

# Inventor Diasporas and the Internationalization of Technology

*Ernest Miguélez*



**WORLD BANK GROUP**

Development Economics Vice Presidency

Operations and Strategy Team

April 2016

## Abstract

This paper documents the influence of diaspora networks of highly-skilled individuals—that is, inventors—on international technological collaborations. Using gravity models, it studies the determinants of the internationalization of inventive activity between a group of industrialized countries and a sample of developing and emerging economies. The paper examines the influence exerted by skilled diasporas

in fostering cross-country co-inventorship as well as R&D offshoring. The study finds a strong and robust relationship between inventor diasporas and different forms of international co-patenting. However, the effect decreases with the level of formality of the interactions. Interestingly, some of the most successful diasporas recently documented—namely, Chinese and Indian ones—do not govern the results.

---

This paper is a product of the Operations and Strategy Team, Development Economics Vice Presidency. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The author may be contacted at [aternest.miguel@u-bordeaux.fr](mailto:aternest.miguel@u-bordeaux.fr).

*The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.*

# INVENTOR DIASPORAS AND THE INTERNATIONALIZATION OF TECHNOLOGY

*Ernest Miguélez*<sup>1</sup>

Key words: diaspora networks, international collaborations, inventors, PCT patents, R&D offshoring

JEL: C8, J61, O31, O33

---

<sup>1</sup> Ernest Miguélez (corresponding author) is Junior Researcher in the CNRS, affiliated to the GREThA UMR CNRS 5113 research group, Université de Bordeaux. He is also visiting fellow at the AQR-IREA, University of Barcelona, and research affiliate at CReAM, UCL. His e-mail address is [ernest.miguelez@u-bordeaux.fr](mailto:ernest.miguelez@u-bordeaux.fr). The research for this article was financed by the Regional Council of Aquitaine (Chaire d'Accueil en Economie de l'Innovation, to Francesco Lissoni) and the Regional Council of Aquitaine (PROXIMO project, to Christophe Carrincazeaux). The author thanks Michel Beine, Gaetan de Rassenfosse, Stuart Graham, William Kerr, Francesco Lissoni, Catalina Martinez, Çağlar Ozgen, Hillel Rapoport, Massimo Riccaboni, Valerio Sterzi, seminar participants at GREThA-Bordeaux University (2013), GATE-LSE Saint Etienne (2014), Temporary Migration Cluster UC Davis (2015), the workshop "The Output of R&D and Innovative Activities: Harnessing the Power of Patent Data" (JRC-IPTS Seville, 2013), the 7th MEIDE conference (Santiago de Chile, 2013), the PSDM 2013 conference (Rio de Janeiro, 2013), the 54<sup>th</sup> ERSa conference (Saint Petersburg, Russia, 2014), and the 2<sup>nd</sup> Geography of Innovation Conference (Utrecht, The Netherlands, 2014) for valuable comments; Julio Raffo for helpful discussions on previous versions of this paper; as well as the contribution of three anonymous referees.

International innovation networks are critical stepping stones for accessing frontier knowledge from the most industrialized economies, both in the form of formal information exchanges as well as through knowledge spillovers (Hall 2011).

International spillovers are essential for developing countries to catch up with advanced economies, and this subject matter ranks high in the development policy agenda (World Bank 2010).

However, geographical and cultural barriers seriously hinder the formation of innovation networks across countries, thereby remaining primarily a national phenomenon—contrary to other features such as trade or Foreign Direct Investment (FDI) (Patel and Pavitt 1991). Guellec and van Pottelsberghe de la Potterie (2001) report that only 4.7% of EPO patents and 6.2% of USPTO patents in 1995 have at least one foreign co-inventor. Picci (2010) estimates this figure to be around 8% for European patents in 2005. Cultural and language differences seriously undermine the internationalization of inventive activity. Other features inherent to geographical separation are also relevant, such as difficulties in screening potential partners, managing and administration of common projects across borders, and differences in legal frameworks and the rule of law—especially regarding the issue of intellectual property rights (Montobbio and Sterzi 2013; Foray 1995).

This paper examines how high-skilled migration and the diasporic networks it creates may overcome cross-country barriers and foster the internationalization of inventive activity. Diaspora networks have been largely studied in the context of trade and FDI. More recent evidence has looked at the international diffusion of ideas (Agrawal et al. 2011; Kerr 2008; Oettl and Agrawal 2008) and firms' internationalization strategy (Foley and Kerr 2013; Saxenian et al. 2002). In parallel,

numerous papers have investigated the internationalization of R&D activities (Guellec and van Pottelsberghe de la Potterie 2001; Patel and Vega 1999; Picci 2010). This paper builds on and extends these two streams of literature by looking at the role of migrant inventors in fostering the internationalization of innovation activities.

Specifically, the paper looks at inventor diasporas and transnational inventive activities between developed and developing countries, which is gaining momentum from a development perspective (Clemens et al. 2014; Montobbio and Sterzi 2013). It also aims to see whether differences emerge across the type of technological internationalization—co-inventorship vs. R&D offshoring. Finally, the extent to which countries' characteristics govern these potential relations is also investigated—that is, whether the least similar countries, for which informal barriers are more acute, have the greatest potential to benefit from diaspora networks.

The study extends the existing literature in a critical way. Contrary to most migration research, which retrieves migration data from decennial censuses, it exploits inventors' information listed in patent applications. Inventors make up a specific class of workers at the upper end of the skills distribution.<sup>2</sup> Differently from former approaches to inventor migration that use ethnic name recognition algorithms to infer their migratory background (Agrawal et al. 2011; Breschi et al. 2014; Breschi et al 2015; Kerr 2008), direct nationality information of the inventors is used, allowing for the introduction of a richer range of origin and destination countries. Admittedly, the use of nationality data does not come without its limitations, the most important being

---

2. In this respect, the PatVal survey shows that 76.9% of European inventors underwent tertiary education (26% with PhD) (Giuri et al. 2007). Similarly, the RIETI-Georgia Tech inventor survey shows that 94% of US inventors have college degrees (46% with PhD), with 88% of the cases for Japanese inventors (13% with PhD) (Walsh and Nagaoka 2009).

the loss of all naturalized inventors who have acquired their host country citizenship. Appendix S3 compares nationality data with ethnic name recognition algorithms (see Breschi et al. 2015), which has the advantage of taking into account all the naturalized immigrants who have acquired their host country nationality. Although some differences emerge—large coefficients when name recognition algorithms are used—the main conclusions of this paper remain unaltered.

To anticipate the results to come, I find a robust effect of highly-skilled diasporas on the internationalization of inventive activity between developed, receiving countries and developing, sending economies: a 10% increase in the inventor diaspora abroad is associated with a 2.0–2.2% increase in international patent collaborations. The evidence found survives the inclusion of a large number of controls, fixed-effects (FE), robustness checks, and identification issues. Moreover, the effect is stronger for inventor-to-inventor collaborations—co-inventorship—than for applicant-to-inventor co-patents—R&D offshoring—suggesting that diaspora effects specifically mediate interpersonal relations between co-workers.<sup>3</sup>

The outline of the paper is as follows: section 2 reviews previous theoretical and empirical contributions on the relationship between migration and other international economic interactions. Section 3 presents the novel dataset on inventor migration flows and develops the methodological setting, including all the econometric concerns. Section 4 presents the results and section 5 concludes. The reader may also find the appendix useful, giving more details on the novel data used and robustness results.

---

3. This paper uses the term “applicant” to describe the owner of the patent, unless otherwise stated, which is normally a firm or research institution for which the inventors usually work. I am aware that in some patent jurisdictions the owner is termed “assignee,” although this term is not used here.

## I. RELATED LITERATURE AND THEORETICAL BACKGROUND

Much of the public debate and academic research on the role of skilled diasporas for development tries to answer a question fraught with political and economic significance: how will growing diasporas—and in particular skilled diasporas—contribute to their home country development?

In the migration literature, diasporas have been defined as “part of a people, dispersed in one or more countries other than its homeland, that maintains a feeling of transnational community among a people and its homeland” (Chander 2001). Potential benefits can be realized exploiting this feeling to the advantage of the home countries, through the individuals’ embedded knowledge as well as through their accessible resources—such as capital or the expatriates’ network of colleagues and acquaintances.

Traditionally linked to diasporas’ role on financial remittances and capital formation, the migration and development literature has also moved to investigating their role in favoring other economic transactions, such as trade (Gould 1994), FDI (Javorcik et al. 2011; Kugler and Rapoport 2007), international diffusion of ideas (Agrawal et al. 2011; Kerr 2008), and firms’ internationalization strategies (Saxenian et al. 2002; Foley and Kerr 2013). Diasporas affect their home countries both directly and indirectly (Kapur and McHale 2005). The direct effect is linked to the diaspora members’ willingness to interact individually with their home countries, in the form of remittances, investments, or sharing ideas and information. This eventually includes the role of returnees’ direct contribution. Highly-skilled migrants may decide to move back or set up entrepreneurial activities in their homelands, while keeping in touch with the destination countries (Wadhwa, Rissing, et al. 2007; Wadhwa, Saxenian, et al. 2007).

The indirect effects refer to the role of diaspora members in leveraging their home countries' reputation in international business networks; facilitating searching and matching between partners, customers-suppliers or in the labor market; and in ensuring the contract fulfillments of the two parties involved (Kapur and McHale 2005). Because of their familiarity with local market needs, diasporas provide information about business opportunities in their home countries, and thus are critical in providing access to relevant information otherwise inaccessible because of cultural, language, institutional, administrative, or geographical barriers. Thus, migrant networks lower transaction costs associated with problems of incomplete information. This is particularly the case when informational difficulties are large, when involving countries with very different social and cultural backgrounds.

They also lower the transaction costs associated with the existence of asymmetric information. As Rauch (2003; 2001) posits, social networks operating across national borders build up or substitute for trust when contract enforcement is weak or nonexistent. Indeed, diasporas create trust by establishing a kind of "moral community," which is used to transmit information about past opportunistic behavior in international business relations. The enforcement mechanism is particularly important in the absence of effective protection of contractual/property rights and should therefore be critical in the relationships of developed-developing countries.

In the trade context, Gould (1994) finds that the stock of migrants in the United States (US) from 47 US trading partners increases US trade with these countries—a pioneering contribution to the topic of migrant networks and trade is due to Greif (1989). This is confirmed by Rauch and Trindade (2002) and Head and Ries (1998), who find that a 10% increase in the number of immigrants increases exports by one



percent and imports by three percent (see also Aleksynska and Peri 2014; Felbermayr and Toubal 2010, 2012). Similar conclusions emerge in the case of FDI. Javorcik et al. (2011) investigate the link between the presence of migrants in the United States and US FDI to the migrants' countries of origin. They find that US FDI to sending countries is positively correlated with the diaspora of that country in the United States—especially migrants with college degree qualifications (see also Kugler and Rapoport 2007).

Less evidence is available on diaspora externalities and international technology cooperation. Saxenian (1999) argues that skilled immigrants in the United States are playing a growing role in linking domestic technology businesses to their countries of origin. Her study on Chinese and Indian immigrant engineers in Silicon Valley shows that these immigrants are uniquely placed to locate foreign partners quickly and manage complex business networks across cultural, institutional, and linguistic boundaries, which is especially relevant in high-tech industries. Systematic evidence for the case of migrant scientists is provided by Scellato et al. (2012) for a group of surveyed scientists from 16 countries. Their study finds that around 40% of foreign-born researchers in these countries maintain research links with their homeland colleagues.

For the specific case of inventors, Agrawal et al. (2011) study knowledge flows between India and the Indian diaspora in the United States, identified through inventors of USPTO patents. Kerr (2008) extends this analysis to nine foreign ethnicities in the United States. By means of citation analysis, he confirms that knowledge diffuses internationally through ethnic networks—especially with regards to the Chinese diaspora, which also has sizeable effects on home country output. Foley and Kerr (2013) find significant effects of US firms' ethnic inventors in promoting linkages

between these firms and their R&D staff home countries, in the form of knowledge flows or R&D alliances. As these authors argue, ethnic inventors in host countries are particularly apposite for helping firms to capitalize on foreign opportunities and overcome barriers to the internationalization of inventive activity. Ethnic inventors usually have the expertise essential for developing products crucial for that particular ethnicity, giving privileged access to foreign markets and business opportunities. Obviously, they possess the language skills and cultural sensitivity necessary to promote international collaborations in their host countries, while at the same time knowing how to conduct business with their homeland colleagues. They also belong to those networks that foster trust and convey information about past opportunistic behavior across national boundaries.

However, this latter evidence might not be conclusive, in the sense that it is limited to the largest receiving country, the United States, and its top providers of foreign scientists and engineers (Breschi et al. 2014). This leads some authors to argue that highly-skilled emigrants may not systematically engage in business networks and knowledge transfers with their homelands but rather that the Indian and Chinese diasporas are so famous in being the exception rather than the rule (Gibson and McKenzie 2012). The present paper sheds some light into this issue too.

## II. RESEARCH METHODS

### *Empirical Approach*

The gravity model to be estimated takes the following form:

$$(1) COPAT_{ijt} = e^{\beta_0} \cdot DIASPORA_{ijt}^{\beta_1} \cdot Z_{ijt}^{\beta_2} \cdot e^{\tau_i} \cdot e^{\tau_j} \cdot e^{\delta_i} \cdot \varepsilon_{ijt}$$

where  $COPAT_{ijt}$  stands for the number of collaborations between i's developing country (out of 67) and j's developed country (out of 20) for year t.<sup>4</sup>  $\beta_j$  is the parameter of interest in this work, while  $DIASPORA_{ijt}$  is the focal variable and is computed as the number of inventor nationals of country i residing in country j for annually repeated 5-year time-windows.  $Z_{ijt}$  is a set of bilateral and attribute control variables, and  $\tau_i$ ,  $\tau_j$ , and  $\delta_t$  are, respectively, developing, developed, and time FE.  $\varepsilon_{ijt}$  denotes the error term.

