	0.00900	-0.000625	-0.00164	-0.00376*
	(0.00530)	(0.00532)	(0.0123)	(0.00187)	(0.00321)	(0.00222)
Rel_dens	0.00578***	0.00359***	0.00781***	0.000487**	0.000447	0.00103***
	(0.000486)	(0.000449)	(0.000733)	(0.000197)	(0.000375)	(0.000297)
Mig*rel	-0.0000484	0.000253**	-0.000455*	0.000214***	0.000150*	0.000157**
	(0.000108)	(0.0000951)	(0.000246)	(0.0000515)	(0.0000773)	(0.0000718)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	93494	51489	41967	318000	63600	254400
R-squared	0.443	0.325	0.485	0.0666	0.0711	0.0808

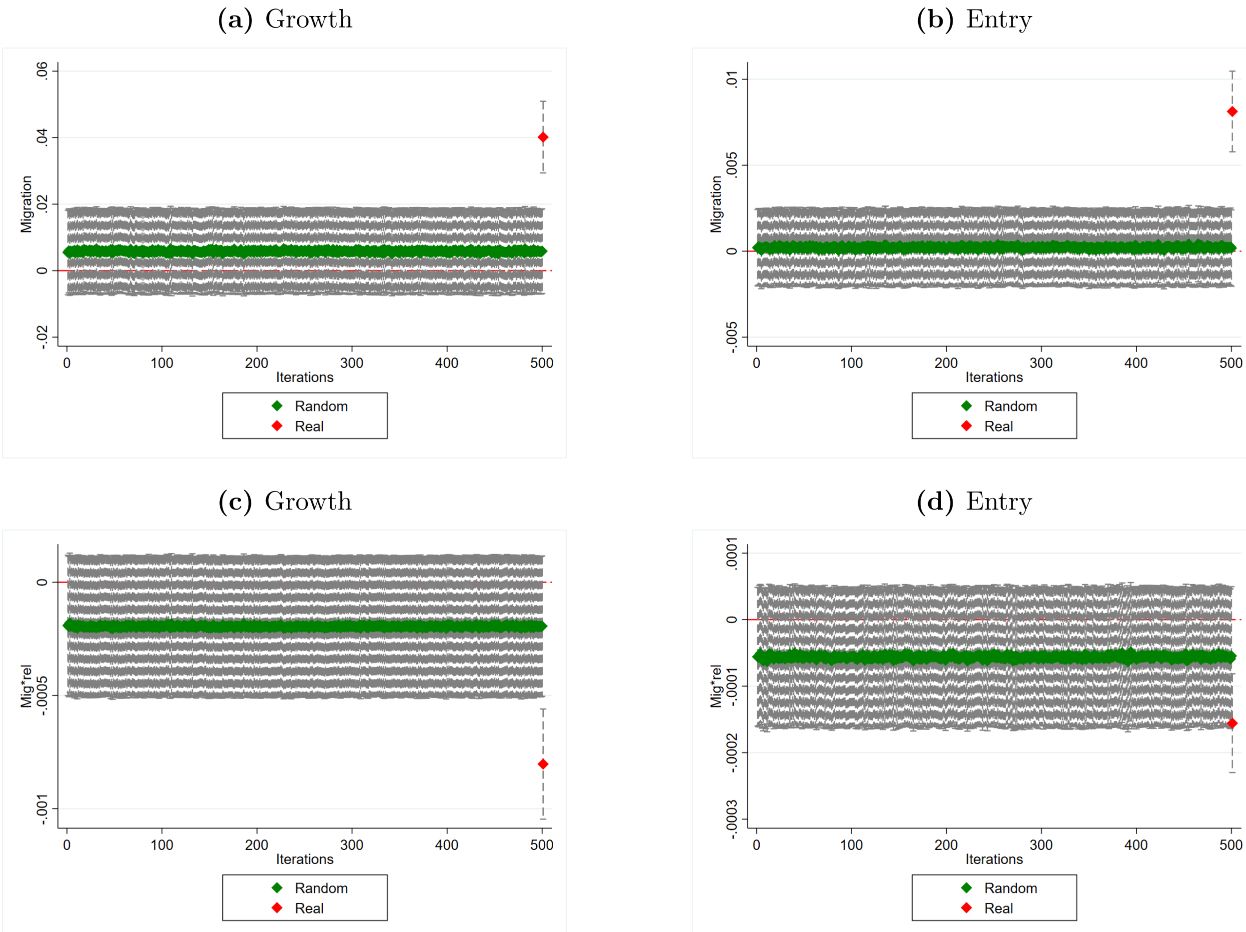
Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Estimations reported in Column 1, 2, and 3 include Tot_pat and Copatents as controls, while estimations in Column 4, 5, and 6 include RTA and Copatents. All RHS variables are lagged of one time window. Migration, Tot_pat, and Copatents are transformed using the inverse hyperbolic sine.

Figure 1: Migrant inventors stock



Source: Author's calculations based on Miguelez and Fink (2017) data.

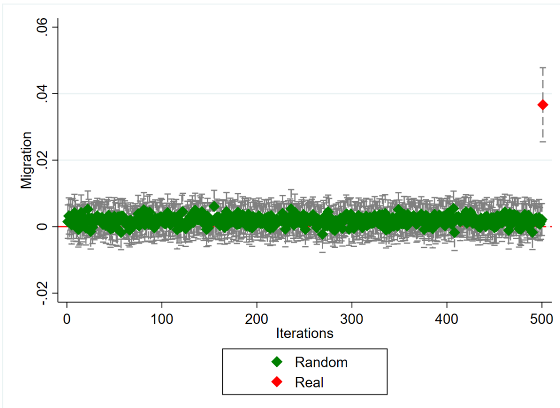
Figure 2: Random model 1



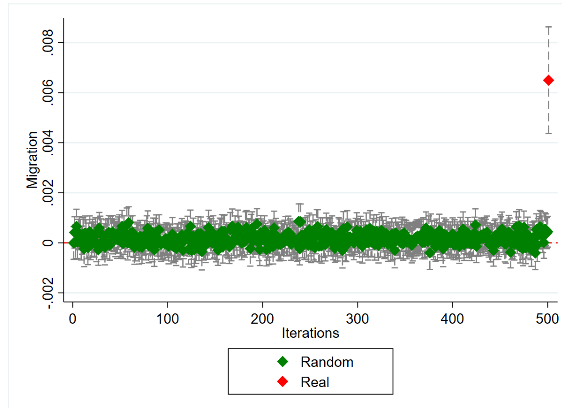
Notes: Summary of 500 estimations using random inventor figures (OLS). Figures a and b plot the estimators of $\beta_1 Migration$ from the baseline equation when substituting the real number of migrant inventors between countries with a random one, for each of 500 iterations. Figures c and d repeat the same exercise and plot the estimators of $\beta_3 Mig * rel$.The figure is based on a randomization approach that replaces the actual number of inventors with a random number, with no restrictions distributed uniformly from 0 to 1. The figure also includes, for reference, the estimation using the actual number of migrant inventors (in blue). Whiskers represent 95% confidence intervals, based on SE clustered at the country level.

