INVENTOR DIASPORAS AND THE INTERNATIONALIZATION OF TECHNOLOGY

Job Market Paper
This version: November 2013

Ernest Miguélez
Economics and Statistics Division, World Intellectual Property Organization (WIPO)[1],
34, Chemin des Colombettes, CH-1211 Geneva 11, Switzerland.
Tel.: (+41) 223388344. Fax: (+41) 223388350
& research fellow at AQR-IREA
& research affiliate at CReAM

E-mail: ernestmiguelez@gmail.com

Abstract
This paper documents the influence of diaspora networks of high-skilled individuals – i.e., inventors – on international technological collaborations. By means of 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 in detail 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 diaspora and different forms of international co-patenting. However, the effect is decreasing with the level of formality of the interactions. Interestingly, some of the most successful diasporas lately documented – namely, Chinese and Indian ones – do not govern the results.

Key words: international collaborations, co-patents, R&D offshoring, inventors, diaspora networks, PCT patents, PPML, GMM

JEL: C8, J61, O31, O33, R0

[1] The views expressed in this article are those of the author and do not necessarily reflect those of the World Intellectual Property Organization or its member states. I acknowledge financial support from the Ministerio de Ciencia e Innovación ECO2011-30260-C03-03. I thank Michel Beine, Gaetan de Rassenfosse, Stuart Graham, Catalina Martinez, Hillel Rapoport, Massimo Riccaboni, Valerio Sterzi, and participants in the workshop "The Output of R&D and Innovative activities: Harnessing the Power of Patents Data" (JRC-IPTS Seville, September 2013), the 7th MEIDE conference (Santiago de Chile, November 2013) and the PSDM 2013 (Rio de Janeiro, November 2013), for valuable comments, and Julio Raffo for helpful discussions on previous versions of this paper. However, any mistake or omission remains mine.

1. Introduction

Firms internationalize their innovation activity in order to adapt their products to foreign production processes and foreign markets (Patel and Vega, 1999), monitor new technology developments (Guellec and van Pottelsberghe de la Potterie, 2001), and exploit technological advantages of foreign countries.[2] Besides, the increasing specialization and complexity of innovation production leads scientists and engineers to look for the most suitable co-inventors worldwide (Katz and Martin, 1997). In general, modern communication technologies have contributed to overcome geographical obstacles to cross-country economic interactions. However, technological collaborations are, still today, primarily a national phenomenon. More than 20 years ago, Patel and Pavitt (1991) observed that the production of technology "remains far from globalized", contrary to other features such as trade or Foreign Direct Investment (FDI). 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. The data used in the present paper confirm this extreme: out of all Patent Cooperation Treaty (PCT) applications listing at least 2 inventors, only 8-9% of them include inventors resident in a minimum of two countries during the 2000s. As for the case of trade (Helliwell, 1998; McCallum, 1995), informal barriers to cross-country economic interactions largely explain these low rates of internationalization. Transnational social networks, such as skilled migrant networks, may overcome these barriers and foster the internationalization of inventive activity. Examining this relationship constitutes the aim of this paper.

Those who belong to the same country likely share the same historical background, cultural roots, and common language (Picci, 2010). Likewise, it facilitates the formation of trust and mutual understanding, which are conducive to the creation of networks. It also eases the screening of potential partners, helps the managing and administration of a common project, and smooths the monitoring of partners' fulfillments. Undeniably, all these factors constitute critical determinants of network formation and imply severe barriers to the internationalization of inventive activity – and largely explain the figures commented above. Other important factors hampering international collaborations may

[2] Examples of such location-specific technological advantages are, e.g., benefiting from a particular science base – including scientists and engineers, university research particularly strong in certain area, or learn from local competitors (Hall, 2011).

refer to differences in legal frameworks and the rule of law, especially regarding the issue of intellectual property rights (Foray, 1995; Montobbio and Sterzi, 2013).