Log-linearizing equation (1) and using OLS techniques would be a straightforward estimation method. However, cross-country co-patents are rare phenomena, which translate into a dependent variable with a very large proportion of zeros, making the logarithmic transformation of these observations impossible. Dropping these zero observations or adding an arbitrary constant to enable logarithmic transformation would be clearly misleading (Burger et al. 2009). In addition, Santos Silva and Tenreyro (2006) show that log-linearizing equation (1) requires  $\varepsilon_{ijt}$ , and therefore  $\ln \varepsilon_{ijt}$ , to be statistically independent of the regressors, otherwise the condition for consistency of OLS would be violated. As the authors show, there is “overwhelming evidence that the error terms in the usual log linear specification are heteroskedastic” (Op. Cit., 642), making the expected value of the error term depend on one or more explanatory variables, leading to inconsistent OLS estimates. Note importantly that this kind of heteroscedasticity cannot be corrected using standard techniques, such as applying a robust covariance estimator, since it does not only affect the estimation of

---

4. Appendix S1 lists the countries used in this study. Note that some high income countries are included among the list of developing economies (e.g., Luxemburg). Removing them from this group does not alter the results.

the standard errors, but also the coefficients. Because of the presence of this heteroscedasticity in the original nonlinear gravity specification, the authors suggest estimating the multiplicative form of the model using nonlinear models, such as the Poisson pseudo-maximum likelihood (PPML), which also provides a natural way of dealing with zero co-patenting and the extreme skewness of the dependent variable, intrinsically heteroscedastic with variance increasing with the mean (Cameron and Trivedi 1998).<sup>5</sup> Thus, equation (1) is estimated by means of PPML using the fact that the conditional expectation of  $COPAT_{ijt}$  in (1) can be written as the following exponential function:

$$(2) E(COPAT_{ijt} | X_{ijt}) = \exp[\beta_0 + \beta_1 \ln DIASPORA_{ijt} + \gamma_n \ln Z_{ijt} + \tau_i + \tau_j + \delta_t] + \varepsilon_{ijt}.$$

## *Data*

### **Inventors' International Migration**

A large part of the migration literature surveyed above has made use of census-based migration datasets becoming available during the last 15 years, broken down by skills—primary, secondary, and tertiary level of education (for a recent data contribution, see Özden et al. 2011).

In contrast, the present analysis is based on a dataset of inventors with migratory backgrounds applying for PCT patent applications between 1990 and 2010. The use of inventor data for migration analysis comes with two main advantages. First, patent data

---

5. Moreover, Santos Silva and Tenreyro (2010, 2011) use simulation techniques to show that the PPML estimator is generally well-behaved in the presence of a large proportion of zeros. As a robustness check, appendix S8 replicates the main estimations using alternative count data methods for zero-inflated models. No important differences with respect to the main results and conclusions emerge.

(together with inventor information) are registered and so can be organized on a yearly basis—contrary to census data, which are collected only every 10 years. Second, the level of education attained may still differ markedly among tertiary educated workers. Tertiary education can include nonuniversity tertiary degrees, undergraduate university degrees, and postgraduate and doctorate degrees, which may not be fully comparable across different countries. Inventors on the other hand constitute a specific class of highly-skilled workers that is more homogeneous than the tertiary-educated workers as a whole. They are behind the production of new knowledge and innovation that encourage economic growth and well-being.

The seminal contribution using inventor data for migration research is due to Kerr (2007) and his successive papers, which analyze immigrants' contribution to US invention. Kerr takes inventors' names from USPTO applications and assigns them an ethnic affiliation using a commercial repository of names and surnames of US residents classified by likely country of origin<sup>6</sup>—for a more recent contribution along the same lines, see Breschi et al. (2014, 2015). Albeit extremely valuable, Kerr's contribution comes with some limitations: first, it is focused entirely on US immigration, while migration is a multifaceted phenomenon including numerous receiving countries. Second, it comprises a limited number of countries of origin—though it suffices for his analysis on skilled immigration in the United States. Finally, ethnic methods cannot distinguish between first and second generation immigrants—for example, African immigrants in Europe—or across countries belonging to the same linguistic group—for

---

6. In particular, nine broad ethnic affiliations are identified: English, European, Russian, Hispanic-Filipino, Chinese, Indian, Japanese, Korean, and Vietnamese

example, inventors from Australia, Canada, the United Kingdom, and the United States.<sup>7</sup>

Recently, the World Intellectual Property Organization (WIPO) released a new dataset on inventors of PCT applications containing not only their current country of residence but also their nationality, representing a promising data source for migration research (Migueluez and Fink 2013). The PCT is an international treaty administered by WIPO offering an advantageous route for seeking for patent protection in more than one jurisdiction (in its 148 contracting states). In general, patent rights only apply in the jurisdiction granting them, be it national (e.g., USPTO) or regional (e.g., the Africa Regional Intellectual Property Organization). To seek for patent protection in multiple countries, applicants need to apply for patents in multiple offices. One simplifying route for doing this is offered by the PCT treaty. In short, by choosing the PCT route, applicants gain additional time—typically 18 months—to decide whether to pursue patent protection internationally and, if so, in which jurisdictions. An international patent right, as such, does not exist, so the applicants still have to apply for patent protection in all countries in which they eventually seek protection. However, the additional time gained by choosing the PCT route is valuable for applicants, as witnessed by the increasing share of international patent protection going through this system (about 54% of multi-jurisdictional applications in 2010).

For the purpose of migration analysis, inventor information retrieved from PCT applications presents several advantages, such as being associated with the most valuable inventions (van Zeebroeck and van Pottelsberghe de la Potterie 2011) or the

---

7. See appendix S3 for a discussion of advantages and disadvantages of name recognition systems to identify the migratory background of inventors, alongside a number of robustness checks carried out using this alternative approach.

fact that the system applies a set of procedural rules common to all participant countries, hence eliminating the “home bias effect” introduced when working with patent data from one single office. Appendix S2 enumerates the advantages of using PCT data for economic analysis, discusses how countries make use of this system, and details the extent of overlap with other patent data sources, such as the USPTO and the EPO.

More importantly for the sake of the present analysis, PCT patent applications are the only ones recording the nationality of the inventors. The reason for that is as follows: because not all countries are PCT contracting states, only national or resident applicants of a PCT contracting state can file PCT applications. In order to verify that applicants meet at least one of the two eligibility criteria, the PCT application form asks for both nationality and residence. In parallel to this, US laws bind the applicant also to be the inventor; US laws also request the applicant to be an individual, not a firm. Thus, if a given PCT application includes the United States as a country in which the applicant has considered pursuing a patent—a so-called designated state in the application—all inventors are listed as “applicants/inventors,” and their residence and nationality information are, in principle, available.

All in all, between 1990 and 2010, the share of inventors’ records for which we can retrieve nationality and residence information is pretty high, around 80% of the cases.<sup>8</sup> Admittedly, this coverage is unevenly distributed over time—around 60–70% during the 1990s and 70–95% during the 2000s—as well as across countries—the United States (66%), Canada (81%), the Netherlands (74%), Germany (95%), the

---

8. The use of the word “record” here signifies the unique combination of “inventor name” and “application number.”

United Kingdom (92%), France (94%), Switzerland (93%), China (92%), and India (90%), among others.<sup>9</sup>

Patent data do not provide unique identifiers for inventors appearing in more than one application. Although this is a drawback, I follow Kerr (2008) and treat each record as if it were a different individual, and compute the migration variables aggregating by country pairs and moving time-windows of five years.<sup>10</sup>

Out of all records with complete information, about 5 million, around 9–10% have a migratory background, that is, residence different from their nationality. Figure 1 depicts the evolution of the share of inventors with migratory background (solid line), alongside the same figures broken down by a number of selected receiving countries/areas. As can be observed, the share of worldwide migrant inventors has steadily increased over time.<sup>11</sup> Among the most receiving countries of the world,

---

9. Compared to other highly patenting countries, the United States, Canada, and the Netherlands present relatively low coverage rates. This is because numerous applications with inventors from these three countries were applied to the USPTO before being extended internationally through the PCT system. In consequence, the United States was not mentioned in the applications as designated state and the inventors were only inventors and not applicants at the same time, which exonerates them from providing nationality information. In principle, there is no reason to believe that this less than complete coverage in these countries may bias the analysis one way or another. However, in order to address this inconsistent coverage of migration information, I repeated the analysis splitting the sample into shorter time windows and excluding some countries. No important differences arise regarding the main conclusions of the study.

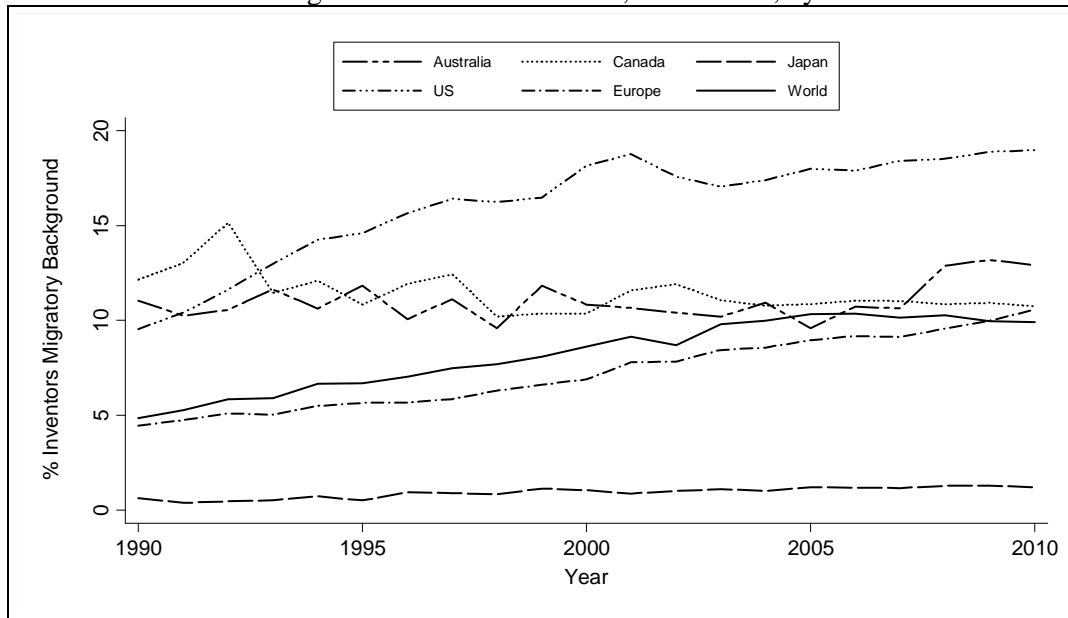
10. The priority date of applications is used to allocate individuals in time. By “priority date” I mean the first year the patent was applied worldwide.

11. In order to make these figures comparable, it is worth looking at differences with other migration datasets. While 8.62% of inventors of PCT patents have a migratory background in 2000, data compiled by Beine et al. (2007) show that general migration rates in 2000 for populations aged 25 years and over



Canada, Australia, and, notably, the United States, stand out as being the primary receiving countries, when compared to their resident stock of inventors. On the other hand, Japan is, and has been over the years, one of the developed countries with a smaller share of inventor immigrant population.

FIGURE 1. Share of Immigrant Inventors over Time, 1990–2010, by Selected Countries



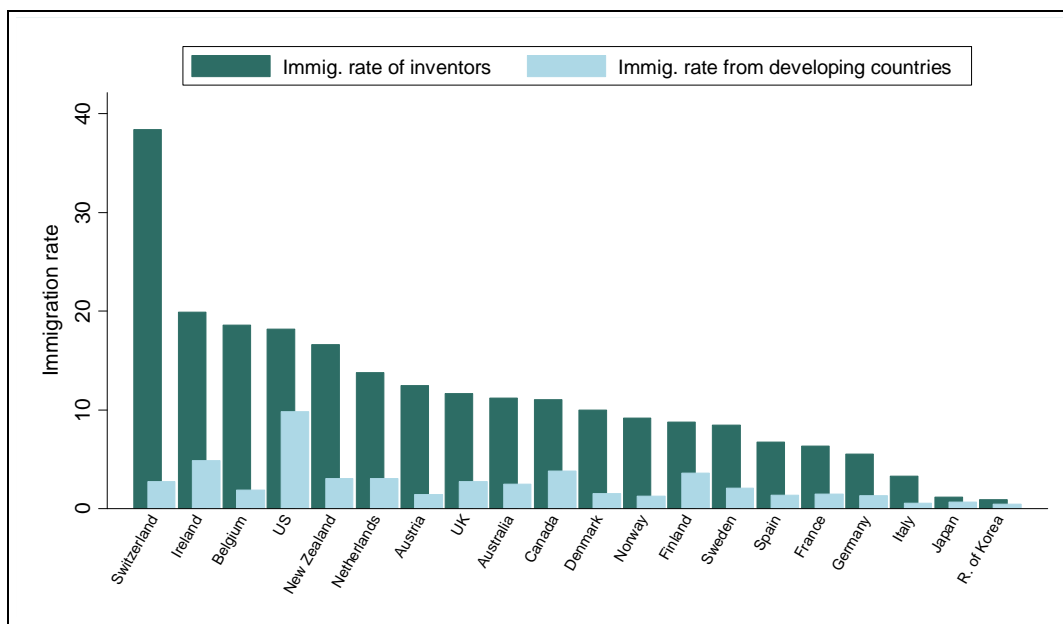
Source: Authors' analysis based on Miguélez and Fink (2013).

Meanwhile, technology-leading European countries, such as Germany or France, lag way behind compared to the United States (figure 2). The exceptional performance of the United States in attracting talent is even more notable when considering only immigrant inventors coming from low- and middle-income economies (see also figure 2). Appendix S4 shows the top 20 most populated inventor migration corridors, where again, the United States stands out as the most typical choice for destination country.

FIGURE 2. Immigration Rates of Inventors, 2001–2010, Receiving Countries

---

were estimated around 1.8%, including 1.1% of immigrants among the unskilled population, 1.8% among populations with secondary education, and 5.4% among populations with tertiary education.



Source: Authors' analysis based on Miguélez and Fink (2013).

### Dependent Variable

International co-patent data are retrieved from PCT applications (WIPO IPSTATS databases). The first focus is on co-patenting at inventor level—co-inventorship. All the co-inventions between inventors residing in country *i* and inventors residing in country *j* are added up by year. To be precise, it includes 67 developing/emerging/transition countries, on the one hand, that co-invent with 20 developed countries, on the other, where diasporas from the former countries reside. If inventors from more than two countries participate in the patent, an international co-inventor for each country-pair is counted, irrespective of the total number of countries involved in that particular invention.<sup>12</sup>

12. Results with alternative ways of computing the dependent variable, where the number of inventors per patent is taken into account, are presented in appendix S5, with no remarkable differences with respect to the results presented in the main text. In contrast, the ranking of inventors listed in patents is not taken on board, due to the difficulties of determining whether the ordering of names bears any similarity with sorting of authors in scientific publications. However, understanding the relationship between immigrant status

Next, I measure R&D offshoring using patent applications in which at least one applicant is a resident from country  $j$  (developed) and simultaneously at least one inventor is a resident from country  $i$  (developing/emerging/transition) —similar R&D offshoring measures in the context of internationalization of inventive activity are used in Guellec and van Pottelsberghe de la Potterie (2001), Harhoff et al. (2013), and Thomson et al. (2013). Again, when inventors come from various countries, a single co-patent for each bilateral  $i$ - $j$  pair is computed.