Figure 3: Random model 2

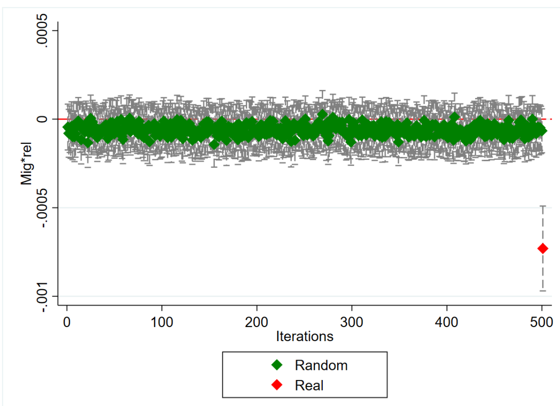
(a) Growth



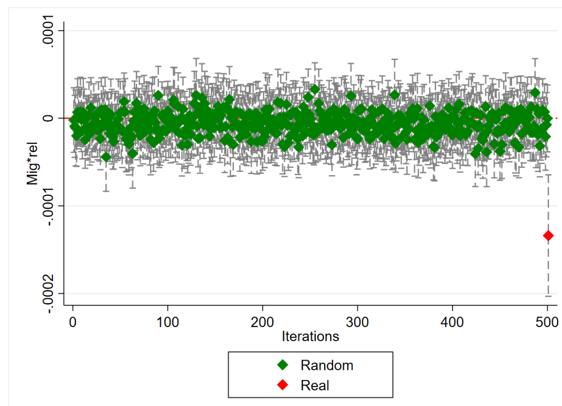
(b) Entry



(c) Growth



(d) Entry



Notes: Summary of 500 estimations using random inventor figures (OLS). Figures a and b plot the estimators of $\beta_1 Migration$ from the baseline equation when substituting the real number of migrant inventors between countries with a random one, for each of 500 iterations. Figures c and d repeat the same exercise and plot the estimators of $\beta_3 Mig * rel$. The figure is based on a randomization approach such that the real and the random number of inventors have the same sample mean and distribution. The figure also includes, for reference, the estimation using the actual number of migrant inventors (in blue). Whiskers represent 95% confidence intervals, based on SE clustered at the country level.

Online Appendix

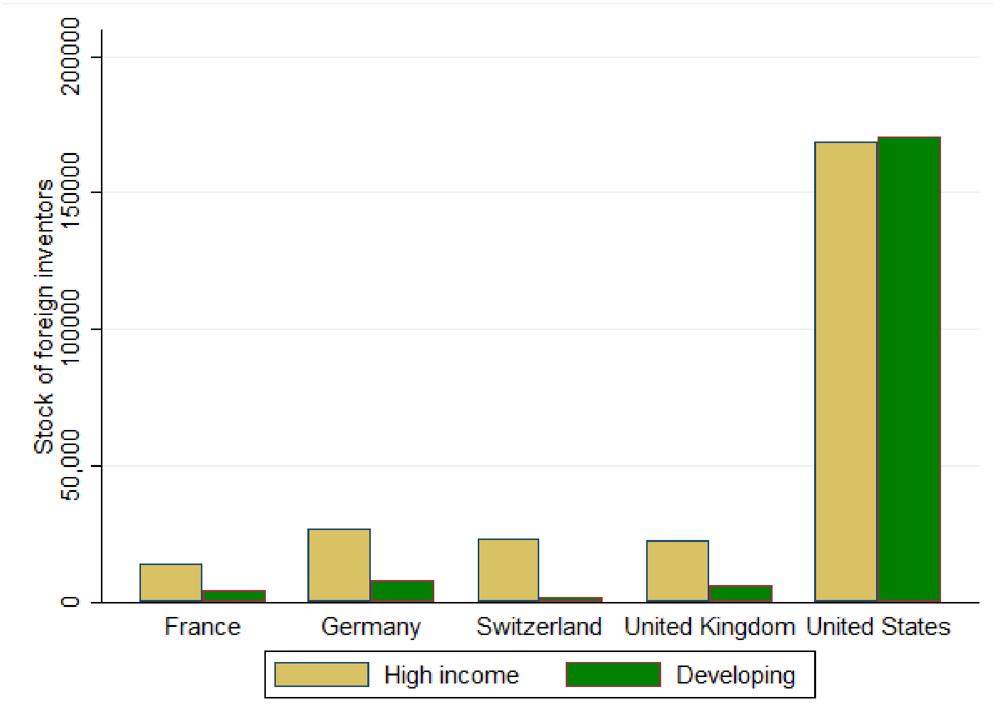
Appendix 1 List of countries in the sample

- **Developing countries:** Albania, Algeria, Argentina, Armenia, Azerbaijan, Bangladesh, Belarus, Belize, Benin , Bolivia, Bosnia and Herzegovina, Botswana, Brazil, Bulgaria, Burkina Faso, Burundi, Cambodia, Cameroon, Central African Republic, Chad, Chile, China, Colombia, Congo, Costa Rica, Cuba, Cote d'Ivoire, Democratic People's Republic of Korea, Democratic Republic of the Congo, Dominican Republic, Ecuador, Egypt, El Salvador, Ethiopia, Gabon, Georgia, Ghana, Guatemala, Guinea, Guinea-Bissau , Haiti, Honduras, India, Indonesia, Iran, Iraq, Jamaica, Jordan, Kazakhstan, Kenya, Kyrgyzstan, Lao People's Democratic Republic, Latvia, Lebanon, Liberia, Lithuania, Madagascar, Malawi, Malaysia, Mali, Mauritania, Mexico, Mongolia, Montenegro, Morocco, Mozambique, Myanmar, Namibia, Nepal, Nicaragua, Niger, Nigeria, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippine, Republic of Moldova, Romania, Russian Federation, Rwanda, Senegal, Serbia, Sierra Leone, South Africa, Sri Lanka, Sudan, Suriname, Swaziland, Syrian Arab Republic, T F Y R of Macedonia, Tajikistan, Thailand, Togo, Tunisia, Turkey, Turkmenistan, Uganda, Ukraine, United Republic of Tanzania, Uruguay, Uzbekistan, Venezuela, Viet Nam, Yemen, Zambia, Zimbabwe.

- **High-income countries:** Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Japan, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Republic of Korea, Spain, Sweden, Switzerland, United Kingdom, United States of America.