Migrant networks may smooth the obstacles to the internationalization of inventive activity. They create trust across national boundaries, provide information on market opportunities and, in general, reduce transaction costs of economic interactions between countries. Diaspora networks have been studied in the context of trade (Gould, 1994), FDI (Javorcik et al., 2011; Kugler and Rapoport, 2007), and international diffusion of ideas (Agrawal et al., 2011; Kerr, 2008). 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). To the best of my knowledge, however, no study has looked at the role of high-skilled diasporas in fostering international technological collaborations.[3]

The present paper looks at the specific issue of transnational inventive activity between developed and developing countries, which still remains an unexplored topic (see, recently, Montobbio and Sterzi, 2013). International technological collaborations constitute a critical mean to access frontier knowledge from industrialized countries, both in the form of formal exchange of information as well as by means of knowledge spillovers (Hall, 2011). In consequence, this subject matter is critical from a development policy perspective. In such framework, the paper investigates whether a link exist between developing countries' inventor diasporas residing in high income economies and the opportunities for technology collaboration between the former and the latter. The paper also aims to answer whether differences emerge across the type of linkages created – co-inventorship vs. R&D offshoring networks, and, finally, the extent to which countries' characteristics govern these potential relations – whether the least similar countries have the greatest potential to benefit from diaspora networks.

Additionally, this study extends the existent literature in a critical way. The large majority of diaspora and migration studies use total immigration data or tertiary

[3] The most related work I found to the present paper is the one by Foley and Kerr (2013), who study how the ethnic composition of US technological firms influences the internationalization of their R&D and inventive activities. I extend their work by looking at diaspora networks in several industrialized countries, not only the US. I also study diasporas coming from a wide range of origin countries, not only nine ethnicities. Further, my analysis at the level of countries let me capture broader effects of diaspora networks on international collaborations, beyond firms' responses to their share of immigrant employees.

educated immigration data retrieved from decennial censuses. However, tertiary education may include non-university tertiary degrees, undergraduate university degrees, and postgraduate and doctorate degrees, which, on top of that, might not be comparable fully across different countries. Further, these datasets usually refer to one single year, which does not allow exploiting time-series variation of the data. Contrarily, I use a novel dataset on inventors with migratory background as a proxy for high-skilled diaspora (Miguelez and Fink, 2013). Inventors constitute a specific class of workers at the upper tail of the skills distribution and arguably a more homogeneous group of employees as compared to the tertiary educated labor force as a whole. In addition, as the original data source come from patent data, I am able to exploit a longitudinal dataset – 21 years, including a large number of sending and receiving countries.

To anticipate the results to come, I find a robust and sizeable effect of high-skilled diasporas on the internationalization of inventive activity between developed, receiving countries and developing, sending economies. The effect is statistically and economically significant: a 10-percent increase in the inventor diaspora abroad is associated with a 1.5 to 2.2 percent 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 mediate particularly interpersonal relations between co-workers.

The outline of the paper is as follows: section 2 reviews previous theoretical and empirical contributions on how diasporas foster international economic interactions. Section 3 presents the novel dataset on inventor diasporas and develops the methodological setting, including all the econometric concerns. Section 4 presents the results and section 5 concludes.

2. Related literature and theoretical background

As standard trade models would predict, migration and international economic interactions – such as trade and FDI – are likely to be substitutes. The free movement of factors equalizes their prices and in consequence commodity prices equalize too, reducing incentives to trade (Egger et al., 2012). In a similar vein, FDI flows to where labor is relatively abundant. If migration reduces human capital endowments of origin countries, migration and FDI flows can be seen as substitute ways to match employers and employees across different countries (Kugler and Rapoport, 2007, 2005). The same logic could apply to international co-inventorship and R&D offshoring. If firms in developed countries internationalize their innovation activities seeking for foreign pools of specialized, high-skilled human capital, large diasporas in the host countries may attenuate their need to locate R&D labs abroad, for instance.

On the other hand, a growing body of literature emphasizes how migrant networks boost international economic transactions and thus counterbalance the negative impact of the brain drain. 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). Diasporas' potential benefits can be realized exploiting this feeling for 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.