It is worth mentioning that previous studies on the determinants of international co-patenting use information from single patent offices only, with few exceptions (Martínez and Rama 2012; Picci 2010). This practice is likely to deliver biased estimates due to the “home bias effect,” which may emerge when using patent data from one single office for cross-country analysis. Since patents at the USPTO, EPO, or JPO, for instance, protect innovation within their respective geographical areas, they are preferred by domestic firms and thus their innovative capability is overestimated with respect to foreign firms. Using data from the PCT mitigates this effect.<sup>13</sup>

### **Control Variables**

Control variables include geographical, linguistic, cultural, and historical barriers to cross-country collaborations. In particular, the great circle distance between the most populated cities of countries (measured in km) is included, as well as a dummy variable indicating whether two countries share a common border, a dummy variable valued 1 if

---

and ranking in a patent is an interesting avenue of research, which goes beyond the scope of the present paper.

13. Other biases inherent to the existence of multiple jurisdictions and patent offices are discussed in de Rassenfosse et al. (2014)—such as the nonrandom choice of patent office. Again, the use of PCT applications should mitigate these biases.

the same language is spoken in both countries, and a dummy variable valued 1 when the two countries share the same colonial past—these variables come from the CEPII distance database (Mayer and Zignago 2011).

Two additional variables, suggested by Melitz and Toubal (2014), help in controlling for cultural similarities between country pairs: first, an index of language similarity. People whose languages share common roots will likely share similar cultural backgrounds. To compute this index, one single language is assigned to every country, and using information on the classification of languages provided by the Ethnologue Project, a language similarity index based on the distance between branches in this classification is computed.<sup>14</sup> Then the number of branches that coincide between each pair of languages are added together and the result is divided by the sum of branches of each of the two languages (in order to take into account the fact that the granularity of branches may not be the same across languages). As a result, an index between 0 and 1 is obtained, where 0 means complete dissimilarity and 1 means that these two languages are almost the same in linguistic terms.<sup>15</sup> Second, the religious heritage of countries is a critical element of their culture and identity (Guiso et al 2009). Countries with similar religious roots, culturally closer, are likely to interact more (for an application in the trade literature, see, among others, Melitz and Toubal 2014).

---

14. For example, the linguistic classification of Portuguese, Swedish, and Danish, from the largest, most inclusive grouping to the smallest, is: Indo-European, Italic/Romance, Italo-Western, Western, Gallo-Iberian, Ibero-Romance, West Iberian, Portuguese-Galician (Portuguese); Indo-European, Germanic, North East, Scandinavian, Danish-Swedish, Swedish (Swedish); Indo-European, Germanic, North East, Scandinavian, Danish-Swedish, Danish-Riksmal, Danish (Danish). [www.ethnologue.com](http://www.ethnologue.com), accessed May 20, 2014.

15. I arbitrarily set to 0 this variable when the countries share exactly the same language, in order to avoid collinearity with the variable “same language.”

Religion similarity is proxied with an index built as follows: for each country, data on the percentage of population adhering to one of eight major religions is retrieved (data from the CIA World Factbook dataset)—these eight religion groups differ slightly from Melitz and Toubal (2014). Then the following formula for each country pair is computed, which results in a variable ranging from 0 (no believers in common) to 1:

$$(3) \text{Religion\_Sim}_{ij} = (\%muslim_i * \%muslim_j) + (\%catholic_i * \%catholic_j) + (\%orthodox_i * \%orthodox_j) + (\%protestant_i * \%protestant_j) + (\%hinduism_i * \%hinduism_j) + (\%buddhist_i * \%buddhist_j) + (\%eastern_i * \%eastern_j) + (\%judaism_i * \%judaism_j)$$

I also control for the intensity of economic linkages between countries using the share of bilateral trade (exports plus imports, EXP+IMP) between a given pair over their total trade (COMTRADE data).<sup>16</sup> Trade is a conduit of information that may foster technological partnerships too, while it might be linked to the presence of migrants at the same time. I also account for the common technological specialization of country-pairs introducing an index of technological distance measured as

$$(4) \text{Tech.distance}_{ij} = 1 - \frac{\sum f_{ih} f_{jh}}{(\sum f_{ih}^2 \sum f_{jh}^2)^{1/2}},$$

where  $f_{ih}$  stands for the share of patents of one technological class h according to the IPC classification of country i and  $f_{jh}$  for the share of patents of one technological class h of country j. Values of the index close to the unity indicate that a given pair of countries are technologically different, and values close to zero indicate that they are technologically similar (Jaffe 1986). Again, PCT patents are used to compute this index.

---

16. Changing this variable to trade in absolute numbers does not alter the main results and conclusions, though it makes the variable EXP+IMP non-significant—results provided upon request from the author.

Finally, two additional attribute variables of individual countries are used. In particular, the number of PCT patents per country, for five-year annually repeated time-windows. This variable controls for the size of country innovation systems, which clearly determines a country's capacity to collaborate with foreigners, as well as its capacity to attract inventors from abroad or send them to other locations. In addition, it includes GDP per capita, from the World Development Indicators (World Bank), expressed in US\$ 2005 at PPP, in order to capture the market potential of countries as well as their capacity to innovate. Table 1 contains summary statistics of the variables included in the models, and appendix S6 the correlation matrix.

TABLE 1. Summary Statistics

	Observations	Mean	St. Dev	Min.	Max.
Collab. inv_i- inv_j	26,160	1.26	10.55	0	678
Collab. app_i- inv_j	26,160	1.73	14.87	0	708
Diaspora size	26,160	22.29	447.18	0	26,661
Distance	26,160	6,889.27	4,539.38	59.62	19,629.50
Contiguity	26,160	0.01	0.10	0	1
Common language	26,160	0.09	0.29	0	1
Lang. similarity	26,160	0.16	0.18	0	0.89
Colonial links	26,160	0.04	0.18	0	1
Religion similarity	26,160	0.13	0.19	0	0.90
EXP+IMP	26,160	0.01	0.03	0	0.41
Tech.distance	26,160	0.48	0.27	0.02	1
# patents_i	26,160	816.86	3,501.12	0	64,990
# patents_j	26,160	43,314.25	99,659.75	52	692,364
GDP p.c._i	26,160	9,258.23	9,566.77	432.05	74,113.90
GDP p.c._j	26,160	29,423.23	6,171.54	11,382.60	48,799.70

Source: Authors' analysis based on data described in the text.

Notes: “\_i” and “\_j” stand for, respectively, migrant inventor's sending country and migrant inventor's receiving country. Per capita GDP at origin presents several missing observations: for 1990, missing data correspond to Azerbaijan, Eritrea, Cambodia, and Latvia; for 1991, to Eritrea; and for 2005, to Cyprus, Gabon, Lesotho, Oman, Rwanda, Thailand, Uzbekistan, and Zimbabwe. Moreover, data for the former Soviet Republics are only available from 1991; data for TFYR of Macedonia, Croatia, and Slovenia only from 1992; and data for the Czech Republic, Slovakia and Eritrea only from 1993. In consequence, the final sample consists of 26,160 observations, instead of 26,800.

Note that all time-variant explanatory variables are lagged one period in order to lessen potential biases caused by system feedbacks. Notwithstanding this common practice, other sources of endogeneity and biased estimates are likely to arise. Hence, I discuss alternative solutions in the results section.

### III. RESULTS

#### *Baseline Estimations*

Table 2 presents the results of baseline PPML estimations, with robust, country-pair clustered standard errors. Column (1) regresses international co-inventorship against the focal variable, plus individual-country and time FE, as well as the list of controls.<sup>17</sup> The effect of the variable is positive and statistically significant. In particular, column (1) shows an elasticity of 0.20. That is to say, a 10% increase in the size of the inventor diaspora abroad is associated with a two-percent increase in international patent collaborations, which is also economically meaningful. This result is of the same order of magnitude as estimates for the case of diasporas and trade (Felbermayr and Toubal 2010; Felbermayr and Toubal 2012; Head and Ries 1998; Rauch and Trindade 2002, although slightly larger than in Aubry et al. 2014) and diasporas and FDI (Aubry et al. 2014).

---

17. In unreported results it is included bilateral (country-country) fixed-effects—results provided from the authors upon request. In a nutshell, in those estimations all time-variant variables dramatically decrease their coefficients, becoming statistically nonsignificant. However, regressions with only country FE and time FE are preferred because: (1) including bilateral FE removes 35% of the observations of the original sample; and (2) both diaspora networks as well as international co-patenting move slowly over time, which makes it difficult to identify any effect coming from within-pair variation—therefore identification relies on cross-country-pair differences only.

TABLE 2. Baseline Specifications without and with Time-Varying Multilateral Resistance

	(1)	(2)	(3)	(4)
	PPML	PPML	PPML	PPML
	Co- inventorship	R&D offshoring	Co- inventorship	R&D offshoring
ln(Diaspora)	0.200*** (0.0229)	0.0929** (0.0407)	0.243*** (0.0233)	0.111*** (0.0423)
ln(Distance)	-0.267*** (0.0523)	-0.105 (0.0721)	-0.0468 (0.0552)	0.173** (0.0760)
Contiguity	0.210* (0.127)	-0.0996 (0.211)	0.116 (0.122)	-0.271 (0.191)
Common language	0.625*** (0.132)	0.931*** (0.220)	0.434*** (0.108)	0.703*** (0.187)
Lang. similarity	0.523** (0.225)	0.822** (0.386)	0.405** (0.202)	0.694** (0.352)
Colonial links	0.0654 (0.124)	0.299* (0.172)	0.0315 (0.103)	0.291** (0.143)
Religion similarity	0.706*** (0.256)	0.219 (0.446)	0.612*** (0.235)	0.185 (0.419)
ln(EXP+IMP)	0.0457** (0.0197)	0.0710*** (0.0249)	0.228*** (0.0335)	0.329*** (0.0495)
ln(Tech.distance)	-0.0637 (0.0469)	-0.245*** (0.0598)	-0.136*** (0.0506)	-0.287*** (0.0689)
ln(# patents_i)	0.334*** (0.0530)	0.354*** (0.0693)		
ln(# patents_j)	0.0568 (0.115)	0.333 (0.205)		
ln(GDP p.c. _i)	1.098*** (0.233)	1.807*** (0.318)		
ln(GDP p.c. _j)	-0.0776 (0.552)	-1.107 (0.823)		
Constant	-6.072 (6.006)	-4.780 (8.651)	-0.410 (0.652)	-1.726** (0.791)
Observations	26,160	26,160	19,676	20,574
Pseudo R2	0.958	0.918	0.978	0.956
Sending FE	Yes	Yes	No	No
Receiving FE	Yes	Yes	No	No
Year FE	Yes	Yes	No	No
Sending FE*Time FE	No	No	Yes	Yes
Receiving FE*Time FE	No	No	Yes	Yes
Log Lik	-18492.28	-23746.92	-16226.61	-20463.96

Source: Authors' analysis based on data described in the text.

Notes: Country-pair clustered robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. '\_i' and '\_j' stand for, respectively, migrant inventor's sending country and migrant inventor's receiving country. Per capita GDP at origin presents several missing observations: for 1990, missing data correspond to Azerbaijan, Eritrea, Cambodia and Latvia; for 1991, to Eritrea; and for 2005, to Cyprus, Gabon, Lesotho, Oman, Rwanda, Thailand, Uzbekistan, and Zimbabwe. Moreover, data for the former Soviet Republics are only available from 1991; data for TFYR of Macedonia, Croatia, and Slovenia only from 1992; and data for the Czech Republic, Slovakia, and Eritrea only from 1993. In consequence, the final sample consists of 26,160 observations, instead of 26,800. The lower number of observations between columns (1), (2) and (3), (4) is due to the inclusion of fixed effects in pseudo-maximum likelihood estimations: the PPML method automatically drops the country-specific fixed-effects (and their corresponding observations) for which the country has zero recorded inventors' flows to every other country in the sample in order to achieve convergence. Results are comparable to other count data methods without removing these observations (see Santos Silva and Tenreiro 2010 for further details).



The results for the remaining explanatory variables are interesting in themselves. As expected, physical distance between the most populated cities exerts a negative influence on the likelihood for cooperation across national boundaries, although sharing a common border barely affects co-inventorship. Common language has a strong positive estimated effect on collaborations between inventors of different countries. Other proxies for cultural similarity, such as language and religion proximity, exert a strong positive effect too. However, historical links between country pairs expressed by their colonial past are not significant. As expected, bilateral trade is positive and significant, while technological distance between countries, that is, how distant countries are in their technological specialization, exerts a negative influence on bilateral co-patents. Finally, both attribute variables—total number of patents and GDP per capita—are significant for the case of origin countries but not for destinations. Thus, it appears that differences across industrialized economies in terms of technological and economic development are relatively minor and are picked up by their country FE.

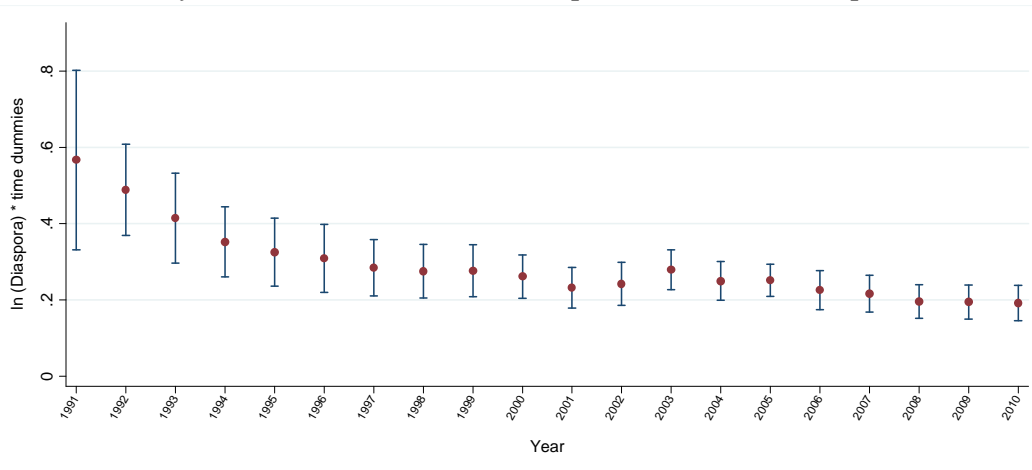
Column (2) looks at R&D offshoring—co-patents between applicants in developed countries and inventors in developing economies. Comparing the estimates with those of column (1), interesting results emerge. First and foremost, the estimated elasticity of inventor diaspora size is notably reduced in these latter estimations—less than a half. That is, diaspora networks particularly mediate interpersonal relations between co-workers. Meanwhile, they have a more nuanced effect on transnational employer-employee linkages.