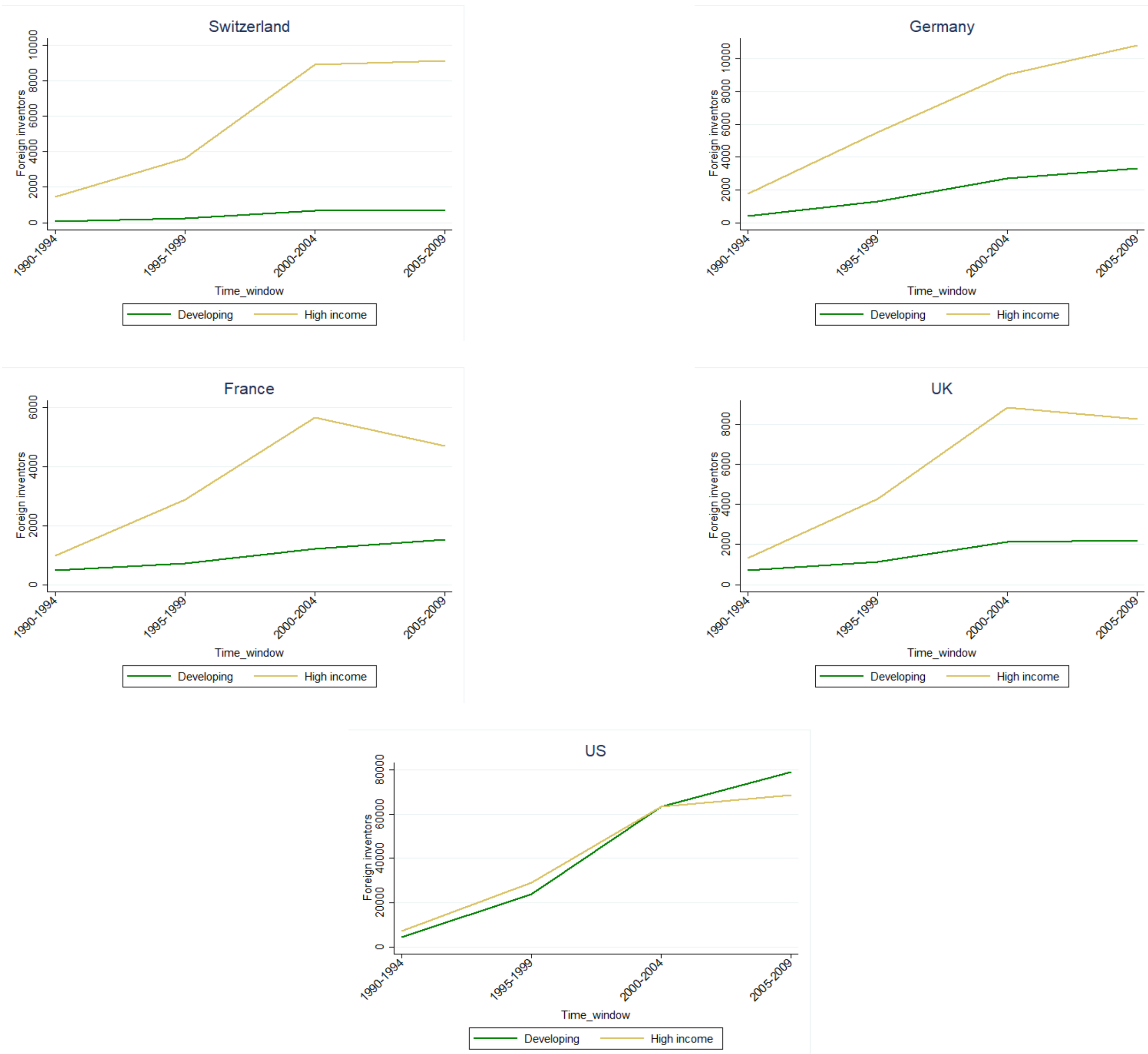
Appendix 2 Additional descriptive statistics

Figure A1: Migrant inventors stocks



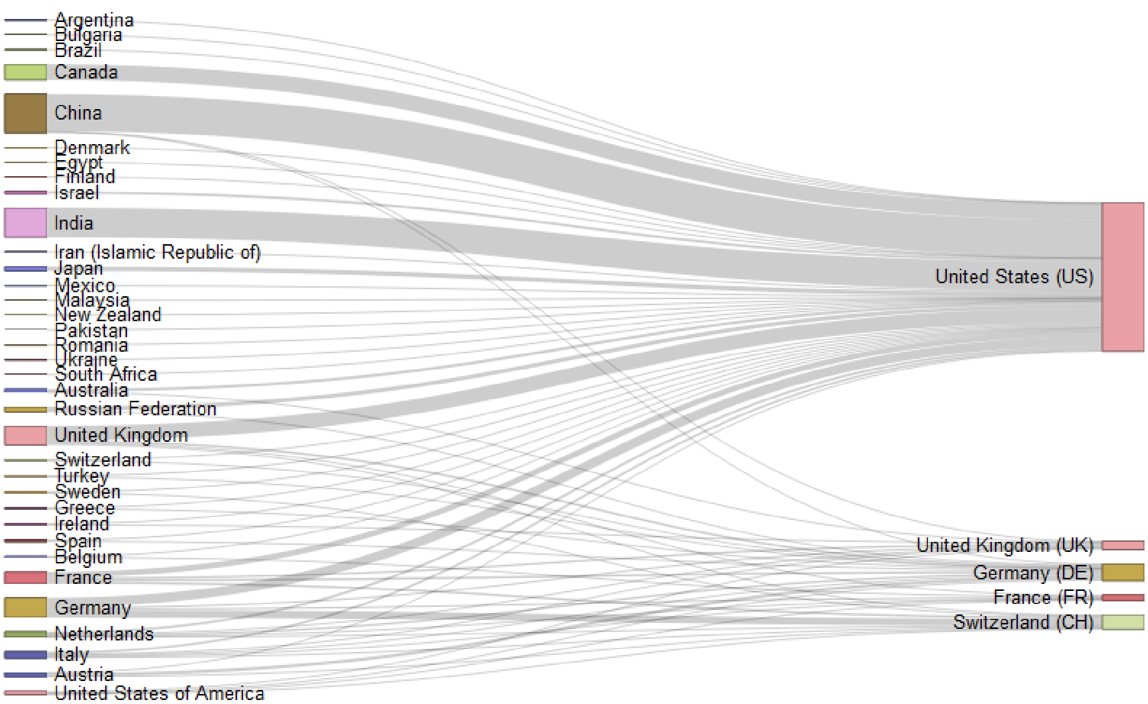
Source: Author's calculations based on Miguelez and Fink (2017) data.

Figure A2: Trends over time



Source: Author's calculations based on Miguelez and Fink (2017) data.

Figure A3: Top cross-country migration corridors



Source: Author's calculations based on Miguelez and Fink (2017) data.