First and foremost, migrant networks lower transaction costs associated to incomplete information problems. Playing such a role, diasporas affect the origin economy both directly and indirectly (Kapur and McHale, 2005). The direct effect refers to the diaspora members' willingness to interact, by themselves, with their home countries, in the form of remittances, investments, or sharing ideas and information. The indirect effects refer to diaspora members' role in leveraging their home countries reputation in international business networks; facilitating searching and matching between partners, customers-suppliers or in the labor market; and finally, in ensuring the contract fulfillments of the two parties involved (op. cit.). Due to their familiarity with local market needs, diasporas provide information about business opportunities in their homelands, and thus are critical to convey access to relevant information otherwise inaccessible because of cultural, language, institutional, administrative, or geographical barriers.

Second, diaspora networks lower transaction costs associated to the existence of asymmetric information. As Rauch (2003, 2001) posits, social networks operating across national borders, build, or substitute for, trust when contract enforcement is weak or non-existent. Indeed, diasporas create trust by establishing a sort of "moral community", which is used to transmit information about past opportunistic behavior in international business relations.

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. This is confirmed by Rauch and Trindade (2002) and Head and Ries (1998), who find that a 10-percent increase in the number of immigrants increases exports by one-percent and imports by three-percent. Several refinements of these studies have critically shown that immigrant networks affect less trade of more homogeneous products – for which prices convey the relevant information – than heterogeneous products – for which non-disclosed information is more relevant (Aleksynska and Peri, 2013). Similar conclusions emerge for the case of FDI. Javorcik et al. (2011) investigate the link between the presence of migrants in the US 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 US – especially migrants with college degree education (see also Kugler and Rapoport, 2007). On their side, Docquier and Lodigiani (2010) find that the elasticity of the capital growth rate to the stock of skilled emigrants is between two- and three-percent.

To the best of my knowledge, this literature is virtually salient on the relationship between international co-patenting and diaspora externalities. However, case studies and anecdotal evidence seem to suggest how important migrant networks for international cooperation are. Saxenian (1999) argues that skilled immigrants in the US 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 positioned to locate foreign partners quickly and manage complex business networks across cultural, institutional and linguistic boundaries, which is especially relevant in high-tech industries. The resulting transnational networks are likely to enhance economic opportunities both for California and for emerging regions in Asia (Saxenian, 1999) – see also Saxenian et al. (2002) and Saxenian and Sabel (2008).

A parallel research stream has documented the link between international skilled migration and cross-country collaborations in science. For instance, Regets (2001) finds a strong positive association between the number of foreign students awarded PhDs in the US and the degree to which scientific articles authored in their sending countries include a US author. More recently, Scellato et al. (2012) report a positive link between mobility and international research networks, for a group of surveyed scientists of 16 countries. Their study finds that around 40-percent of foreign-born researchers in these countries maintain research links with their homeland colleagues. Meanwhile, non-mobile researchers are less likely to collaborate with someone from outside the country than either immigrant or returnee scientists. Finally, Jonkers and Cruz-Castro (2013) show how high-skilled returnees – Argentinian scientists –likely cooperate with their former host country.

On their side, Foley and Kerr (2013) find significant effects of US firms' ethnic inventors in promoting linkages of these firms with the innovators' home countries, in the form of knowledge flows or R&D alliances. The use of inventors' data to studying diaspora externalities closely relates to the present work. Agrawal et al. (2011) use the likely cultural origin of inventors' names in USPTO patents to estimate the size of the Indian diaspora in the US. Afterwards, they test the prevailing source of knowledge for inventors resident in India, that is, the role of the Indian diaspora vis-à-vis the agglomeration of Indian inventors in their origin country. Interestingly, they find inventor emigration to harm domestic knowledge access. However, the knowledge transferred by the diaspora is more valuable than domestic knowledge when looking at the most important innovations measured by citations received. In a broader research agenda, Kerr (2008) estimates the ethnic origin of all USPTO inventors' names, for nine ethnicities: Chinese, English, European, Hispanic, Indian, Japanese, Korean, Russian, and Vietnamese. 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.