Second, geography per se does not play a significant role in explaining R&D offshoring. The diaspora and geography results put together seem to suggest that personal face-to-face relations and trust building are critical in explaining co-

inventorship—where contracts are usually more tacit and contract enforcement is difficult—but less important in explaining more formal and hierarchical relationships, such as those represented by offshoring relationships, where probably explicit, written contracts are the rule.

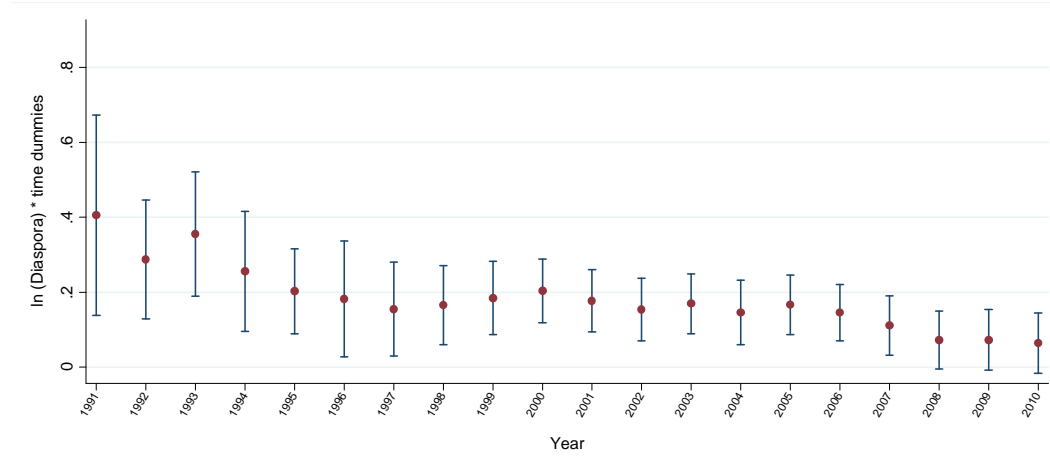
Other remarkable differences are worth reporting. For instance, the coefficient associated with a colonial past increases its point estimate and now becomes significant. That is, historical ties between the former metropolis and its formal colonies seem to have left an enduring effect over time that, still today, influences innovation networks across national borders. Finally, the common specialization of countries seems to play a greater role too when looking at applicant-to-inventor co-patents, as compared to inventor-to-inventor collaborations.

FIGURE 3. Yearly Estimated Coefficients of Diaspora on Co-inventorship, 1991–2010



Source: Authors' analysis based on data described in the text.

FIGURE 4. Yearly Estimated Coefficients of Diaspora on R&D Offshoring, 1991–2010



Source: Authors' analysis based on data described in the text.

Furthermore, these specifications (columns [1] and [2]) are used to include the focal variable interacted with time dummies, in order to explore the effects of skilled diasporas over time. Figures 3 and 4 present the estimated coefficients of the interaction variables, and show quite interesting, although expected, results. As can be seen, diaspora effects over time on co-inventorship and R&D offshoring present a marked decreasing trend. Quite likely, the use of information and communication technologies (ICT) has made the utilization of international ethnic ties less relevant now than 20 years ago. Similarly, the economic and institutional development of emerging economies has also contributed to mitigate the role of skilled diasporas in reducing the costs of asymmetric information—for example, think of the role of intellectual property in these emerging countries in the aftermath of their subscription to TRIPs, the Trade Related Intellectual Property Agreements that comes with membership of the World Trade Organization (WTO).

Columns (3) and (4) of table 2 mimic estimations (1) and (2) but control for time-variant multilateral resistance. While country FE controls for average multilateral resistance to collaborating over time (Feenstra 2004), some elements of this multilateral

resistance are likely to be time-variant and may not be picked up by the attribute variables included (Adam and Cobham 2007).<sup>18</sup> Consequently, these columns include country-specific time dummies, and repeat the main estimations, focusing attention only on bilateral variables. Some nameable differences with respect to the baseline regressions emerge, like a reduced role of distance in explaining inventor-to-inventor collaborations. However, the focal variable remains positive and strongly significant, and it presents coefficients that are slightly larger than previously.

*Identification: Cultural Proximity and Instrumental Variables*

Table 3 adds interaction terms between the inventor diaspora variable and different dimensions of cultural proximity between countries—common language, language similarity, common colonial past, and religion similarity. Given that transnational migrant networks mitigate the costs of incomplete information beyond country boundaries, one would expect their impact to be stronger for country pairs exhibiting larger informational frictions. Hence, negative and significant interaction terms will provide evidence on the least similar countries relying more on diaspora externalities than pairs of countries that are culturally closer. Results (table 3) partially confirm this extreme: the interaction with colonial ties is negative and significant—also the interaction with language similarity in the last specification.<sup>19</sup>

As in Kugler et al. (2013), I interpret these negative coefficients as evidence of a causal link between inventor diasporas and international co-inventive activity. If unobserved confounding factors drive both migration and co-patenting at the same time,

---

18. For an application to the migration literature, see Bertoli and Fernández-Huertas Moraga 2013.

19. The same estimation procedure using R&D offshoring as the dependent variable delivers not significant coefficients. This is further evidence of the critical role of diasporas for worker-to-worker collaborations and their more nuanced effects for the case of more hierarchical, R&D offshoring relations.

they should work in such a way that they are capable of explaining not only the main results—the diaspora-co-patenting relation—but also the differentiated effect of diaspora networks across different cultural dimensions, which is unlikely.

TABLE 3. Inventor Diaspora Effect between Culturally Closer/More Distant Countries

	(1)	(2)	(3)	(4)	(5)
	PPML	PPML	PPML	PPML	PPML
	Co-inventorship				
ln(Diaspora)	0.204*** (0.0247)	0.199*** (0.0229)	0.198*** (0.0230)	0.201*** (0.0230)	0.206*** (0.0249)
ln(Distance)	-0.263*** (0.0535)	-0.270*** (0.0517)	-0.267*** (0.0520)	-0.275*** (0.0560)	-0.262*** (0.0557)
Contiguity	0.213* (0.127)	0.217* (0.128)	0.233* (0.127)	0.216* (0.127)	0.250** (0.127)
Common language	0.690*** (0.193)	0.613*** (0.132)	0.630*** (0.131)	0.628*** (0.131)	0.776*** (0.198)
Lang. similarity	0.508** (0.227)	0.781*** (0.273)	0.528** (0.224)	0.532** (0.224)	0.858*** (0.285)
Colonial links	0.0502 (0.131)	0.0774 (0.124)	0.473* (0.248)	0.0724 (0.126)	0.468* (0.253)
Religion similarity	0.702*** (0.254)	0.676*** (0.262)	0.692*** (0.256)	0.782*** (0.267)	0.639** (0.287)
ln(Dia.)*Com. language	-0.0122 (0.0223)				-0.0302 (0.0241)
ln(Dia.)*Language sim.		-0.0829 (0.0522)			-0.117* (0.0606)
ln(Diaspora)*Colonial			-0.106** (0.0498)		-0.110** (0.0494)
ln(Diaspora)*Religion				-0.0526 (0.0708)	0.00141 (0.0758)
Controls	Yes	Yes	Yes	Yes	Yes
Constant	-6.102 (6.019)	-5.847 (6.000)	-6.941 (5.985)	-5.701 (6.080)	-6.745 (6.044)
Observations	26,160	26,160	26,160	26,160	26,160
Pseudo R2	0.958	0.958	0.957	0.957	0.958
Sending FE	Yes	Yes	Yes	Yes	Yes
Receiving FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Log Lik	-18490.83	-18482.05	-18478.32	-18490.28	-18459.73

Source: Authors' analysis based on data described in the text.

Notes: Country-pair clustered robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. 'i' and 'j' stand for, respectively, migrant inventor's sending country and migrant inventor's receiving country. Per capita GDP at origin presents several missing observations: for 1990, missing data correspond to Azerbaijan, Eritrea, Cambodia, and Latvia; for 1991, to Eritrea; and for 2005, to Cyprus, Gabon, Lesotho, Oman, Rwanda, Thailand, Uzbekistan, and Zimbabwe. Moreover, data for the former Soviet Republics are only available from 1991; data for TFYR of Macedonia, Croatia, and Slovenia only from 1992; and data for the Czech Republic, Slovakia, and Eritrea only from 1993. In consequence, the final sample consists of 26,160 observations, instead of 26,800.

More importantly, instrumental variable estimates are also provided. Potential and available candidates for such a role are (i) the size of the bilateral diaspora between countries  $i$  and  $j$  in 1960 (data from Özden et al. 2011) and its square, and (ii) the size of the unskilled diaspora (migrants with only primary education) originally from country  $i$  residing in country  $j$  in 1990 (data from Docquier et al. 2009) and its square. First, the stocks of migrants by country of origin in the 1960 censuses (therefore immigrants arrived between the end of the Second World War and 1960) are likely to affect the current stocks of highly-skilled migrants through network effects favoring further migration flows over the long run. Note that these figures include foreign-born people counts in dates closer to the age of mass migration than to the technological revolution of the 1990s and the 2000s. Quite probably, they are uncorrelated with current levels of cross-country collaborations, apart from influence through current skilled diasporas. Similarly, the current stocks of migrants with primary or lower levels of education correlate with current stocks of highly-skilled diasporas. The relation between existing diasporas and existing migration flows not only operates at a labor market level, but also among ethnic communities operating across different skills groups. Large stocks of unskilled immigrants in a given country will mean the existence of attractive factors—for example, amenities—which are also attractive to highly-skilled immigrants (Hunt and Gauthier-Loiselle 2008). On the other hand, uneducated migrants should play a non-existent role in boosting co-inventorship or R&D offshoring with their homelands—justifying their exclusion from the main equations, apart from their effects through inventor diasporas. Moreover, unskilled diaspora data come from the 1990 census—which accounts for the unskilled migrant flows of the 1980s—so as to be more confident that they are unaffected by unobserved factors influencing co-patenting patterns between 1990 and 2010.

Table 4 presents GMM estimations of the PPML—see Windmeijer and Silva (1997). Column (1) presents the first-stage results. Note that the value of the F-test statistic of the first stage, 344.58, is well above 10, which is usually considered a good threshold, and so the instruments cannot be judged as weak. Moreover, Hansen J statistics for mutual consistency of available instruments are provided at the bottom of columns (2) and (3), and they do not reject the null hypothesis that the excluded instruments are valid and uncorrelated with the error term, so there are no over-identification problems. Column (2) shows GMM estimates using co-inventorship as the dependent variable. It shows a positive and statistically significant relationship between inventor diasporas and international co-inventorship. The GMM results are slightly stronger in terms of magnitude of estimated coefficient relative to former PPML. Thus, analysis suggests that ignoring the endogeneity issue tends to underestimate the effect of migration on international co-inventorship. Note, however, that the difference is small and therefore the coefficients are comparable. On the other hand, results for the case of R&D offshoring remain positive but not significant (column [3]). I interpret these differentiated results as further evidence of the critical role of highly-skilled diasporas for worker-to-worker co-inventorship and their less important effect for applicant-to-inventor relations. Personal face-to-face relations and trust building are critical to explain co-inventorship, where contracts are usually more tacit and contract enforcement is difficult. However, in more formal and explicit relationships where written contracts are the rule, such as R&D offshoring, highly-skilled, technical immigrant workers do not play any role. Of course, one cannot rule out the possibility of measurement error of the diaspora variable due to the naturalization issue discussed above—in fact, results in appendix S3 using ethnic name recognition algorithms point to a larger coefficient on the estimated diaspora-R&D offshoring relation.

TABLE 4. GMM Estimates with Instrumented Diaspora

	(1)	(2)	(3)
	1 <sup>st</sup> stage results	GMM	GMM
		Co-inventorship	R&D offshoring
ln(Diaspora)		0.223*** (0.0774)	0.120 (0.170)
ln(Distance)	-0.268*** (0.0315)	-0.247*** (0.0649)	-0.132 (0.0985)
ln(diaspora 1960s)	0.0128 (0.0102)		
ln(diaspora 1960s)^2	0.00534*** (0.00163)		
ln(low-skilled diaspora)	0.0549*** (0.00971)		
ln(low-skilled diaspora)^2	0.0137*** (0.00195)		
Controls	Yes	Yes	Yes
Constant	14.86*** (1.967)	-6.866 (5.998)	-0.0369 (8.833)
Observations	26,160	26,160	26,160
Controls	Yes	Yes	Yes
Sending FE	Yes	Yes	Yes
Receiving FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
F-test	344.58		
p-value	0.0000		
Hansen's J chi2		3.00468	6.13152
p-value		0.3909	0.1054

Source: Authors' analysis based on data described in the text.

Notes: Country-pair clustered robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. '\_i' and '\_j' stand for, respectively, migrant inventor's sending country and migrant inventor's receiving country. Per capita GDP at origin presents several missing observations: for 1990, missing data correspond to Azerbaijan, Eritrea, Cambodia, and Latvia; for 1991, to Eritrea; and for 2005, to Cyprus, Gabon, Lesotho, Oman, Rwanda, Thailand, Uzbekistan, and Zimbabwe. Moreover, data for the former Soviet Republics are only available from 1991; data for TFYR of Macedonia, Croatia, and Slovenia only from 1992; and data for the Czech Republic, Slovakia, and Eritrea only from 1993. In consequence, the final sample consists of 26,160 observations, instead of 26,800. Instruments are centered around their mean, and they are as follows: (i) the size of the bilateral diaspora between countries i and j in the 1960s—and its square, and (ii) the size of the unskilled diaspora original from country i residing in country j in 1990 (migrants with only primary or lower levels of education)—and its square.

One way to see whether my interpretation of the different effects of inventor diaspora is plausible is to include tertiary educated stocks of migrants as extracted from census data. The data come from Docquier et al. (2009) and were originally built from two census rounds, 1990 and 2000, in which the regressions were used, respectively, for the years 1991–2000 and the years 2001–2010 (recall that censuses record data only every 10 years). As can be seen in table 5, column (2), the coefficient for the inventor



diaspora variable on international co-inventorship remains strongly significant and largely unaltered when tertiary educated stocks of migrants are included among the regressors—which is not significant (column [1] reproduces the baseline result for comparison purposes). On the contrary, inventor diaspora is no longer significant for the case of R&D offshoring—column (4) as opposed to column (3), while tertiary educated stocks of migrants becomes positive and strongly significant. Although not conclusive, I interpret this result as evidence that other types of skilled migrants may play an important role in boosting R&D offshoring networks, beyond what inventors (most of them scientists and engineers) do. Managers, executives, CEOs, trade agents, venture capitalists, and in general, migrant entrepreneurs, armed with their knowledge of technology markets and their global social capital, might be better placed to favor international R&D businesses and R&D offshoring back home. Unfortunately, these migrant entrepreneurs do not necessarily apply for patents and therefore may remain hidden to our proxy for highly-skilled diasporas. However, confirming this extreme would require further empirical examination that goes beyond the scope of this paper.