Table A1: Top-20 MSA with more migrants and top-20 MSA with largest share of migrants, 2000 - 2009

Tot migrant inventors		**Share of migrant inventors**	
San José, CA	19123	San Diego, CA	0.352
San Diego, CA	18870	Evansville-Henderson, IN	0.323
Boston-Worcester, MA	15247	San José, CA	0.313
San Francisco, CA	8515	Trenton, NJ	0.294
Oakland, CA	6255	Champaign-Urbana, IL	0.288
Chicago, IL	5565	Middlesex-Somerset, NJ	0.287
New York-Newark, NY-NJ-PA	4983	New Albany-Schenectady-Troy, NY	0.284
Middlesex-Somerset, NJ	4463	Gainesville, FL	0.254
Philadelphia, PA-NJ	4427	Yolo, CA	0.254
Houston, TX	4175	Dallas, TX	0.254
New Haven-Bridgeport, CT	4129	New Haven-Bridgeport, CT	0.252
Newark, NJ	3900	New York-Newark, NY-NJ-PA	0.245
Los Angeles-Long Beach, CA	3857	Dutchess County, NY	0.239
Washington, DC-MD-VA-WV	3649	Santa Barbara - Santa Maria, CA	0.233
Raleigh-Durham-Chapel Hill, NC	3127	San Francisco, CA	0.228
Dallas, TX	2862	Oakland, CA	0.226
Minneapolis-St. Paul, MN-WI	2622	Ann Harbor, MI	0.213
Seattle-Bellevue-Everett, WA	2564	Boston-Worcester, MA	0.211
Trenton, NJ	2117	State College, PA	0.210
Orange County, CA	2059	Bergen-Pessaic, NJ	0.206

Source: Author's calculations based on Miguelez and Fink (2017) data and OECD REGPAT database.
Notes: In the second column we only include MSAs with at least 1000 inventors.

Table A2: Top-20 metro regions with more migrants, and top-20 metro-regions with largest share of migrants, 2000 - 2009

Tot migrant inventors		**Share of migrant inventors**	
London, UK	6936	Genéve, CH	0.479
Basel, CH	4251	Lausanne, CH	0.450
Paris, FR	4222	Zürich, CH	0.385
Zurich, CH	3526	Basel	0.330
Lausanne, CH	2595	Southampton, UK	0.0.221
München, DE	2474	London, UK	0.209
Mannheim-Lüdwigshafen, DE	2371	Edinburgh, UK	0.182
Cambridge, UK	1882	Cambridge, UK	0.176
Frankfurt Am Main, DE	1493	Sheffield, UK	0.159
Stuttgart, DE	1322	Bern, CH	0.142
Heidelberg, DE	1147	Birmingham, UK	0.138
Düsseldorf, UK	1108	Mulhouse, FR	0.132
Berlin, DE	971	Strasbourg, FR	0.127
Genéve, CH	916	Newcastle Upon Tyne, UK	0.113
Köln, DE	739	Reading, UK	0.113
Nürnberg, DE	696	Mannheim-Lüdwigshafen, DE	0.105
Ruhrgebiet, DE	637	Glasgow, UK	0.103
Reading, UK	545	Aberdeen, UK	0.102
Lyon, FR	531	Halle an der Saale, DE	0.096
Grenoble, FR	526	Nice, FR	0.096

Source: Author's calculations based on Miguelez and Fink (2017) data and OECD REGPAT database. **Notes:** In the second column we only include metropolitan areas with at least 1000 inventors.

Appendix 3 Additional computations

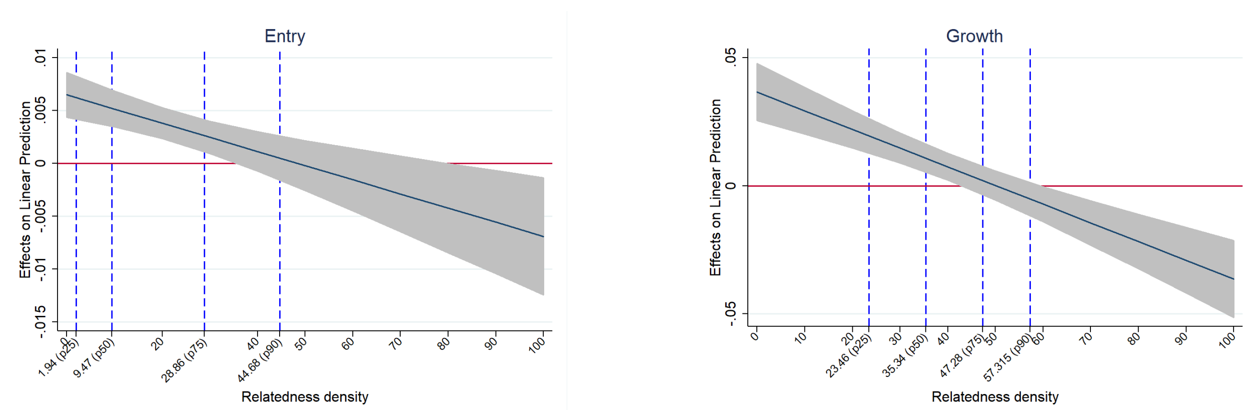
Figure A4 plots the marginal effect of *Migration* for different levels of relatedness density highlighting the 25 percentile, the median, and the 75 percentile, and showing that the effect of doubling the stock of migrants is positive and significant for the 75% of our observations, and either positive or non-significant for the 90% of our observations. Concerning *Growth*, doubling the stock of migrants increases the number of patents filed by 3.6% when relatedness density is equal to 0, 1.1% when it's equal to the mean. Also in this case we have a negative value when relatedness density is equal to 100, and Figure A4 shows that the effect of Migration is either positive or non-significant for 90% of our observations.

Table A3: Marginal effect of Migration for different levels of relatedness

		Entry	
	Pooled	Developing	High-income
Rel_dens = 0	0.006***	0.007***	0.005
Rel_dens = Mean	0.004***	0.005***	0.002
Rel_dens = 100	-0.006**	-0.010***	-0.003
		Growth	
	Pooled	Developing	High-income
Rel_dens = 0	0.036***	0.051***	0.019*
Rel_dens = Mean	0.011***	0.018***	0.008**
Rel_dens = 100	-0.036***	-0.056***	-0.007

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure A4: Interaction migration and relatedness



Notes: This figure plots the estimated β_3 from equation 1 and their corresponding 95% confidence intervals.

Appendix 4 Including additional controls

Table A4: Estimations including trade and FDI

	(Growth)			(Entry)		
	Pooled	High-income	Developing	Pooled	High-income	Developing
	(1)	(2)	(3)	(4)	(5)	(6)
Migration	0.0368***	0.0186**	0.0516***	0.00663***	0.00501	0.00705***
	(0.00576)	(0.00720)	(0.00698)	(0.00108)	(0.00460)	(0.00117)
Rel_dens	0.00887***	0.00555***	0.0106***	0.00173***	0.00128**	0.00186***
	(0.000690)	(0.00102)	(0.000793)	(0.000183)	(0.000406)	(0.000270)
Mig*rel	-0.000749***	-0.000254	-0.00109***	-0.000195***	-0.0000831	-0.000195***
	(0.000120)	(0.000149)	(0.000175)	(0.0000309)	(0.0000854)	(0.0000467)
Trade	0.000788	0.0000398	0.00128	-0.000103	-0.0000497	-0.000176
	(0.000652)	(0.000577)	(0.00125)	(0.000239)	(0.000786)	(0.000228)
FDI	-0.000416	0.00153	-0.00171	-0.00120	-0.000366	-0.000816
	(0.00178)	(0.00154)	(0.00477)	(0.000952)	(0.00195)	(0.00117)
Tot_pat	-0.129***	-0.118***	-0.139***			
	(0.00630)	(0.00895)	(0.00847)			
RTA				-0.000200***	-0.00442**	-0.000158***
				(0.0000543)	(0.00127)	(0.0000445)
Country per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	93494	51489	41967	318000	63600	254400
R-squared	0.445	0.325	0.487	0.0645	0.0708	0.0807

Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All RHS variables are lagged of one time window. Migration, Tot_pat, Copatents, Trade, and FDI are transformed using the inverse hyperbolic sine.