As Foley and Kerr (2013) argue, ethnic inventors in host countries are particularly apposite to help firms to capitalize in foreign opportunities and overcome barriers to the internationalization of inventive activity. Ethnic inventors have usually essential expertise for developing products crucial for that particular ethnicity, giving a 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 in parallel they also know 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. Finally, high-skilled individuals themselves, inventors in particular, are likely to return home while maintaining their linkages with their former host country, enabling the formation of further collaborative networks across national borders (Alnuaimi et al., 2012; Nanda and Khanna, 2010).

To conclude this review, it is worth pointing out that some scholars have argued that lessons from the most successful Asian diasporas – namely Indian and Chinese diasporas – do not straightly extrapolate to other migrant communities. In a nutshell, they argue that high-skilled emigrants do not systematically engage in business networks and knowledge transfers with their homelands, but rather, the Indian and Chinese diasporas are so famous for being the exception rather than the rule (Gibson and McKenzie, 2012). Others argue that, while the related literature is extensive for the case of the largest destination country – the US, it is limited for other receiving areas (Breschi et al., 2013). In light of these arguments, the empirical approach presented here explores the extent to which the US experience and its top providers of foreign talent govern diaspora effects on co-inventorship and R&D offshoring, or else results can be generalized.

3. Research methods

3.1. Inventors' international migration

In large part, the surge of empirical analysis described in the former section responds to census-based migration datasets becoming available in the last 15 years (Carrington and Detragiache, 1998; Docquier and Marfouk, 2006; Özden et al., 2011). These datasets, broken down by skills – primary, secondary and tertiary level of schooling, have allowed researchers to investigate empirically the role of skilled diasporas in fostering transnational interactions, such as trade or FDI.

Contrariwise, the present analysis is based on a new dataset on inventors with migratory background, applying for PCT patent applications, from 1990 to 2010. The use of inventors' data with migratory background comes with two main advantages as compared to the existing datasets. First, patent data (and thus inventors' information) is collected on a yearly basis – contrary to census data, collected every 10 years – and it is available for a large number of sending and receiving countries. Second, attained education may still differ markedly among tertiary educated workers – ranging from non-university tertiary degrees to PhDs, for instance. Moreover, inventors constitute a specific class of high-skilled workers which is more homogeneous than the tertiary educated workers as a whole. Indeed, they are behind the production of new knowledge and innovation that spur economic growth and well-being. On top of all that, using inventor information from PCT applications implies probably capturing the most skilled inventors. PCT patent applications are clearly aimed at being extended worldwide and may hence be associated with the most valuable inventions (Guellec and Van Pottelsberghe de la Potterie, 2002; Jensen et al., 2011; van Zeebroeck and van Pottelsberghe de la Potterie, 2011).[4]

Information on inventors with migratory background is retrieved from patent applications under the PCT treaty (WIPO IPSTATS databases).[5] To the best of my knowledge, PCT patent applications are the only ones recording this type of information. Behind that is the fact that not all countries are PCT contracting states, while only nationals or residents 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, it turns out that US patent application procedures bind the applicant of a patent also to be the

[4] The use of patent data does not come without limitations, though. Aside from the well-known issues of varying quality of patents and that not all inventions are patented, more worrisome for the present analysis is the fact that the observation of both migration and collaborations is based on successful outcomes (the patent application). However, potential biases created by the impossibility to observe migration and collaborations without a successful output are addressed by means of the instrumentalization strategy described below.

[5] See Miguelez and Fink (2013) for a detailed description of the dataset.

inventor. If a given PCT application included the US as a country in which the applicant considered pursuing a patent – a so-called designated state in the application – all inventors were listed as applicants and their residence and nationality information are, in principle, available – in fact, this is the case for the majority of applications.