TABLE 5. Baseline Regressions with Bilateral Highly-Skilled (HS) Migration Stocks

	(1)	(2)	(3)	(4)
	Co-inventorship		R&D offshoring	
ln(Inventor diaspora)	0.200*** (0.0229)	0.189*** (0.0236)	0.0929** (0.0407)	0.0639 (0.0408)
ln(HS migration)		0.0409 (0.0252)		0.103*** (0.0362)
Controls	Yes	Yes	Yes	Yes
Observations	26,160	26,160	26,160	26,160
Sending FE	Yes	Yes	Yes	Yes
Receiving FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Source: Authors' analysis based on data described in the text.

Notes: Country-pair clustered robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

### *Are China and India Different after All?*

Next, I look at the robustness of the results once the main players are removed from the analysis. This is motivated by the observation that the majority of studies look at the case of the largest receiving country, that is, the United States, and its main providers of skilled talent—that is, China, India, and other Asian economies (Kerr 2008; Agrawal et al. 2011; Breschi et al. 2015; Saxenian 1999, 2006). In the light of this, some scholars argue that lessons from case studies on China and India cannot be extrapolated to other migrant communities—that is, it is difficult to say whether highly-skilled emigrants systematically engage in business networks and knowledge transfers with their homelands or rather that the Indian and Chinese diasporas are so famous for being an exception rather than the rule (Gibson and McKenzie 2012).

In order to explore this issue, table 6 repeats co-inventorship estimations—with and without country-specific time dummies—but removing from the sample either the BRICS countries (Brazil, Russia, India, China, and South Africa) or the United States or both. Contrary to the arguments posited by Gibson and McKenzie (2012), among others, the coefficient accompanying the diaspora variable remains strongly significant and economically meaningful in all models and barely lower when compared to previous estimates.

TABLE 6. Are China and India the Exception Rather than the Rule?

	(1)	(2)	(3)	(4)
	PPML	PPML	PPML	PPML
Co-inventorship				
	No BRICS	No US	No BRICS, no US	No BRICS, no US
ln(Diaspora)	0.192*** (0.0384)	0.204*** (0.0358)	0.200*** (0.0419)	0.190*** (0.0465)
ln(Distance)	-0.447*** (0.0631)	-0.325*** (0.0582)	-0.515*** (0.0654)	-0.269*** (0.0759)
Bilateral controls	Yes	Yes	Yes	Yes
Attribute controls	Yes	Yes	Yes	No
Constant	-5.626 (7.610)	-7.494 (5.929)	-5.900 (7.425)	1.753*** (0.667)
Observations	24,180	24,852	22,971	14,752
Pseudo R2	0.900	0.828	0.658	0.725
Sending FE	Yes	Yes	Yes	No
Receiving FE	Yes	Yes	Yes	No
Year FE	Yes	Yes	Yes	No
Sending FE*Time FE	No	No	No	Yes
Receiving FE*Time FE	No	No	No	Yes
Controls	Yes	Yes	Yes	Yes
Log Lik	-14236.62	-15541.05	-11836.79	-10463.32

Source: Authors' analysis based on data described in the text.

Notes: Country-pair clustered robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Per capita GDP at origin presents several missing observations: for 1990, missing data correspond to Azerbaijan, Eritrea, Cambodia and Latvia; for 1991, to Eritrea; and for 2005, to Cyprus, Gabon, Lesotho, Oman, Rwanda, Thailand, Uzbekistan, and Zimbabwe. Moreover, data for the former Soviet Republics are only available from 1991; data for TFYR of Macedonia, Croatia and Slovenia only from 1992; and data for the Czech Republic, Slovakia, and Eritrea only from 1993. In consequence, the final sample consists of 26,160 observations, instead of 26,800.

### *Further Robustness Analysis*

To further check the robustness of the results, table 7 runs the baseline specification using different estimation methods and dependent variables—for the case of co-inventorship only. In particular, it includes OLS and Negative Binomial models. For both cases the estimated coefficients are slightly larger compared to table 2, but they are fairly comparable. Column (3) takes on board the potential heterogeneity of international co-patents in terms of quality and weights them according to the number of forward citations received within five years after their priority year. As can be seen, results are robust for this different approach to computing the dependent variable—although further research, possibly at the micro level, should shed more light on this

issue. Finally, in order to be sure that the choice of PCT applications to build the co-inventorship variable does not bias the results, the baseline regression is replicated using alternative patent data sources, such as the USPTO, the EPO, and “Triadic Patent Families” (TPF) (OECD Triadic Patent Families database, January 2014). TPF consist of a set of patents filed at the EPO, the Japan Patent Office (JPO), and granted by the USPTO that share one or more priority applications. If anything (columns [4] through [6]), it seems that using alternative sources of patent applications may overestimate the relationship between high-skilled migration and international co-patenting.

Appendix S7 replicates the baseline regressions but splits the count of co-patents and foreign inventors into five broad technology fields (Schmoch 2008). Appendix S8 uses alternative count data methods intended to zero-inflated dependent variables, with no important differences with respect to the main results.

TABLE 7. Robustness Checks. Co-inventorship

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS co-inv.+1	NegBin	Citation weighted	USPTO	EPO	TPF
	Co-inventorship					
ln(Diaspora)	0.232*** (0.0136)	0.221*** (0.0267)	0.165*** (0.0268)	0.229*** (0.0230)	0.174*** (0.0234)	0.219*** (0.0263)
ln(Distance)	-0.0136 (0.0156)	-0.453*** (0.0486)	-0.291*** (0.0621)	-0.213*** (0.0546)	-0.243*** (0.0626)	-0.207*** (0.0663)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Constant	1.089 (0.835)	-6.659 (5.554)	-0.0461 (7.122)	-10.27 (6.523)	-2.879 (5.891)	-1.395 (6.584)
Obs.	26,160	26,160	26,160	26,160	25,760	24,700
R2	0.571	0.354	0.903	0.980	0.937	0.883
Sending FE	Yes	Yes	Yes	Yes	Yes	Yes
Receiv. FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Source: Authors' analysis based on data described in the text.

Notes: Country-pair clustered robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Per capita GDP at origin presents several missing observations: for 1990, missing data correspond to Azerbaijan, Eritrea, Cambodia and Latvia; for 1991, to Eritrea; and for 2005, to Cyprus, Gabon, Lesotho, Oman, Rwanda, Thailand, Uzbekistan, and Zimbabwe. Moreover, data for the former Soviet Republics are only available from 1991; data for TFYR of Macedonia, Croatia and Slovenia only from 1992; and data for the Czech Republic, Slovakia, and Eritrea only from 1993. In consequence, the final sample consists of 26,160 observations, instead of 26,800.

#### IV. CONCLUSION

This paper examines the impact of highly-skilled migrant networks in high income countries on the internationalization of inventive activity between high income and developing economies—measured as cross-country PCT co-patenting (co-inventorship and R&D offshoring). In order to study this relationship, it makes use of a unique dataset on inventors with a migratory background. To my knowledge, there have been few previous attempts to measure the mentioned links, and therefore this constitutes the main contribution of the paper.

The results show a strong and positive association between highly-skilled diasporas and the internationalization of inventive activity between developed and developing countries. The effect is statistically and economically significant: a 10% increase in the inventor diaspora abroad is associated with a 2.0-2.2% increase in international patent collaborations at the level of inventors. The effect found is robust for the large list of controls and robustness checks performed—such as alternative ways of measuring the dependent and the focal explanatory variables. Regressions also include country FE and country-specific time dummies, in order to control for any time-invariant and time-variant characteristics of the country of origin or the country of destination of the migrants that may affect the relationship between inventor diasporas and international co-patenting. Hence, any source of bias should come from country-pair (omitted) variables, for which an IV approach is used in order to avoid the possibility of the focal regressors picking up any confounding effect that might bias their point estimates. These findings do not suffice to conclude that a “brain gain” exists that makes up for the loss of highly-skilled human capital of sending economies, although they are undeniably necessary elements. Note, however, that boosting

international co-inventorship and team formation is only one of the multiple brain *gain* effects of emigrant inventors, which may eventually include the international diffusion of knowledge (Kerr 2008; Breschi et al. 2015) or the accumulation of human capital in sending economies (Beine et al. 2008)—higher expected returns to the inventor activity.

In developing/emerging countries, accessing foreign technologies via international collaborations is a hot political issue (Sterzi and Montobbio 2013). From a development policy perspective, these findings support the idea that exploiting highly-skilled diaspora networks in technology frontier economies might be an instrumental way of engaging in international innovation networks. This subject matter ranks high among policymakers in these countries, as witnessed by the recent visit of the Indian Prime Minister to Silicon Valley.<sup>20</sup>

Interestingly enough, the effect, although relatively diminished, does not depend on the remarkable performance of particular diasporas abroad, such as Chinese or Indian inventors. Equally, results are not particularly driven by the country both attracting the largest number of migrant inventors and concentrating a significant proportion of North-South international collaborations, that is, the United States. It seems therefore that inventor diaspora effects are exhausted at relatively low levels of highly-skilled diasporas (for similar results for the trade-migration relationship, see Egger et al. 2012).

The IV results suggest that highly-skilled diaspora effects weaken dramatically in the case of R&D offshoring—collaborations between applicants in developed countries and inventors in developing ones. These results seem to suggest that personal

---

20. See: <http://www.theguardian.com/world/2015/sep/23/narendra-modi-aims-to-bolster-indias-tech-credentials-on-us-visit> (accessed 28th September 2015).

relations and trust building are critical to explain co-inventorship—where contracts are usually more tacit, contract enforcement is difficult, and diaspora networks may play a referral role—but are less important in explaining more formal and hierarchical relationships—where probably explicit, written contracts are the rule. Technical highly-skilled workers, such as scientists and engineers, are possibly not the best placed to boost R&D offshoring networks to their countries of origin, but other skilled workers, such as executives, managers, trade agents, or entrepreneurs, may play a more critical role. Although some preliminary evidence seems to give support to this hypothesis, further research, possibly at the firm and inventor levels, will shed more light on these particular issues.

Of course, using patent data for migration research does not come without limitations. One important caveat is that inventors are only observed when they seek patents. However, not all inventions are patented; indeed, the propensity to patent for each dollar invested in research and development differs considerably across industries. In addition, studies have documented a skewed distribution of patent values, with relatively few patents yielding high economic returns. Another limitation of the dataset used in this study is that it misses inventors with a migratory background that have become nationals of their host country. To the extent that it is easier to gain citizenship in some countries than in others, this introduces a bias in the data. A related bias stems from the possibility that migrants of some origins may be more inclined to adopt the host country's nationality than migrants from other origins. Unfortunately, the data do not allow the severity of these biases to be assessed. Notwithstanding these caveats, we believe that this database meaningfully captures a phenomenon of growing importance.

## REFERENCES

- Adam, C., and D. Cobham. 2007. *Modelling Multilateral Trade Resistance in a Gravity Model with Exchange Rate Regimes*. CDMA Conference Paper Series 0702. Centre for Dynamic Macroeconomic Analysis.
- Agrawal, A., D. Kapur, J. McHale, and A. Oettl. 2011. “Brain Drain or Brain Bank? The Impact of Skilled Emigration on Poor-Country Innovation.” *Journal of Urban Economics* 69 (1): 43–55.
- Aleksynska, M., and G. Peri. 2014. “Isolating the Network Effect of Immigrants on Trade.” *The World Economy* 37 (3): 434–55.
- Aubry, A., M. Kugler, and H. Rapoport. 2014. “Migration, FDI and the Margins of Trade.” unpublished manuscript  
[http://econ.biu.ac.il/files/economics/seminars/amandine\\_aubry.pdf](http://econ.biu.ac.il/files/economics/seminars/amandine_aubry.pdf).
- Beine, M., F. Docquier, and H. Rapoport. 2007. “Measuring International Skilled Migration: A New Database Controlling for Age of Entry.” *The World Bank Economic Review* 21 (2): 249–54.
- . 2008. “Brain Drain and Human Capital Formation in Developing Countries: Winners and Losers.” *The Economic Journal* 118 (528): 631–52.
- Bertoli, S., and J. Fernández-Huertas Moraga. 2013. “Multilateral Resistance to Migration.” *Journal of Development Economics* 102 (May): 79–100.
- Breschi, S., F. Lissoni, and E. Miguelez. 2015. *Foreign Inventors in the US: Testing for Diaspora and Brain Gain Effects*. Cahiers du GREThA 2015-25. Groupe de Recherche en Economie Théorique et Appliquée.
- Breschi, S., F. Lissoni, and G. Tarasconi. 2014. *Inventor Data for Research on Migration and Innovation: A Survey and a Pilot*. WIPO Economic Research



- Working Paper 17. World Intellectual Property Organization - Economics and Statistics Division.
- Burger, M., F. van Oort, and G.-J. Linders. 2009. "On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-Inflated Estimation." *Spatial Economic Analysis* 4 (2): 167–90.
- Cameron, A. C., and P. K. Trivedi. 1998. *The Analysis of Count Data*. Cambridge: Cambridge University Press.
- Chander, A.. 2001. "Diaspora Bonds." *New York University Law Review* 76: 1005–43.
- Clemens, M. A., Ç. Özden, and Hillel Rapoport. 2014. "Migration and Development Research Is Moving Far Beyond Remittances." *World Development* 64 (December): 121–24.
- de Rassenfosse, G., A. Schoen, and A. Wastyn. 2014. "Selection Bias in Innovation Studies: A Simple Test." *Technological Forecasting and Social Change* 81: 287–99.
- Docquier, F., B. L. Lowell, and A. Marfouk. 2009. "A Gendered Assessment of Highly Skilled Emigration." *Population and Development Review* 35 (2): 297–321.
- Egger, P. H., M. von Ehrlich, and D. R. Nelson. 2012. "Migration and Trade." *The World Economy* 35 (2): 216–41.
- Feenstra, R. C. 2004. *Advanced International Trade: Theory and Evidence*. Princeton, N.J.: Princeton University Press.
- Felbermayr, G. J., and F. Toubal. 2010. "Cultural Proximity and Trade." *European Economic Review* 54 (2): 279–93.
- . 2012. "Revisiting the Trade-Migration Nexus: Evidence from New OECD Data." *World Development* 40 (5): 928–37.