Appendix 5 Correlation matrix

Table A5: Correlation matrix

Variables	Migration	Rel_dens	Copatents	Tot_pat	RTA
Migration	1.000				
Rel_dens	0.151	1.000			
Copatents	0.166	0.0713	1.000		
Tot_pat	0.115	0.097	0.252	1.000	
RTA	0.001	0.018	0.000	0.001	1.000

Appendix 6 Alternative definition of Entry

Table A6: Alternative definition of entry

	(Pooled)	(Developing)	(High-income)
	(1)	(2)	(3)
	Entry	Entry	Entry
Migration	0.0011	0.0023**	0.0014
	(0.0009)	(0.0011)	(0.0016)
Rel_dens	0.0005***	0.0010***	0.0001
	(0.0001)	(0.0003)	(0.0001)
Mig*rel	-0.0000**	-0.0000***	-0.0000
	(0.0000)	(0.0000)	(0.0000)
Copatents	-0.0199***	-0.0305***	-0.0013
	(0.0024)	(0.0067)	(0.0028)
RTA	-0.0002***	-0.0001***	-0.0030**
	(0.0000)	(0.0000)	(0.0008)
Observations	93494	318000	93494
R-squared	0.4450	0.0661	0.4444

Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All RHS variables are lagged of one time window. Migration and Copatents are transformed using the inverse hyperbolic sine transformation.

Appendix 7 Estimations with RHS at the country level

Table A7: RHS at the country level

	(Growth)			(Entry)		
	Pooled	High-income	Developing	Pooled	High-income	Developing
	(1)	(2)	(3)	(4)	(5)	(6)
Migration	0.0074*	0.0096*	0.0036	-0.0003	-0.0004	-0.0019
	(0.0027)	(0.0046)	(0.0025)	(0.0008)	(0.0008)	(0.0029)
Rel_dens	0.0058***	0.0068***	0.0046***	0.0011***	0.0014***	0.0010***
	(0.0005)	(0.0009)	(0.0005)	(0.0002)	(0.0003)	(0.0002)
Mig*rel	-0.0001*	-0.0002	-0.0001	-0.0000	0.0000	0.0000
	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0001)
Copatents	-0.0104**	0.0034	-0.0017	-0.0216***	-0.0498***	-0.0015
	(0.0039)	(0.0087)	(0.0045)	(0.0020)	(0.0097)	(0.0025)
Tot_pat	-0.1272***	-0.1399***	-0.1197***			
	(0.0065)	(0.0085)	(0.0093)			
RTA				-0.0002***	-0.0002***	-0.0044**
				(0.0001)	(0.0000)	(0.0013)
Country per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	93494	51489	41967	318000	63600	254400
R-squared	0.4385	0.4839	0.3134	0.0659	0.0806	0.0739

Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All RHS variables are lagged of one time window. Migration, Tot_pat, and Copatents are transformed using the inverse hyperbolic sine.

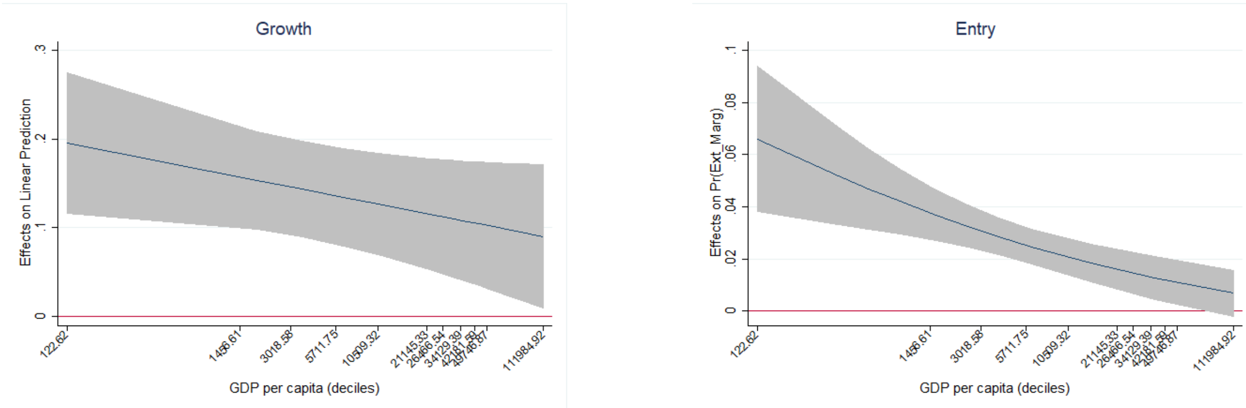
Appendix 8 Interactions with GDP

In this section we expand our analysis on differences across levels of development by interacting our main variables with real GDP per capita. To introduce this new variable we use data from the World Bank. Figure A5 plots the estimation results of the following equation where we add the interaction between $Migration$ and GDP_pc:

$$\begin{aligned} Y_{c,t,tw} = & \alpha + \beta_1 Migration_{c,t,tw-1} + \beta_2 Rel\_dens_{c,t,tw-1} + \beta_3 GPD\_pc_{c,tw-1} + \beta_4 Controls_{c,t,tw-1} \\ & + \beta_5 Migration_{c,t,tw-1} * Rel\_dens_{c,t,tw-1} + \beta_6 Migration_{c,t,tw-1} * GPD\_pc_{c,tw-1} + \\ & \gamma_c + \delta_{t,tw} + \epsilon_{c,t,tw} \end{aligned} \tag{10}$$

Figure A5 show that the effect of the interaction between GDP per capita and $Migration$ on $Entry$ is at maximum for the first income decile. Although the confidence interval suggests that the estimations are not very precise, it is worth noticing that the lower band is well above zero, and above to the lower band of the 4th and 5th income decile. When moving to higher levels of income, the average effect gets close to zero and becomes not significant between the 9th and 10th decile. This confirms our conclusions, in which we argue that having a high-skilled diaspora in specialized areas is particularly beneficial for developing countries, less for high-income countries. We are aware that the effect is always positive and significant for the $Growth$ equation, and only not-significant at the highest deciles for the entry equation. However, we attribute these results to the absence of country*time FEs, so these graphs are not fully comparable to our baseline regressions.