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. [6] 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 – US (66%), Canada (81%), the Netherlands (74%), Germany (95%), the United Kingdom – UK (92%), France (94%), Switzerland (93%), China (92%) and India (90%), among others.[7]

Once individual-level data are retrieved, I aggregate across pairs of countries and years.[8] In particular, I treat each record in the patent database as if it were a different individual and compute diaspora variables for annually repeated time-windows of five years.

Out of all records with complete information – about 5 million, around 9-10% have migratory background – i.e., residence different from nationality. Figure 1 depicts the evolution of the share of inventors with migratory background – dashed line, alongside the same figures broken down by a number of selected receiving countries – in different colors. As can be observed, the share of worldwide migrant inventors has steadily increased over time.[9] Among the most receiving countries of the world, Canada, Australia and, notably, the US, stand out as being the primary receiving countries, as compared to their resident stock of inventors. Meanwhile, technology leading European countries, such as Germany or France, are lagging behind. Of special interest is the case of the UK, which has experienced a substantial increase in its stock of immigrant

[6] By "record" should be understood a unique combination of "inventor name" and "application number".
[7] To address this inconsistent coverage of migration information over time, I repeated the analysis splitting the sample in shorter time windows. No important differences regarding my main explanatory variables need to be reported.
[8] I use the priority date of applications to allocate individuals in time. By "priority date" I mean the first year the patent was applied worldwide.
[9] 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 migratory background in 2000, data compiled by Docquier and Marfouk (2006) or Beine et al. (2007) show that general migration rates in 2000 for population 25 years old and over were estimated around 1.8%, including 1.1% of immigrants among the unskilled population, 1.8% among population with secondary education, and 5.4% among population with tertiary education.

inventor population. On the other side, Japan is, and has been over the years, one of the developed countries with a smaller share of inventor immigrant population.

[Figure 1 about here]

Other European countries rank even better than the US in terms of immigration rates of inventors (Figure 2) – notably, Switzerland, Ireland or Belgium. However, the exceptional performance of the US in attracting talent is notorious when considering only immigrant inventors coming from low and middle income economies, as can be seen also in Figure 2.

[Figure 2 about here]

Indeed, the US concentrates around 60 percent of the overall inventor migration for the 2001-2010 period, and close to 75 percent of migrant inventors from low and middle income economies (Table 1, panel a). Other high income countries follow way behind. It is also possible to disaggregate these figures and show the top-20 most populated corridors (Table 1, panel b). 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 constitute important sources of inventors during the period 2001-2010 too – e.g., Russia, Turkey, Iran, Romania or Mexico.

[Table 1 about here]

Admittedly, the bulk of inventor migration is concentrated in North-North corridors, although the South-North corridors are also sizeable (Figure 3). In fact, the South-North corridor has gained prominence over the North-North one, greatly due to the massive migration flows of Indian and Chinese inventors to the US. Meanwhile, inventor migration in North-South or South-South corridors is anecdotal.

[Figure 3 about here]

A similar pattern emerges when looking at cross-country PCT co-patenting figures at the inventor level (Figure 4) – co-inventorship. The Figure depicts the share of international collaborations when (i) inventors are only from the OECD area, (ii) inventors are both from OECD and non-OECD countries; (iii) only non-OECD inventors are included. Clearly, international inventions between OECD countries' inventors dominate. However, PCT co-inventions listing at least 1 non-OECD inventor has risen exponentially from 14% at the beginning of the 1990s to around 40% in 2012. The lower costs of conducting R&D activities in developing economies largely explain this surge (Montobbio and Sterzi, 2013). Moreover, due to the recent experiences of economic growth and technological development of some emerging economies, access to qualified personnel and location-specific knowledge pools have driven co-inventorship and R&D offshoring with these countries too (Thursby and Thursby, 2006).