- Foley, C. F., and W. R. Kerr. 2013. "Ethnic Innovation and U.S. Multinational Firm Activity." *Management Science* 59 (7): 1529–44.
- Foray, D. 1995. "The Economics of Intellectual Property Rights and Systems of Innovation: The Persistence of National Practices versus the New Global Model of Innovation." *Technical Change and the World Economy: Convergence and Divergence in Technology Strategies*, 109–33.
- Gibson, J., and D. McKenzie. 2012. "The Economic Consequences of 'Brain Drain' of the Best and Brightest: Microeconomic Evidence from Five Countries." *The Economic Journal* 122 (560): 339–75.
- Giuri, P., M. Mariani, S. Brusoni, G. Crespi, D. Francoz, A. Gambardella, W. Garcia-Fontes, et al. 2007. "Inventors and Invention Processes in Europe: Results from the PatVal-EU Survey." *Research Policy* 36 (8): 1107–27.
- Gould, D. M. 1994. "Immigrant Links to the Home Country: Empirical Implications for U.S. Bilateral Trade Flows." *The Review of Economics and Statistics* 76 (2): 302–16.
- Greif, A. 1989. "Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders." *The Journal of Economic History* 49 (04): 857–82.
- Guellec, D., and B. van Pottelsberghe de la Potterie. 2001. "The Internationalisation of Technology Analysed with Patent Data." *Research Policy* 30 (8): 1253–66.
- Guiso, L., P. Sapienza, and L. Zingales. 2009. "Cultural Biases in Economic Exchange?" *The Quarterly Journal of Economics* 124 (3): 1095–131.
- Hall, B. H. 2011. *The Internationalization of R&D*. UNU-MERIT Working Paper Series 049. United Nations University, Maastricht Economic and social Research and training centre on Innovation and Technology.

- Harhoff, D., E. Mueller, and J. Van Reenen. 2013. *What Are the Channels for Technology Sourcing? Panel Data Evidence from German Companies*. CEP Discussion Paper dp1193. Centre for Economic Performance, LSE.
- Head, K., and J. Ries. 1998. "Immigration and Trade Creation: Econometric Evidence from Canada." *The Canadian Journal of Economics / Revue Canadienne d'Economique* 31 (1): 47–62.
- Hunt, J., and M. Gauthier-Loiselle. 2008. *How Much Does Immigration Boost Innovation?* Working Paper 14312. National Bureau of Economic Research.
- Jaffe, A. B. 1986. "Technological Opportunity and Spillovers of R & D: Evidence from Firms' Patents, Profits, and Market Value." *The American Economic Review* 76 (5): 984–1001.
- Javorcik, B. S., Ç. Özden, M. Spatareanu, and C. Neagu. 2011. "Migrant Networks and Foreign Direct Investment." *Journal of Development Economics* 94 (2): 231–41.
- Kapur, D., and J. McHale. 2005. *Give Us Your Best and Brightest: The Global Hunt for Talent and Its Impact on the Developing World*. Brookings Inst Press.
- Kerr, W. R. 2007. *The Ethnic Composition of US Inventors*. Harvard Business School Working Paper 08-006. Harvard Business School.
- Kerr, W. R. 2008. "Ethnic Scientific Communities and International Technology Diffusion." *Review of Economics and Statistics* 90 (3): 518–37.
- Kugler, M., O. Levintal, and H. Rapoport. 2013. *Migration and Cross-Border Financial Flows*. CReAM Discussion Paper Series 1317. Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London.
- Kugler, M., and H. Rapoport. 2007. "International Labor and Capital Flows: Complements or Substitutes?" *Economics Letters* 94 (2): 155–62.

- Martínez, C., and R. Rama. 2012. "Home or next Door? Patenting by European Food and Beverage Multinationals." *Technology Analysis and Strategic Management* 24 (7): 647–61.
- Mayer, T., and S. Zignago. 2011. *Notes on CEPII's Distances Measures: The GeoDist Database*. SSRN Scholarly Paper ID 1994531. Rochester, NY: Social Science Research Network.
- Melitz, J., and F. Toubal. 2014. "Native Language, Spoken Language, Translation and Trade." *Journal of International Economics* 93 (2): 351–63.
- Migueluez, E., and C. Fink. 2013. *Measuring the International Mobility of Inventors: A New Database*. WIPO Economic Research Working Paper 8. World Intellectual Property Organization, Economics and Statistics Division.
- Montobbio, F., and V. Sterzi. 2013. "The Globalization of Technology in Emerging Markets: A Gravity Model on the Determinants of International Patent Collaborations." *World Development* 44 (April): 281–99.
- Oettl, A., and A. Agrawal. 2008. "International Labor Mobility and Knowledge Flow Externalities." *Journal of International Business Studies* 39 (8): 1242–260.
- Özden, Ç., C. R. Parsons, M. Schiff, and T. L. Walmsley. 2011. "Where on Earth Is Everybody? The Evolution of Global Bilateral Migration 1960–2000." *The World Bank Economic Review* 25 (1): 12–56.
- Patel, P., and K. Pavitt. 1991. "Large Firms in the Production of the World's Technology: An Important Case of 'Non-Globalisation.'" *Journal of International Business Studies* 22 (1): 1–21.
- Patel, P., and M. Vega. 1999. "Patterns of Internationalisation of Corporate Technology: Location vs. Home Country Advantages." *Research Policy* 28 (2–3): 145–55.

- Picci, L. 2010. "The Internationalization of Inventive Activity: A Gravity Model Using Patent Data." *Research Policy* 39 (8): 1070–81.
- Rauch, J. E. 2003. "Diasporas and Development: Theory, Evidence, and Programmatic Implications." *Department of Economics, University of California at San Diego*.
- . 2001. "Business and Social Networks in International Trade." *Journal of Economic Literature* 39 (4): 1177–203.
- Rauch, J. E., and V. Trindade. 2002. "Ethnic Chinese Networks in International Trade." *Review of Economics and Statistics* 84 (1): 116–30.
- Santos S., J. M. C., and S. Tenreyro. 2006. "The Log of Gravity." *Review of Economics and Statistics* 88 (4): 641–658.
- . 2010. "On the Existence of the Maximum Likelihood Estimates in Poisson Regression." *Economics Letters* 107 (2): 310–12.
- . 2011. "Further Simulation Evidence on the Performance of the Poisson Pseudo-Maximum Likelihood Estimator." *Economics Letters* 112 (2): 220–22.
- Saxenian, A. 1999. "Silicon Valley's New Immigrant Entrepreneurs." In .
- . 2006. *The New Argonauts: Regional Advantage in a Global Economy*. Harvard University Press.
- Saxenian, A., Y. Motoyama, and X. Quan. 2002. *Local and Global Networks of Immigrant Professionals in Silicon Valley*. San Francisco, CA: Public Policy Institute of California.
- Scellato, G., C. Franzoni, and P. Stephan. 2012. *Mobile Scientists and International Networks*. Working Paper 18613. National Bureau of Economic Research.
- Schmoch, U. 2008. "Concept of a Technology Classification for Country Comparisons." *Final Report to the World Intellectual Property Organization (WIPO), Fraunhofer Institute for Systems and Innovation Research, Karlsruhe*.

- Thomson, R., G. de Rassenfosse, and E. Webster. 2013. "Quantifying the Effect of R&D Offshoring on Industrial Productivity at Home." *Mimeo, University of Melbourne*.
- Van Zeebroeck, N., and B. van Pottelsberghe de la Potterie. 2011. "Filing Strategies and Patent Value." *Economics of Innovation and New Technology* 20 (6): 539–561.
- Wadhwa, V., B. A. Rissing, A. Saxenian, and G. Gereffi. 2007. *Education, Entrepreneurship and Immigration: America's New Immigrant Entrepreneurs, Part II*. SSRN Scholarly Paper ID 991327. Rochester, NY: Social Science Research Network.
- Wadhwa, V., A. Saxenian, B. A. Rissing, and G. Gereffi. 2007. *America's New Immigrant Entrepreneurs: Part I*. SSRN Scholarly Paper ID 990152. Rochester, NY: Social Science Research Network.
- Walsh, J. P., and S. Nagaoka. 2009. "Who Invents?: Evidence from the Japan-U.S. Inventor Survey."
- Windmeijer, F. A. G., and J. M. C. Santos Silva. 1997. "Endogeneity in Count Data Models: An Application to Demand for Health Care." *Journal of Applied Econometrics* 12 (3): 281–94.
- World Bank. 2010. *Innovation Policy : A Guide for Developing Countries*. World Bank Publications. The World Bank.

## V. APPENDIX

### *Appendix S1: List of countries included in the analysis*

#### **List of developed countries**

Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Republic of Korea, Spain, Sweden, Switzerland, United Kingdom, and United States of America.

#### **List of developing/emerging/transition countries**

Algeria, Argentina, Armenia, Bangladesh, Belarus, Bosnia and Herzegovina, Brazil, Bulgaria, Cameroon, Chile, China, Colombia, Costa Rica, Croatia, Cyprus, Czech Republic, Ecuador, Egypt, Estonia, Ethiopia, Georgia, Ghana, Greece, Guatemala, Hungary, Iceland, India, Indonesia, Iran (Islamic Republic of), Iraq, Israel, Jamaica, Jordan, Kenya, Latvia, Lebanon, Lithuania, Luxembourg, Malaysia, Mauritius, Mexico, Moldova, Morocco, Nepal, Nigeria, Pakistan, Peru, Philippines, Poland, Portugal, Romania, Russian Federation, Singapore, Slovakia, Slovenia, South Africa, Sri Lanka, Syrian Arab Republic, T F Y R of Macedonia, Thailand, Trinidad and Tobago, Tunisia, Turkey, Ukraine, Uruguay, Venezuela, and Viet Nam.

## *Appendix S2: Patents and the PCT system*

Inventor nationality information to measure the migratory background of inventors comes from patent applications filed under the Patent Cooperation Treaty (PCT). Accordingly, some background of the PCT system, which facilitates the process of seeking patent protection in multiple jurisdictions, is provided.

A patent is the legal right of an inventor to exclude others from using a particular invention. To obtain a patent right, firms or individuals must file an application to the patent office. The PCT is just an international treaty administered by the World Intellectual Property Organization (WIPO) that facilitates patent protection in more than one jurisdiction.

The key here is to realize that patent rights only apply in the jurisdiction of the patent office that grants the right (USPTO, EPO, etc.). A patent applicant seeking to protect an invention in more than one country has two options. First, he can file applications directly at the patent offices in the jurisdictions in which the applicant wishes to pursue a patent –the so-called “Paris route”. Second, the applicant can file an application under the PCT, which gives the applicant additional time – as compared to the “Paris route” – to decide whether to continue to seek patent protection in more than one jurisdiction – and in which ones. The additional time gained can be valuable for applicants at a relatively early stage of the patenting process, at which the commercial significance of an invention is still uncertain.

For the purpose of economic analysis, using PCT applications present several advantages. First, the system applies one set of procedural rules to applicants from around the world and collects information based on uniform filing standards. This



reduces potential biases that would arise if one were to collect similar information from different national sources applying different procedural rules and filing standards.

Working with only a single national source (e.g., USPTO) may be a viable alternative for studying, say, inventor immigration for a particular country, but this approach could not reliably track migrating inventors on a global basis.

Second, the literature argues that PCT patents are associated generally to the patents with larger economic value. Several studies looking at EPO patents find a strong association between extending the protection to the PCT system and patent quality – see, for example, Guellec and Van Pottelsberghe de la Potterie, 2002; Jensen et al., 2011; van Zeebroeck and van Pottelsberghe de la Potterie, 2011. The reasons behind point to the fact that applicants with high-quality inventions tend to pursue a more proactive strategy to protect their intellectual property, including pursuing patent protection internationally. For this same reason, some evidence seems to suggest that large firms are over-represented among PCT applicants (Fernández-Ribas, 2010).

However, there is no reason to think that any systematic difference across countries in terms of quality of the patents or representativeness of large firms among applicants exists.

The number of PCT applications has been increasing, being around 190,000 in 2012 – for comparison purposes, there were around 165,000 applications at the EPO in 2012, and 519,000 at the USPTO, by far the most attractive market (see Table S2.1.). During 2005-2010, of all unique applications worldwide (that is to say, “patent families”) (column (2)), 15-17% went through the PCT system – this share is 17% for the case of the EPO, and 47-54% for the case of the USPTO.

However, out of all applications seeking international protection (that is, applying for patent protection in more than one jurisdiction), in 2010, around 54% of them went through the PCT system. The PCT share has continuously risen over the past two decades; in 1995 it only stood at 25.4% of all international patents. Admittedly, considerable differences in how residents of different countries use the system emerge. First, the propensity of patent applicants to seek protection beyond their national jurisdiction differs markedly. For instance, in 2011, residents of China filed fewer than 20,000 applications outside of China, or only 4.54% of all the applications by Chinese residents worldwide. In contrast, the 20 developed countries included in this paper show in general large shares of their patent activity in other countries: e.g., the Republic of Korea (26.4%), Japan (39.1%), US (42.7%), Germany (57.6%), the UK (59.7%), France (62.8%), the Netherlands (74.7%), and Switzerland (78.6%). Second, among those applicants seeking patent protection abroad, substantial variation around the 54% mentioned above emerges: the PCT share was between two-thirds and three-quarters for Finland, France, the Netherlands, Sweden, and the US; it was between one-half and two-thirds for Australia, Germany, Switzerland and the UK; and it was between one-quarter and one-half for Canada, Japan, and the Republic of Korea.

Hence, the importance of the PCT system among patentees has steadily increased, making this source of information a valuable resource for economic analysis.

Besides, it is also insightful to see, not only the share of worldwide patents (and international worldwide patents) that go through the PCT system, but also how much overlap exist between PCT patents and some other important patent repositories, such as the EPO and the USPTO. Table S2.2. does this. In particular, it looks at the number of EPO patents that go also through the PCT system (column (1)) and the number of USPTO patents that also go through the PCT system (column (3)) – as well as,

respectively, the share of overlap EPO-PCT (column (2)) and USPTO-PCT (column (4)). As can be seen, the overlap between EPO and PCT patents is notably high, while it is considerably lower for the USPTO-PCT case – although increasing during the 2000s, when the use of the PCT boomed worldwide.

Reassuringly, section 4.4 of the main text shows as a robustness check the estimation of the baseline co-inventorship model using alternative data sources to compute the variable, such as triadic patent families (TPF), which covers patent applications filed to the EPO, the JPO and the USPTO that share a same set of priorities. The same applies when the dependent variable is built using data from the EPO or from the USPTO.

Results clearly show that the inventor diaspora is strongly significant (and with comparable coefficients) irrespective of the data source used to compute the dependent variable.