Figure A5: Interaction migration and income



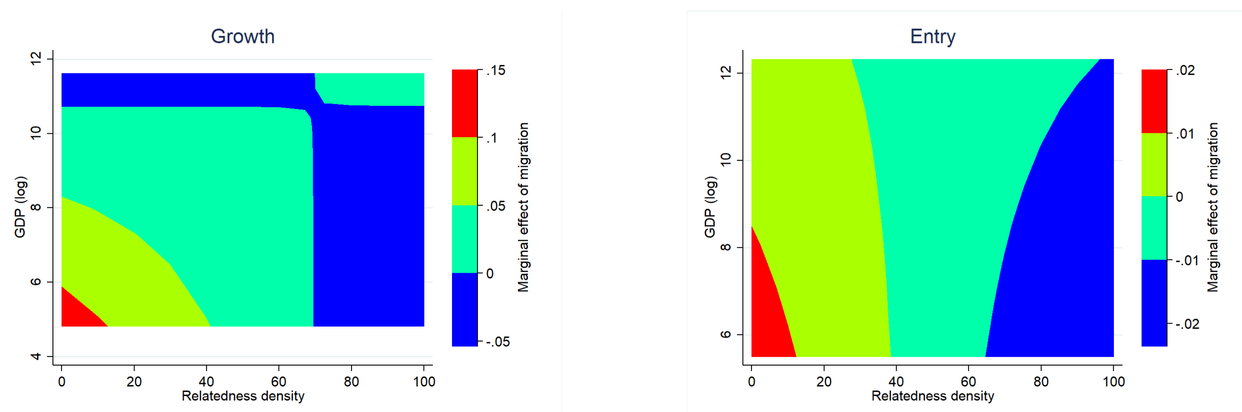
Notes: This figure plots the estimated β_6 from equation 10 and their corresponding 95% confidence intervals.

Then, we re-estimate the equation by introducing a triple interaction between $Migration$, GDP_pc, and Rel_dens:

$$\begin{aligned} Y_{c,t,tw} =& \alpha + \beta_1 Migration_{c,t,tw-1} + \beta_2 Rel\_dens_{c,t,tw-1} + \beta_3 GPD\_pc_{c,tw-1} + \beta_4 Controls_{c,t,tw-1} \\ &+ \beta_5 Migration_{c,t,tw-1} * GPD\_pc_{c,tw-1} * Rel\_dens_{c,t,tw-1} + \\ &\beta_6 Migration_{c,t,tw-1} * Rel\_dens_{c,t,tw-1} + \gamma_c + \delta_{t,tw} + \epsilon_{c,t,tw} \end{aligned} \tag{11}$$

The plot for *Growth* shows that migration has the highest marginal effect in the left corner of the plot, for low levels of income and relatedness. The plot for *Entry* shows the same thing, and also has lots of regions in which the marginal effect of $Migration$ is, on average, null or negative. However, it's worth noticing that the average level of relatedness in our sample is 16.50 with a standard deviation of 17.93. As a consequence the majority of our observations will fall between the second and third band (and this explains the average coefficients we get from the original estimations).

Figure A6: Interaction migration, income, and relatedness



Notes: This figure plots the estimated β_5 from equation 11 and their corresponding 95% confidence intervals.

Appendix 9 IV First stage

Table A8: IV First stage - Migration

	(Migration - Growth sample)			(Migration - Entry sample)		
	(Pooled sample)	(Developing)	(High income)	(Pooled sample)	(Developing)	(High income)
	(1)	(2)	(3)	(4)	(5)	6)
IV	0.1539***	0.1594***	0.1415***	0.1879***	0.1950***	0.1379***
	(0.0177)	(0.0253)	(0.0193)	(0.0087)	(0.0104)	(0.0180)
IV*rel	0.0015***	0.0015***	0.0016***	0.0020***	0.0021***	0.0016***
	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0003)	(0.0002)
Rel_dens	-0.0242***	-0.0188***	-0.0281***	-0.0234***	-0.0155***	-0.0228***
	(0.0022)	(0.0025)	(0.0034)	(0.0018)	(0.0020)	(0.0025)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	93494	41967	51489	318000	254400	63600
R-squared	0.9074	0.9008	0.8979	0.8789	0.8159	0.8954
	(Mig*rel - Growth sample)			(Mig*rel - Entry sample)		
	(Pooled sample)	(Developing)	(High income)	(Pooled sample)	(Developing)	(High income)
	(1)	(2)	(3)	(4)	(5)	6)
IV	-0.6875	-0.4841	-1.2073	-0.2283	-0.0625	-0.7200
	(0.6277)	(0.9103)	(0.8443)	(0.1989)	(0.2261)	(0.7284)
IV*rel	0.1425***	0.1408***	0.1486***	0.1563***	0.1625***	0.1492***
	(0.0048)	(0.0076)	(0.0068)	(0.0042)	(0.0079)	(0.0066)
Rel_dens	1.8142***	1.6083***	1.7539***	1.3168***	1.0178***	1.7279***
	(0.1360)	(0.2587)	(0.1479)	(0.1018)	(0.1565)	(0.1269)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	93494	41967	51489	318000	254400	63600
R-squared	0.9611	0.9491	0.9667	0.9605	0.9369	0.9617

Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Estimations reported in Column 1, 2, and 3 include Tot_pat and Copatents as controls, while estimations in Column 4, 5, and 6 include RTA and Copatents. All RHS variables are lagged of one time window. Migration, the IV, Tot_pat, and Copatents are transformed using the inverse hyperbolic sine transformation.

Appendix 10 Estimation of the gravity model

Table A9: Gravity estimation

	(1)
	Migration
Mig_less_skilled*share_pop	0.2716***
	(0.0264)
Distance	-0.0001***
	(0.0000)
Contiguity	0.6491***
	(0.1721)
Common language	0.3883***
	(0.0756)
Common religion	-1.2482***
	(0.2848)
Colony	-0.2705**
	(0.0880)
Metropolitan area FE	Yes
Country of origin FE	Yes
Time window FE	Yes
Observations	34421
Log pseudolikelihood	-386686.0783

Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Mig_less_skilled*share_pop is transformed using the inverse hyperbolic sine.

Appendix 11 International inventions

Table A10: Pooled sample (international inventions)

	Growth			Entry		
	(1)	(2)	(3)	(4)	(5)	(6)
Migration	0.00957***	0.0374***	0.0370***	0.000115	0.00509***	0.00493***
	(0.00280)	(0.00588)	(0.00579)	(0.000870)	(0.000963)	(0.000960)
Rel_dens	0.00515***	0.00861***	0.00833***	0.000485**	0.00112***	0.000836***
	(0.000441)	(0.000750)	(0.000765)	(0.000184)	(0.000178)	(0.000182)
Mig*rel		-0.000779***	-0.000745***		-0.000195***	-0.000120***
		(0.000114)	(0.000117)		(0.0000301)	(0.0000338)
Copatents			-0.0159***			-0.0268***
			(0.00421)			(0.00166)
Tot_pat	-0.115***	-0.114***	-0.110***			
	(0.00666)	(0.00654)	(0.00678)			
RTA				-0.000208***	-0.000209***	-0.000205***
				(0.0000562)	(0.0000566)	(0.0000556)
Country per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	79596	79596	79596	318000	318000	318000
R-squared	0.444	0.446	0.447	0.0703	0.0706	0.0730

Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All RHS variables are lagged of one time window. Migration, Tot_pat, and Copatents are transformed using the inverse hyperbolic sine.