[Figure 4 about here]

3.2. Empirical approach

All in all, the model to be estimated takes the following form:

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

where $COPAT_{ijt}$ stands for the number of collaborations between i's developing country (out of 99) and j's developed country (out of 20), for year t. $DIASPORA_{ijt}$ is the focal variable and is computed in two main ways. First, the number of inventors nationals of country i residing in country j, for annually repeated 5-year time-windows; second, the share of inventors nationals of country i residing in country j out of all inventors residing in country j, for annually repeated 5-year time-windows. Z_{ijt} is a set of bilateral and attribute control variables, and τ_i, τ_j, and δ_t are, respectively, developing, developed and time FE. ε_{ijt} denotes the error with the usual desired properties.

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 allow the logarithmic transformation would be clearly misleading (Burger et al., 2009). In addition, Santos Silva and Tenreyro (2006) show that log-linearizing equation (1) may induce a form of heteroskedasticity of the error term because of the log-transformation of the data, making OLS estimations inconsistent. Instead, the authors suggest estimating the multiplicative form of the model by Poisson pseudo-maximum likelihood (PPML), which provides also a natural way to deal with zero co-patenting and the extreme skewness of the dependent variable, intrinsically heteroskedastic with variance increasing with the mean (Cameron and Trivedi, 1998). In sum, I estimate equation (1) by means of PPML using the fact that the conditional expectation of $COPAT_{ijt}$ in (1) can be written as the following exponential function

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

For robustness, I also run all models by OLS and other estimation methods, finding no differences regarding the main conclusions, though significantly different point estimates of some of the coefficients.

3.3. Data

Dependent variable

International co-patent data is retrieved from PCT applications (WIPO IPSTATS databases). I first focus on co-patenting at the inventor level – co-inventorship. The use of the inventor level is preferred rather than the applicant level in order to attribute the actual co-invention process to the countries in which it has been developed. Moreover, knowledge is more likely to pass through interpersonal links at the inventor level rather than at the applicant level (Montobbio and Sterzi, 2013). I annually add up all the co-inventions between inventors residing in country i and inventors residing in country j.

To be precise, I include 99 developing/emerging/transition countries, on the one side, that co-invent with 20 developed countries, on the other side, where diasporas from the former countries reside. If inventors from more than two countries participate in the patent, I count an international co-inventor for each country-pair, irrespective of the total number of countries involved in that particular invention.[10]

Next, I also look at the role of inventor diasporas in fostering R&D offshoring to their homelands. 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, I compute a single co-patent for each bilateral i-j pair of countries.

It is worth mentioning that previous studies on the determinants of international co-patenting use information from single patent offices, with few exceptions (Picci, 2010). This practice when studying the internationalization of inventive activity between a large number of countries is likely to deliver biased estimates due to the 'home bias effect'. The 'home bias effect' emerges 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 area, they are preferred by domestic firms, and thus their innovative capability is overestimated with respect to foreign firms. Instead, using data from the PCT mitigates this effect because these patents are by definition international and applicants from all countries are equally likely to apply through the PCT system – provided that the applicants are either national or resident of a PCT member state.[11] For this same reason, this paper provides additional added value to the literature on inventor migration, which so far has focused on the US and using data from the USPTO only (Breschi et al., 2013).

[10] Notice that I computed other more complex measures of international co-invention, following Picci (2010) or Hoekman et al. (2010). No important changes arise – results provided upon request.

[11] Other biases inherent to the existence of multiple jurisdictions and patent offices are discussed in de Rassenfosse et al. (2013) – such as the non-random choice of the patent office. Again, the use of PCT applications should mitigate these biases.

Control variables

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

I also control for the intensity of economic linkages between countries using the share of bilateral trade (imports plus exports) between a given pair over their total trade (COMTRADE data). Trade is a conduit of information which may foster technological partnerships too, whilst 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

$$t_{ij} = 1 - \frac{\sum f_{ih} f_{jh}}{\left(\sum f_{ih}^2 \sum f_{jh}^2\right)^{1/2}}, \tag{3}$$

where f_{ih} stands for the share of patents of one technological class h according to the IPC classification (out of 300 technological classes in the subdivision chosen) 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 would indicate that a given pair of countries are technologically different, and values close to zero indicate that they are technologically similar (Jaffe, 1986). Again, I use PCT patents to compute this index (WIPO IPSTATS databases).