**Table S2.1. Number of patent applications worldwide and by selected patent offices (1995-2013)**

Year	Applications worldwide	Patent Families	Applications through PCT	Share PCT over applications	Share PCT over families	Applications at EPO	Share EPO over applications	Share EPO over families	Applications at USPTO	Share USPTO over applications	Share USPTO over families
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1995	1,047,400	572,565	39,016	3.73%	6.81%	71,861	6.86%	12.55%	192,638	18.39%	33.64%
1996	1,088,400	592,413	47,060	4.32%	7.94%	79,104	7.27%	13.35%	194,631	17.88%	32.85%
1997	1,163,200	627,835	55,892	4.81%	8.90%	90,047	7.74%	14.34%	226,298	19.45%	36.04%
1998	1,214,800	665,622	65,657	5.40%	9.86%	102,159	8.41%	15.35%	236,149	19.44%	35.48%
1999	1,268,400	695,482	74,942	5.91%	10.78%	112,525	8.87%	16.18%	267,705	21.11%	38.49%
2000	1,377,400	772,496	91,441	6.64%	11.84%	126,947	9.22%	16.43%	321,306	23.33%	41.59%
2001	1,456,900	803,126	106,214	7.29%	13.23%	135,919	9.33%	16.92%	375,503	25.77%	46.76%
2002	1,443,600	806,431	108,658	7.53%	13.47%	134,700	9.33%	16.70%	385,722	26.72%	47.83%
2003	1,490,300	841,558	113,403	7.61%	13.48%	140,760	9.45%	16.73%	402,947	27.04%	47.88%
2004	1,574,400	863,716	120,638	7.66%	13.97%	149,644	9.50%	17.33%	453,647	28.81%	52.52%
2005	1,702,900	905,780	134,616	7.91%	14.86%	159,748	9.38%	17.64%	495,523	29.10%	54.71%
2006	1,791,200	929,842	149,529	8.35%	16.08%	166,677	9.31%	17.93%	490,291	27.37%	52.73%
2007	1,876,900	956,422	157,556	8.39%	16.47%	168,116	8.96%	17.58%	500,638	26.67%	52.34%
2008	1,929,200	984,022	160,538	8.32%	16.31%	168,860	8.75%	17.16%	488,789	25.34%	49.67%
2009	1,861,700	955,880	152,730	8.20%	15.98%	158,985	8.54%	16.63%	457,481	24.57%	47.86%
2010	1,996,800	1,009,086	161,443	8.09%	16.00%	164,431	8.23%	16.30%	478,533	23.96%	47.42%
2011	2,157,900	1,097,829	179,453	8.32%	16.35%	171,071	7.93%	15.58%	507,167	23.50%	46.20%
2012	2,356,500		191,561	8.13%		165,870	7.04%		519,214	22.03%	

Source: World Intellectual Property Indicators – 2013 edition (Wipo, 2013) and author's calculations from PATSTAT (EPO Worldwide Patent Statistical Database)

**Table S2.2. Overlap between PCT and EPO/USPTO applications**

<b>Year</b>	<b>EPO-PCT</b>	<b>% EPO with PCT</b>	<b>USPTO-PCT</b>	<b>% USPTO with PCT</b>
	(1)	(2)	(3)	(4)
<b>1990</b>	6,362	9.54%	250	0.19%
<b>1991</b>	9,713	15.49%	900	0.65%
<b>1992</b>	13,212	20.42%	3,080	2.11%
<b>1993</b>	17,775	27.43%	5,004	3.29%
<b>1994</b>	21,275	31.52%	6,918	4.13%
<b>1995</b>	25,396	35.34%	8,565	4.45%
<b>1996</b>	31,284	39.55%	10,111	5.19%
<b>1997</b>	36,758	40.82%	12,440	5.50%
<b>1998</b>	43,460	42.54%	15,279	6.47%
<b>1999</b>	49,972	44.41%	19,833	7.41%
<b>2000</b>	56,638	44.62%	23,211	7.22%
<b>2001</b>	61,991	45.61%	25,454	6.78%
<b>2002</b>	64,510	47.89%	30,771	7.98%
<b>2003</b>	68,602	48.74%	43,977	10.91%
<b>2004</b>	71,929	48.07%	75,334	16.61%
<b>2005</b>	77,840	48.73%	96,305	19.44%
<b>2006</b>	82,785	49.67%	72,257	14.74%
<b>2007</b>	81,622	48.55%	70,799	14.14%
<b>2008</b>	80,314	47.56%	76,496	15.65%
<b>2009</b>	80,368	50.55%	80,812	17.66%
<b>2010</b>	84,837	51.59%	87,180	18.22%

Source: Author's calculations from PATSTAT (EPO Worldwide Patent Statistical Database)

*Appendix S3: Inventor nationality vs. alternative measures of inventor migration*

The present paper uses inventor nationality information from PCT applications to measure bilateral migration stocks, which presents several advantages (see Appendix S2). The use of this new data source constitutes an important step forward for the empirical research on skilled migration and its consequences for sending and receiving countries.

Admittedly, the use of nationality data does not come without limitations, the more important one being the loss of all naturalized inventors who have acquired their host country citizenship and therefore do not show up in PCT statistics. Likely, inventor nationality data is going to under-estimate inventor diaspora figures.

An alternative approach to exploit inventors' information from patent data for migration and innovation research is to use name recognition algorithms to infer the potential migratory background of the listed inventors (Kerr, 2008; Agrawal et al., 2008, 2011; and more recently, Breschi et al., 2014). Name recognition algorithms use large repositories of names and surnames associated to a country of birth, as a sort of dictionary, to provide a list of possible countries of origin to each name and surname of the inventors, with an associated probability. This approach has the advantage of taking into account all the naturalized immigrants who have acquired their host country nationality. However, the method has its own disadvantages, such as (i) it may capture second and third generation migrants with little linkages with their ancestors' homelands – e.g., African immigrants in Europe, Turkish immigrants in Germany, therefore over-estimating the size of the diaspora; (ii) it usually captures a limited number of countries of origin (nine ethnicities in Kerr, 2008, and only Indians in

Agrawal et al., 2008, 2011); and (iii) it is not able to distinguish between, for instance, English, Irish, US, or Australian inventors; between German, Swiss and Austrian inventors; and between Spanish and Latin American inventors, among others.

Yet, in order to be sure that the chosen approach in the present paper does not bias the results and conclusions to a large extent, Appendix S3 makes use of the likely country of origin (the so-called “ethnicity”) of inventors, built using inventors’ information from the EPO and the IBM-GNR system, a commercial database for name disambiguation (Breschi et al., 2014). The IBM-GNR system is a large repository of names and surnames, associated to a country of origin, collected by the US immigration authorities during the 1990s. In the following paragraphs, several robustness analyses exploiting this alternative dataset are performed.

First of all, one of the appendices in Breschi et al. (2015) is partially reproduced. The authors compare immigrant inventor figures of EPO inventors residing in the US, for the 10 countries of origin they are able to tell apart using IBM-GNR, with inventor nationality data from PCT, again for US resident inventors of the same 10 countries of origin. These are compared at the same time to census data on college educated US residents as well as US residents born in the US, but with foreign ancestry. Data from the latter two come from IPUMS-USA, for the year 2000 (<https://usa.ipums.org/usa/>), and specifically computes:

- (i) The percentage share of US residents with 4+ years of college education, born outside the US, by country of birth (aged 15 and above)
- (ii) The percentage share of US residents (all education levels, aged 15 and above), born in the US but of foreign ancestry, by ancestors' country. This type of information is important, since it may indicate the presence of many non-English



surnames, and possibly names, which may induce name recognition algorithms to classify an inventor as of foreign origin, when in fact he or she maybe the descendant of 19th-20th century migrants.

Comparing columns (1) and (2) of Table S3.1., it can be seen that the percentage of EPO US-resident inventors of foreign origin, active in 2000 by country of origin and the percentage of PCT US-resident inventors of foreign nationality, 1995-2005, by nationality, are of similar magnitude. At the same time we observe the share of college-educated foreign born to be very similar to that of inventors of foreign origin (either according to their name or their nationality), for the majority of the countries of origin considered (column (4)) – China and India being the exceptions.

As expected, shares of foreign inventors built using the name recognition algorithm (IBM-GNR) are systematically larger as compared to shares of foreign inventors computed using nationality information. The former may over-estimate foreign inventors due to second and third generation migrants, while the latter may underestimate them, due to naturalized migrants. The country with the largest difference is Iran (column (3)), which might be considered an special case, since, as Breschi et al. (2015) argue, many Iranian inventors may be part of the migration wave following the 1979 revolution, later to acquire the US citizenship. Aside from this special case, the other countries with ratios larger than 1.5 are Italy, Germany and Poland. Interestingly, these three countries, together with France, present disproportionate shares of US residents, born in the US, with foreign ancestry (column (5)). This leads us to conclude that name recognition algorithms may misclassify inventors to migrant communities with long tradition of immigration to a given host country, such as Italians, Germans and Polish into the US, and that probably, nationality data are closer to the real figures.

**Table S3.1. Comparison of inventor and censual data, by Country of Origin**

	(1)	(2)	(3)	(4)	(5)
	% US-resident inventors of foreign origin, active in 2000, by country of origin <sup>(1)</sup>	% US-resident inventors of foreign nationality, 1995-2005; by nationality <sup>(2)</sup>	(1) / (2)	% 4+college-educated US residents, born outside the US, by country of birth <sup>(3)</sup>	% US residents (all education levels), born in the US, by ancestors' country <sup>(3)</sup>
China	3.879	3.673	1.06	1.346	0.189
Germany	2.07	1.038	1.99	0.598	13.457
France	0.752	0.589	1.28	0.159	2.912
India	3.839	2.984	1.29	1.547	0.067
Iran	0.351	0.110	3.19	0.28	0.016
Italy	0.459	0.228	2.01	0.164	4.861
Japan	0.589	0.483	1.22	0.345	0.252
Korea	0.534	0.482	1.11	0.631	0.059
Poland	0.202	0.111	1.82	0.196	2.452
Russia	0.582	0.469	1.24	0.597	0.874

(1) Source: EP-INV database (Breschi et al., 2015)

(2) Source: author's calculation from WIPO-PCT dataset.

(3) Source: IPUMS-USA census data

Second, in order to study the robustness of our estimates, irrespective of the data source used, additional regressions using inventor migration information coming from the Breschi's et al. (2014, 2015) name recognition algorithm (the IBM-GNR) are presented here. In particular, I take all the EPO inventors residing in one of the 20 developed countries of the sample, and infer a unique country of origin on the basis of his name and surname. It is worth mentioning that the IBM-GNR system does not provide a unique country of origin for each inventor, but a list of potential countries, plus a sort of probability associated to them. Thus, it was necessary to reduce this list to a unique country of origin – or groups of countries of origin sharing the same or similar linguistic groups. Thus, for instance, all the Chinese, Singaporean, or Taiwanese inventors were grouped under the “ethnicity” of origin “Chinese” and all the Indians and Bangladeshi inventors were grouped under the label “Indians”. This necessary grouping (which is one of the caveats of name recognition algorithms) leads to group the dependent

variable too, as well as the controls – e.g., I added up all the co-inventor and R&D offshoring relationships for the cases of China-US, Singapore-US, and Taiwan-US; I computed the average distance between the three cases; and I took the maximum for variables such as contiguity, same language, etc.

All in all, I end up with a group of 20 developed countries (the same as before), and a number of broad ethnicities of origin (18), or geographical/linguistic regions of origin, including Arabic, English, French, African countries with British colonization, African countries with French colonization, Greek, Hebrew, Portuguese, Romanian, Scandinavian, Spanish, Eastern European countries plus the Republics of the Former Soviet Union, Chinese, Indian, Indonesian, Turkish, Vietnamese, as well as other Asian – for a 20-year period. Table S3.2. reproduces the baseline regressions using EPO inventors from 18 origins according to their names and surnames. Note that the coefficients accompanying the diaspora variable are of the same order of magnitude than the results obtained using inventor nationality data – if anything, the coefficients are slightly larger in table S3.2. Thus, it can be safely concluded that, on average, inventor nationality information can be used as a proxy for highly-skilled worker migration.

**Table S3.2. Baseline regressions using name recognition algorithms to build the diaspora variable**

	(1)	(2)
	PPML	PPML
	Co-inventorship	R&D offshoring
ln(Diaspora – “ethnic”)	0.245*** (0.0377)	0.253*** (0.0562)
ln(Distance)	-0.185** (0.0742)	0.134 (0.113)
Controls	Yes	Yes
Constant	-2.049 (5.801)	-3.874 (10.02)
Observations	6,900	6,620
Pseudo R2	0.969	0.925
Sending FE	Yes	Yes
Receiving FE	Yes	Yes
Year FE	Yes	Yes
Log Lik	-9429.09	-12469.39

Finally, an additional robustness check is performed. In particular, Table S3.3. includes a set of regressions splitting the sample of developed countries according to their type of immigration laws. The first group includes countries that apply the principle of *jus sanguine*, by which citizenship is determined by parents’ citizenship, and not place of birth. On the contrary, in the second group applies the principle of *jus soli*, that is, citizenship is determined by place of birth. The rationale behind this robustness exercise is to see whether differences emerge across these groups of countries, which could be attributed to differences in how easy is to obtain host country’s citizenship.

The classification of receiving countries between *jus soli* and *jus sanguine* is taken from Bertocchi and Strozzi (2010), who assembled a database where countries are classified according to their immigration laws in 1948, 1975 and 2001, that is, whether a country’s legislation contains elements of the *jus soli/jus sanguine* principle in their immigrations laws.

In 1975, *jus sanguine* countries are Austria, Belgium, Switzerland, Denmark, Spain, Finland, Germany, Italy, Japan, the Republic of Korea, Norway and Sweden. *Jus soli* countries are Australia, Canada, United Kingdom, Ireland, New Zealand, and the US. Meanwhile, France and the Netherlands, apply a mixed regime reflecting elements of both *jus soli* and *jus sanguine*.

In 2001, *jus sanguine* countries are Japan, the Republic of Korea, Norway and Switzerland. *Jus soli* countries are Canada, Ireland, New Zealand, and the US. Finally, 12 countries apply a mixed regime in 2001: Australia, Austria, Belgium, Denmark, Spain, Finland, France, United Kingdom, Germany Italy, the Netherlands, and Sweden.

Three dummy variables for each tier of the Bertocchi's and Strozzi's (2010) classification is built, both for 1975 and for 2001, and interacted with the focal variable of this paper, inventor diasporas. Table S3.3. shows the results for both co-inventorship and R&D offshoring. As can be seen, differences across groups with respect to co-inventorship are minor in relation to the baseline regressions, which confirms the main hypothesis. Differences across groups of countries with respect to R&D offshoring are more pronounced when using the 1975 classification, but again, all the coefficients remain significant.