Table A11: Developing and high-income (international inventions)

			Growth			
		High-income			Developing	
	(1)	(2)	(3)	(4)	(5)	(6)
Migration Migration	0.00847**	0.0211**	0.0210**	0.0136**	0.0522***	0.0521***
	(0.00317)	(0.00677)	(0.00677)	(0.00588)	(0.00800)	(0.00814)
Rel_dens	0.00401***	0.00562***	0.00558***	0.00618***	0.0108***	0.0106***
	(0.000459)	(0.000983)	(0.000992)	(0.000904)	(0.000861)	(0.000953)
Mig*rel		-0.000342**	-0.000337**		-0.00116***	-0.00114***
		(0.000133)	(0.000134)		(0.000153)	(0.000168)
Copatents			-0.00282			-0.0242*
			(0.00444)			(0.0142)
Tot_pat	-0.105***	-0.106***	-0.105***	-0.115***	-0.113***	-0.112***
	(0.00932)	(0.00947)	(0.00946)	(0.00992)	(0.00986)	(0.0102)
Country per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	49348	49348	49348	30141	30141	30141
R-squared	0.298	0.299	0.299	0.511	0.514	0.515
			Entry			
		High-income			Developing	
	(1)	(2)	(3)	(4)	(5)	(6)
Migration Migration	0.00245	0.00519	0.00498	0.00183*	0.00598***	0.00596***
	(0.00212)	(0.00350)	(0.00343)	(0.000996)	(0.00101)	(0.00102)
Rel_dens	0.000479**	0.000797**	0.000761**	0.000516	0.00106***	0.000991***
	(0.000222)	(0.000260)	(0.000248)	(0.000329)	(0.000262)	(0.000274)
Mig*rel		-0.0000761	-0.0000627		-0.000216***	-0.000200***
		(0.0000597)	(0.0000586)		(0.0000494)	(0.0000461)
Copatents			-0.00654**			-0.0295***
			(0.00270)			(0.00586)
RTA	-0.00473**	-0.00473**	-0.00467**	-0.000161***	-0.000162***	-0.000161***
	(0.00136)	(0.00136)	(0.00136)	(0.0000456)	(0.0000458)	(0.0000457)
Country per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	63600	63600	63600	254400	254400	254400
R-squared	0.0676	0.0677	0.0679	0.0907	0.0911	0.0914

Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All RHS variables are lagged of one time window. Migration, Tot_pat, and Copatents are transformed using the inverse hyperbolic sine.

Table A12: IV estimations - international inventions

	(Pooled sample)		(Developing)		(High income)	
	Growth	Entry	Growth	Entry	Growth	Entry
	(1)	(2)	(3)	(4)	(5)	6)
Migration	0.0566***	0.0034	0.0425**	0.0085**	0.0701**	-0.0083
	(0.0123)	(0.0029)	(0.0156)	(0.0026)	(0.0262)	(0.0185)
Mig*rel	-0.0010***	-0.0001	-0.0010***	-0.0002***	-0.0011*	0.0001
	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0004)	(0.0003)
Rel_dens	0.0095***	0.0008**	0.0102***	0.0011***	0.0089***	-0.0000
	(0.0010)	(0.0003)	(0.0012)	(0.0003)	(0.0022)	(0.0011)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	79613	318000	30261	254400	49352	63600
R-squared	0.4469	0.0730	0.5169	0.0914	0.2929	0.0675
P-value underid test	0.000	0.000	0.000	0.000	0.000	0.000
Kleibergen-Paap statistic	98.215	460.413	52.482	351.053	44.851	66.456

Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Estimations reported in Column 1, 3, and 5 include Tot_pat and Copatents as controls, while estimations in Column 2, 4, and 6 include RTA and Copatents. All RHS variables are lagged of one time window. Migration, the IV, Tot_pat, and Copatents are transformed using the inverse hyperbolic sine transformation.

Appendix 12 Excluding outliers

Table A13: Estimations without China

	(Growth)		(Entry)	
	Pooled	Developing	Pooled	Developing
	(1)	(2)	(3)	(4)
Migration	0.0345***	0.0458***	0.00599***	0.00618***
	(0.00545)	(0.00679)	(0.000996)	(0.000999)
Rel_dens	0.00861***	0.0102***	0.00148***	0.00181***
	(0.000710)	(0.000858)	(0.000194)	(0.000276)
Mig*rel	-0.000664***	-0.000871***	-0.000113***	-0.000135**
	(0.000110)	(0.000130)	(0.0000321)	(0.0000401)
Copatents	-0.0117***	-0.0255**	-0.0219***	-0.0285***
	(0.00339)	(0.0125)	(0.00176)	(0.00723)
Tot_pat	-0.126***	-0.137***		
	(0.00656)	(0.00933)		
RTA			-0.000197***	-0.000157***
			(0.0000535)	(0.0000445)
Country per time FE	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes
Observations	90991	39443	315456	251856
R-squared	0.442	0.480	0.0669	0.0823

Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All RHS variables are lagged of one time window. Migration, Tot_pat, and Copatents are transformed using the inverse hyperbolic sine.

Table A14: Estimations without India

	(Growth)		(Entry)	
	Pooled	Developing	Pooled	Developing
	(1)	(2)	(3)	(4)
Migration	0.0349***	0.0484***	0.00648***	0.00705***
	(0.00565)	(0.00769)	(0.00111)	(0.00123)
Rel_dens	0.00855***	0.0102***	0.00148***	0.00178***
	(0.000700)	(0.000854)	(0.000193)	(0.000272)
Mig*rel	-0.000712***	-0.00105***	-0.000136***	-0.000189***
	(0.000126)	(0.000216)	(0.0000363)	(0.0000494)
Copatents	-0.0103**	-0.0221*	-0.0219***	-0.0283***
	(0.00353)	(0.0122)	(0.00176)	(0.00744)
Tot_pat	-0.124***	-0.134***		
	(0.00622)	(0.00830)		
RTA			-0.000196***	-0.000156***
			(0.0000532)	(0.0000441)
Country per time FE	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes
Observations	91859	40332	315456	251856
R-squared	0.444	0.483	0.0666	0.0821

Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All RHS variables are lagged of one time window. Migration, Tot_pat, and Copatents are transformed using the inverse hyperbolic sine.

Table A15: Estimations without US as destination country

	Growth			Entry		
	(Pooled)	(High-income)	(Developing)	(Pooled)	(High-income)	(Developing)
Migration	0.0250***	0.0126**	0.0418***	0.00625***	0.00436	0.00777***
	(0.00508)	(0.00525)	(0.00887)	(0.00116)	(0.00305)	(0.00151)
Rel_dens	0.00745***	0.00520***	0.00904***	0.00150***	0.00126***	0.00179***
	(0.000596)	(0.000804)	(0.000709)	(0.000183)	(0.000301)	(0.000268)
Mig*rel	-0.000582***	-0.000222*	-0.00102***	-0.000186***	-0.0000972	-0.000280***
	(0.000121)	(0.000128)	(0.000238)	(0.0000395)	(0.0000617)	(0.0000561)
Copatents	-0.0109**	0.000607	-0.0232*	-0.0213***	-0.00350	-0.0276***
	(0.00342)	(0.00383)	(0.0122)	(0.00181)	(0.00262)	(0.00769)
Tot_pat	-0.126***	-0.118***	-0.138***			
	(0.00644)	(0.00886)	(0.00870)			
RTA				-0.000197***	-0.00438**	-0.000157***
				(0.0000535)	(0.00128)	(0.0000444)
Country per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	93494	51489	41967	318000	63600	254400
R-squared	0.444	0.325	0.486	0.0661	0.0709	0.0810

Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All RHS variables are lagged of one time window. Migration, Tot_pat, and Copatents are transformed using the inverse hyperbolic sine.