Finally, two additional attribute variables of individual countries are used in order to not bias the point estimates of my focal regressors. In particular, I introduce the number of PCT patents per country, for 5-year annually repeated time-windows. This variable controls for the size of the countries' innovation system, which clearly determines the 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, I retrieve GDP per capita from the World Development Indicators – World Bank, expressed in US$ 2005 at PPP, in order to capture market potential of countries as well as their capacity to innovate. Appendix 2 contains summary statistics of the variables included in the models, as well as the correlation matrix.

Note that I lag one period all time-variant explanatory variables 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.

4. Results

4.1. Baseline estimations

Table 2 presents the results of the baseline PPML estimations, with robust, country-pair clustered standard errors. I employ two alternative focus explanatory variables: the size of the bilateral diaspora and the share of the bilateral diaspora over the number of inventors in receiving countries. Columns (1) and (2) regress international co-patenting between inventors against these two focal variables separately, plus individual-country and time FE. The effect of the two variables is positive and statistically significant. Columns (3) and (4) further introduce a number of control variables. The focal variables remain statistically significant, although their point estimates are somewhat reduced. In particular, column (3) shows an elasticity of 0.18. That is to say, a 10-percent increase in the size of the inventor diaspora abroad is associated with a 1.8-percent increase in international patent collaborations, which is also economically meaningful. This result is of the same order of magnitude than estimates for the case of trade and diasporas (Head and Ries, 1998; Rauch and Trindade, 2002). The estimated coefficient for the diaspora variable as a proportion of the local inventors is of similar magnitude (in statistical terms). However, the exact interpretation of its elasticity is somewhat tricky – the variable is the log-transformation of a ratio. I therefore will focus my attention on the total diaspora coefficients hereafter.

[Table 2 about here]

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 to cooperate across national boundaries, though sharing a common border does not. Common language has a strong positive estimated effect on collaborations between inventors of different countries. However, historical links between country pairs expressed by their colonial past is not significant. As expected, bilateral trade is positive and significant, whilst technological distance between countries, i.e. how distant are countries in their technological specialization, exerts a negative influence on bilateral co-patents. Finally, both attribute variables – the total number of patents and the GDP per capita – are significant for the case of origin countries, but not for destinations. Thus, it appears that differences across developed countries in terms of technological and economic development are relatively minor and are picked up by their country FE.

In columns (5) and (6) I look at R&D offshoring – co-patents between applicants in developed countries and inventors in developing economies. By comparing the estimates with those of columns (3) and (4), interesting results emerge. First and foremost, the estimated elasticity of inventor diaspora size is notably reduced in these later estimations – less than a half as compared to columns (3) and (4). That is, diaspora networks particularly mediate interpersonal relations between co-workers. Meanwhile, they have a more nuanced effect on transnational employer-employee linkages.

Second of all, geography per se does not play a significant role to explain R&D offshoring. The diaspora and geography results put together seem to suggest that 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, but less important to explain more formal and hierarchical relationships – such as the ones represented by offshoring relationships, where probably explicit, written contracts are the rule.

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

Table 3 mimics estimations (3) through (6) of Table 2 but controlling for time-variant multilateral resistance. While country FE control for average multilateral resistance to collaborate over time (Feenstra, 2004), some elements of this multilateral resistance are likely to be time-variant and might not be picked up by the attribute variables included (Adam and Cobham, 2007). In consequence, Table 3 includes country FE plus country-specific time dummies, and repeats the main estimations, focusing the attention only on bilateral variables. Some nameable differences with respect to Table 2 emerge, like the non-significant role of distance in explaining inventor-to-inventor collaborations. However, the focal variables remain positive and strongly significant, and they present coefficients slightly larger than before.