**Table S3.3. Baseline regressions split by country group, according to their immigration laws, 1975 and 2001**

	(1)	(2)	(3)	(4)
	Co-	R&D	Co-	R&D
	inventorship	offshoring	inventorship	offshoring
	1975		2001	
ln(Diaspora)*Jus sanguinis	0.190*** (0.0369)	0.158** (0.0629)	0.168*** (0.0515)	0.189** (0.0801)
ln(Diaspora)*Mixed regime	0.241*** (0.0482)	0.311*** (0.0603)	0.208*** (0.0353)	0.189*** (0.0557)
ln(Diaspora)*Jus soli	0.201*** (0.0232)	0.104*** (0.0374)	0.201*** (0.0230)	0.105*** (0.0390)
ln(Distance)	-0.275*** (0.0569)	-0.0922 (0.0808)	-0.288*** (0.0647)	-0.0633 (0.0889)
Controls	Yes	Yes	Yes	Yes
Constant	-6.023 (6.114)	-6.320 (8.404)	-5.747 (5.998)	-4.904 (8.823)
Observations	26,160	26,160	26,160	26,160
Pseudo R2	0.957	0.925	0.958	0.924
Sending FE	Yes	Yes	Yes	Yes
Receiving FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

**Notes:** Country-pair clustered robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. ‘\_i’ and ‘\_j’ stand for, respectively, migrant inventor’s sending country and migrant inventor’s receiving country’.

*Appendix S4: The top-20 most populated inventor migration corridors*

Table S4.1. shows the top-20 most populated inventor migration corridors for the period 2001-2010 (left panel). As expected, the US stands out as the most typical choice for destination country, while most origins are other high income economies. The nameable exceptions are the top two corridors – China-US and India-US – with middle income country origins. Other middle income economies are also important sources of inventors during the period 2001-2010 e.g., Russia, Turkey, Iran, Romania and Mexico (right panel).

**Table S4.1. Top-20 most populated corridors, 2001-2010**

Largest inventor migration corridors			Largest inventor migration corridors, limited to non-OECD sending countries		
Sending	Receiving	Counts	Sending	Receiving	Counts
China	United States	44,444	China	United States	44,444
India	United States	35,607	India	United States	35,607
Canada	United States	18,745	Russia	United States	4,347
U.K.	United States	14,897	China	Japan	2,514
Germany	United States	10,290	China	Singapore	1,925
Germany	Switzerland	8,199	Turkey	United States	1,923
R. of Korea	United States	7,264	Iran	United States	1,442
France	United States	6,540	Romania	United States	1,229
Japan	United States	5,065	Russia	Germany	1,217
Russia	United States	4,347	Mexico	United States	1,164
Australia	United States	3,243	Brazil	United States	1,116
Israel	United States	2,968	Malaysia	Singapore	1,094
France	Switzerland	2,748	Ukraine	United States	977
Netherlands	United States	2,708	China	U.K.	921
Austria	Germany	2,676	China	Germany	889
France	Germany	2,601	India	Singapore	847
China	Japan	2,514	Argentina	United States	821
Italy	United States	2,503	Singapore	United States	771
Germany	Netherlands	2,289	Malaysia	United States	728
Netherlands	Germany	2,140	South Africa	United States	721

*Appendix S5: Alternative measures of international co-patenting*

Appendix S5 repeats the baseline estimations using alternative ways to compute the co-inventorship and the R&D offshoring variables using PCT applications. In particular, two alternative approaches are considered:

- (i) A first approach is taken from Hoekman et al. (2010), and exponentially multiplies the number of collaborations as more inventors are involved. Thus, for instance, if an international co-patent is made by two inventors in the US and three in India, this translates into a value of six collaborations for the US-India pair.
- (ii) A second, more complex measure is taken from Picci (2010), and is better explained through an example: imagine an international co-patent, with two inventors from the US, one from India and one from China (patent A); then another patent (B), with an inventor in the US and an inventor in China. The aggregated co-inventorship figures will be computed as follows:

- $InvInv_{US,IN} = 0.5*0.25 + 0*0=0.125$
- $InvInv_{US,CN} = 0.5*0.25 + 0.5*0.5=0.375$

This computation results in non-integer dependent variables

Results in the table S5.1 show that conclusions do not change to a large extent when different approaches to compute the two different variables are used, with virtually inexistent changes in what refers to international co-inventorship, and more ambiguous results concerning R&D offshoring (in line with the general conclusions of the paper regarding the latter).



**Table S5.1. Alternative ways to compute the dependent variable**

	(1)	(2)	(3)	(4)
	Co-inventorship		R&D offshoring	
	(i)	(ii)	(i)	(ii)
ln(Diaspora)	0.191*** (0.0314)	0.232*** (0.0241)	0.0511 (0.0491)	0.0840* (0.0496)
ln(Distance)	-0.320*** (0.0724)	-0.282*** (0.0569)	-0.0765 (0.0877)	-0.0936 (0.0893)
Controls	Yes	Yes	Yes	Yes
Constant	-1.395 (8.391)	-8.685 (6.665)	2.239 (10.09)	-3.847 (10.40)
Observations	26,160	26,160	26,160	26,160
Pseudo R2	0.934	0.961	0.898	0.901
Sending FE	Yes	Yes	Yes	Yes
Receiving FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Appendix S6: Correlation matrix

**Table S6.1. Correlation matrix**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1														
2	0.93	1													
3	0.37	0.37	1												
4	0.01	0.03	-0.12	1											
5	0.05	0.02	0.11	-0.27	1										
6	0.07	0.08	0.13	0.1	0.05	1									
7	-0.03	-0.03	-0.01	-0.09	0.07	-0.19	1								
8	0.01	0.02	0.15	-0.05	0.13	0.32	-0.08	1							
9	-0.03	-0.03	-0.04	-0.03	0.12	0.09	0.33	0.15	1						
10	0.11	0.11	0.33	-0.13	0.08	0.06	-0.02	0.13	0.12	1					
11	-0.15	-0.16	-0.41	0.13	-0.07	0.03	-0.07	-0.02	-0.08	-0.24	1				
12	0.21	0.21	0.46	-0.16	0.1	-0.08	0.07	-0.01	0.03	0.23	-0.73	1			
13	0.18	0.17	0.54	0.00	0.02	-0.02	-0.03	0.04	-0.07	0.41	-0.21	0.26	1		
14	0.05	0.04	0.14	-0.23	0.12	-0.07	0.15	0.00	0.17	0.15	-0.44	0.52	0.09	1	
15	0.11	0.11	0.35	-0.13	0.02	0.00	0.12	-0.03	0.01	0.15	-0.22	0.26	0.51	0.09	1

**Notes:** 1. Collab. inv<sub>i</sub>- inv<sub>j</sub>; 2. Collab. app<sub>i</sub>- inv<sub>j</sub>; 3. ln(Diaspora size); 4. ln(Distance); 5. Contiguity; 6. Common language; 7. Language similarity; 8. Colonial links; 9. Religion similarity; 10. ln(EXP+IMP); 11. ln(tech. distance); 12. ln(# patents)<sub>i</sub>; 13. ln(# patents)<sub>j</sub>; 14. ln(GDP p.c.)<sub>i</sub>; 15. ln(GDP p.c.)<sub>j</sub>. ‘<sub>i</sub>’ and ‘<sub>j</sub>’ stand for, respectively, migrant inventor’s sending country and migrant inventor’s receiving country’.

### *Appendix S7: Sector heterogeneity*

Use of patent data allows me to identify sector-specific particularities in the relationship between international co-patenting and diaspora networks. In particular, I follow Schmoch's (2008) classification of IPC codes into 35 technology fields, and group them into 5 broad sectors – namely, electrical engineering, instruments, chemistry, mechanical engineering, and others.

Columns (1) through (5) of Table S7.1. split collaborations and diaspora data into the five sectors and re-estimate the baseline model on the determinants of international co-inventorship. The coefficient on the inventor diaspora is positive and statistically significant at a one-percent level in all the specifications. Moreover, and contrary to other variables, differences across sectors are not large, witnessing to the importance of networks regardless of the technology being analyzed. The lower part of Table S7.1. repeats the estimates for the R&D offshoring case. Again, the focal variable in this study presents positive coefficients in all sectors analyzed, although non-significant point estimates for some of them. Moreover, differences in coefficients are remarkable.

**Table S7.1. The effect of inventor diasporas, by technology field**

	(1)	(2)	(3)	(4)	(5)
	PPML	PPML	PPML	PPML	PPML
	Electrical engineering	Instruments	Chemistry	Mechanical	Other sectors
<b>Co-inventorship</b>					
ln(Diaspora)	0.248*** (0.0437)	0.180*** (0.0294)	0.161*** (0.0248)	0.149*** (0.0364)	0.183*** (0.0420)
ln(Distance)	-0.139 (0.0850)	-0.402*** (0.0672)	-0.319*** (0.0673)	-0.385*** (0.0774)	-0.696*** (0.122)
Controls	Yes	Yes	Yes	Yes	Yes
Constant	-31.16*** (11.71)	-8.392 (11.76)	7.643 (5.602)	11.72 (9.777)	33.93* (18.85)
Observations	23,500	24,300	25,760	23,780	20,740
Pseudo R2	0.932	0.889	0.912	0.796	0.699
Sending FE	Yes	Yes	Yes	Yes	Yes
Receiving FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
<b>R&amp;D offshoring</b>					
ln(Diaspora)	0.203*** (0.0665)	0.129*** (0.0331)	0.0892** (0.0402)	0.152*** (0.0415)	0.106 (0.0709)
ln(Distance)	0.0618 (0.123)	-0.216** (0.0934)	-0.0473 (0.105)	-0.175 (0.123)	-0.285 (0.173)
Controls	Yes	Yes	Yes	Yes	Yes
Constant	-8.768 (17.39)	14.61 (10.47)	0.175 (8.569)	9.154 (12.28)	17.99 (21.34)
Observations	24,700	24,960	25,760	24,560	21,140
Pseudo R2	0.872	0.884	0.838	0.766	0.505
Sending FE	Yes	Yes	Yes	Yes	Yes
Receiving FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

**Notes:** Country-pair clustered robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. ‘\_i’ and ‘\_j’ stand for, respectively, migrant inventor’s sending country and migrant inventor’s receiving country’.

### *Appendix S8: Zero-inflated estimations*

An inspection of the two dependent variables (co-inventorship and R&D offshoring) reveals that the proportion of zero values is notably high, around 80% in both cases.

Even though Santos Silva and Tenreyro (2011) use simulation techniques to show that the PPML estimator is generally well-behaved in the presence of many zeros, Appendix S8 replicates the main regressions using two alternative count data models which take explicitly into account this zero-inflated probability distribution of the dependent variables. In particular, the zero-inflated Poisson and the zero-inflated Negative Binomial models are used. As revealed by Table S8.1., results do not change to a large extent and therefore do not alter our main conclusions.

**Table S8.1. Zero-inflated count data models**

	(1)	(2)	(3)	(4)
	Inflation model (logit)	Zero-inflated Poisson	Inflation model (logit)	Zero-inflated NegBin
<b>Co-inventorship</b>				
ln(Diaspora)	-0.116* (0.0700)	0.261*** (0.0254)	-0.203* (0.110)	0.256*** (0.0284)
ln(Distance)	0.774*** (0.140)	-0.109* (0.0566)	0.646*** (0.201)	-0.280*** (0.0607)
Controls	Yes	Yes	Yes	Yes
Constant	1.506 (12.71)	1.449 (6.603)	2.905 (17.70)	2.203 (6.881)
Observations	26,160	26,160	26,160	26,160
Sending FE	Yes	Yes	Yes	Yes
Receiving FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
<b>R&amp;D offshoring</b>				
ln(Diaspora)	-0.116* (0.0700)	0.261*** (0.0254)	-0.203* (0.110)	0.256*** (0.0284)
ln(Distance)	0.774*** (0.140)	-0.109* (0.0566)	0.646*** (0.201)	-0.280*** (0.0607)
Controls	Yes	Yes	Yes	Yes
Constant	1.506 (12.71)	1.449 (6.603)	2.905 (17.70)	2.203 (6.881)
Observations	26,160	26,160	26,160	26,160
Sending FE	Yes	Yes	Yes	Yes
Receiving FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

**Notes:** Country-pair clustered robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. ‘\_i’ and ‘\_j’ stand for, respectively, migrant inventor’s sending country and migrant inventor’s receiving country’.

## References

- Agrawal, A., D. Kapur, and J. McHale. 2008. "How do spatial and social proximity influence knowledge flows? Evidence from patent data" *Journal of Urban Economics* 64(2), 258–269.
- Agrawal, A., D. Kapur, J. McHale, and A. Oettl. 2011. "Brain drain or brain bank? The impact of skilled emigration on poor-country innovation." *Journal of Urban Economics* 69(1), 43–55.
- Bertocchi, G., and C. Strozzi. 2010. "The Evolution of Citizenship: Economic and Institutional Determinants." *Journal of Law and Economics* 53(1), 95–136.
- Breschi, S., F. Lissoni, and E. Miguelez. 2015. *Foreign inventors in the US: Testing for Diaspora and Brain Gain Effects*. CReAM Discussion Paper Series number 1509, Centre for Research and Analysis of Migration (CReAM).
- Breschi, S., F. Lissoni, and G. Tarasconi. 2014. "Inventor Data for Research on Migration and Innovation: A Survey and a Pilot." WIPO Economic Research Working Paper No. 17, World Intellectual Property Organization - Economics and Statistics Division.
- Fernández-Ribas, A. 2010. "International Patent Strategies of Small and Large Firms: An Empirical Study of Nanotechnology." *Review of Policy Research* 27(4), 457–473.
- Guellec, D., and B. Van Pottelsberghe de la Potterie. 2002. "The Value of Patents and Patenting Strategies: Countries and Technology Areas Patterns." *Economics of Innovation and New Technology* 11(2), 133–148.
- Hoekman, J., K. Frenken, and R.J.W. Tijssen. 2010. "Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe." *Research Policy* 39(5), 662–673.
- Jensen, P.H., R. Thomson, and J. Yong. 2011. "Estimating the patent premium: Evidence from the Australian Inventor Survey" *Strategic Management Journal* 32(10), 1128–1138.
- Kerr, W.R. 2008. "Ethnic Scientific Communities and International Technology Diffusion." *Review of Economics and Statistics* 90(3), 518–537.
- Picci, L. 2010. "The internationalization of inventive activity: A gravity model using patent data." *Research Policy* 39(8), 1070–1081.
- Santos Silva, J.M.C., and S. Tenreyro. 2011. "Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator." *Economics Letters* 112(2), 220–222.
- Schmoch, U. 2008. "Concept of a technology classification for country comparisons." *Final report to the World Intellectual Property Organization (WIPO)*, Fraunhofer Institute for Systems and Innovation Research, Karlsruhe.
- van Zeebroeck, N., and B. van Pottelsberghe de la Potterie. 2011. "Filing strategies and patent value" *Economics of Innovation and New Technology* 20(6), 539–561.
- Wipo. 2013. "World Intellectual Property Indicators, 2013 edition." WIPO Economics & Statistics Series, World Intellectual Property Organization - Economics and Statistics Division.