Appendix 13 Alternative dependent variables

Table A16: Excluding collaborations

	Growth			Entry		
	(Pooled)	(High-income)	(Developing)	(Pooled)	(High-income)	(Developing)
Migration	0.0429***	0.0198**	0.0802***	0.00466***	0.00613	0.00428***
	(0.0104)	(0.00838)	(0.0159)	(0.00109)	(0.00424)	(0.00124)
Rel_dens	0.00925***	0.00597***	0.0135***	0.00105***	0.00129**	0.00135***
	(0.00119)	(0.00102)	(0.00206)	(0.000182)	(0.000383)	(0.000247)
Mig*rel	-0.000596**	-0.000159	-0.00111***	-0.0000338	-0.0000948	-0.0000283
	(0.000183)	(0.000165)	(0.000317)	(0.0000348)	(0.0000795)	(0.0000528)
Copatents	0.0316***	0.0226**	0.0496*	-0.0152***	-0.00286	-0.0137**
	(0.00865)	(0.00913)	(0.0253)	(0.00151)	(0.00268)	(0.00581)
Tot_pat	-0.246***	-0.214***	-0.333***			
	(0.0187)	(0.0198)	(0.0347)			
RTA				-0.000158***	-0.00400**	-0.000122***
				(0.0000441)	(0.00125)	(0.0000360)
Country per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	73122	47589	25429	318000	63600	254400
R-squared	0.298	0.272	0.367	0.0735	0.0714	0.0898

Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All RHS variables are lagged of one time window. Migration, Tot_pat, and Copatents are transformed using the inverse hyperbolic sine.

Table A17: Excluding PCT applications

	Growth			Entry		
	(Pooled)	(High-income)	(Developing)	(Pooled)	(High-income)	(Developing)
Migration	0.0400***	0.0225**	0.0605***	0.00618***	0.00253	0.00657***
	(0.00645)	(0.00903)	(0.00654)	(0.00114)	(0.00369)	(0.00140)
Rel_dens	0.00890***	0.00645***	0.0104***	0.00132***	0.00101**	0.00149***
	(0.000791)	(0.00118)	(0.000805)	(0.000175)	(0.000363)	(0.000223)
Mig*rel	-0.000762***	-0.000334*	-0.00119***	-0.0000540	-0.0000124	-0.0000640
	(0.000135)	(0.000171)	(0.000165)	(0.0000404)	(0.0000769)	(0.0000680)
Copatents	-0.00936*	-0.000469	-0.0220	-0.0168***	-0.00786**	-0.0167**
	(0.00539)	(0.00685)	(0.0136)	(0.00200)	(0.00321)	(0.00651)
Tot_pat	-0.118***	-0.113***	-0.135***			
	(0.00769)	(0.00919)	(0.0118)			
RTA				-0.000106**	-0.00250**	-0.0000840**
				(0.0000324)	(0.000809)	(0.0000279)
Country per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	84124	47338	36745	318000	63600	254400
R-squared	0.377	0.262	0.453	0.0753	0.0694	0.0875

Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All RHS variables are lagged of one time window. Migration, Tot_pat, and Copatents are transformed using the inverse hyperbolic sine.

Appendix 14 Excluding zero cells

Table A18: Estimations excluding zero cells

	(Growth)			(Entry)		
	Pooled	High-income	Developing	Pooled	High-income	Developing
	(1)	(2)	(3)	(4)	(5)	(6)
Migration	0.0268***	0.0121*	0.0421***	0.0069***	0.0027	0.0096***
	(0.0046)	(0.0051)	(0.0075)	(0.0018)	(0.0036)	(0.0022)
Rel_dens	0.0088***	0.0051***	0.0122***	0.0017***	0.0012**	0.0021***
	(0.0007)	(0.0007)	(0.0010)	(0.0003)	(0.0004)	(0.0004)
Mig*rel	-0.0006***	-0.0002	-0.0012***	-0.0002**	-0.0000	-0.0003***
	(0.0001)	(0.0001)	(0.0002)	(0.0000)	(0.0001)	(0.0001)
Copatents	-0.0079*	0.0003	-0.0160	-0.0185***	-0.0006	-0.0245***
	(0.0035)	(0.0033)	(0.0103)	(0.0025)	(0.0032)	(0.0060)
Tot_pat	-0.1395***	-0.1071***	-0.1731***			
	(0.0089)	(0.0099)	(0.0113)			
RTA				-0.0005	-0.0070***	-0.0004
				(0.0003)	(0.0017)	(0.0002)
Country per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	47528	27170	20131	93402	31352	61791
R-squared	0.3741	0.3070	0.4369	0.0565	0.0901	0.0727

Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All RHS variables are lagged of one time window. Migration, Tot_pat, and Copatents are transformed using a regular logarithmic transformation.

Appendix 15 Migration at the ipc4 level

Table A19: Migration at the ipc4 level

	(Growth)			(Entry)		
	Pooled	High-income	Developing	Pooled	High-income	Developing
	(1)	(2)	(3)	(4)	(5)	(6)
Migration	0.0412***	0.0151*	0.0730***	0.00743	0.00447	0.0168**
	(0.00819)	(0.00850)	(0.0137)	(0.00462)	(0.00545)	(0.00723)
Rel_dens	0.00650***	0.00483***	0.00783***	0.00131***	0.00108***	0.00161***
	(0.000484)	(0.000597)	(0.000748)	(0.000180)	(0.000229)	(0.000299)
Mig*rel	-0.00129***	-0.000510**	-0.00203***	-0.000539***	-0.000195	-0.000797***
	(0.000199)	(0.000206)	(0.000364)	(0.000101)	(0.000116)	(0.000178)
Copatents	-0.00757**	0.00150	-0.0163*	-0.0159***	-0.00300	-0.0203**
	(0.00325)	(0.00368)	(0.00869)	(0.00197)	(0.00240)	(0.00861)
Tot_pat	-0.124***	-0.118***	-0.136***			
	(0.00640)	(0.00864)	(0.00920)			
RTA				-0.000196***	-0.00437**	-0.000157***
				(0.0000534)	(0.00128)	(0.0000444)
Country per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Tech per time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	93494	51489	41967	318000	63600	254400
R-squared	0.445	0.325	0.487	0.0667	0.0709	0.0815

Notes: Standard errors clustered at the country level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All RHS variables are lagged of one time window. Migration, Number of patents, Mig_rel_tech, Mig_nrel_tech, and Copatents are transformed using the inverse hyperbolic sine.