[Table 3 about here]

4.2. Identification: cultural proximity and instrumental variables

Table 4 comes back to the baseline specifications (Table 2) and adds interaction terms between the inventor diaspora variable and different dimensions of cultural proximity between countries – common language and common colonial past. 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 culturally closer. Results (Table 4) partially confirm this extreme: the two interaction terms included are negative. Admittedly, though, only the interaction with colonial past is statistically significant, but not the interaction with common language.[12]

[12] The same estimation procedure using R&D offshoring as the dependent variable delivers similar results (negative coefficients of the interactions), but not significant. This is further evidence on the critical role

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. Indeed, a main concern of my analysis is the possibility of omitted variables driving both migration and co-patenting at the same time. However, I claim that the large list of control variables and FE included arguably reduces this possibility to a minor extent. Other factors, historical in nature, may remain unaccounted, which in turn might lead to biased estimates. However, such characteristics will likely affect the levels of co-patenting and migration, as opposed to recent changes, and could be largely controlled for including country-pair FE – for a discussion, see Parsons (2012). Subsection 4.5 addresses this point and introduces FE in the estimations. If unobserved confounding factors remain, they should work in such a way that they are capable to explain the main results – the diaspora-co-patenting relation – but also the differentiated effect of diaspora networks across different cultural dimensions, which I find unlikely.

[Table 4 about here]

For robustness, I also provide instrumental variables estimates and check the validity of the results. Plausible candidates to play such role are (i) the size of the bilateral diaspora between countries i and j in the 1960s – and its square (data from Özden et al., 2011), (ii) whether the two countries i and j were subject to a temporary guest-worker agreement in the 1960s and 1970s (Beine et al., 2011),[13] and (iii) the size of the unskilled diaspora original from country i residing in country j in 1990 (emigrants with only primary education) – and its square (data from Docquier et al., 2009). First, the stocks of migrants in the 1960s are likely to affect the current stocks of migrants and, in particular, the stock of high-skilled migrants, through network effects favoring further migration flows over the long run. Quite likely, they are uncorrelated with current levels of cross-country collaborations, aside from its influence through current skilled diasporas. In a similar vein, temporary guest-worker agreements are likely to exert a strong influence again on the formation of migrant stocks in those years – especially unskilled workers, hence once more influencing current flows of skilled workers through network effects. Finally, the current stocks of migrants with only primary

of diasporas for worker-to-worker collaborations, and their more nuanced effects for the case of more hierarchical, R&D offshoring relations.

[13] I thank Michel Beine for sharing the data on temporary guest-worker agreements.

education likely correlate with the current stocks of high-skilled diasporas. The relation between existing diasporas and existing migration flows does not only operate at the 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 – e.g., cultural amenities – which are attractive also to high-skilled immigrants (Hunt and Gauthier-Loiselle, 2008). On the other hand, uneducated migrants should play an inexistent role in boosting co-inventorship or R&D offshoring with their homelands – justifying their exclusion from the main equations, aside from their effects through inventor diasporas. Moreover, I use unskilled diaspora data taken from 1990 census – which accounts for 1980s unskilled migrant flows – so as to be more confident that they are unaffected by unobserved factors influencing co-patenting patterns between 1990 and 2010.

Columns (4) and (5) of Table 4 present GMM estimations of the PPML – see Windmeijer and Silva (1997). Note that the Shea partial R2 of the first stage is 0.499 and the value of the F-tests statistic, 1,102.94, is well above 10, which is usually considered a good threshold, and so the instruments cannot be judged as weak. Column (4) shows the 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 the magnitude of the estimated coefficient relative to the former PPML. I attribute this difference to the impact of technology internationalization on the incentives to migrate in the first place, as already commented in section 2. Increased job and business opportunities in diasporas' homelands would discourage migrants of leaving the country and would make the relationship between migration and international co-patenting negative, instead of positive. Likely, GMM estimates show the direct impact of diasporas on co-patenting, net of counterbalancing effects of co-inventorship on emigration from the origin countries. Thus, the 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 side, results for the case of R&D offshoring remain positive, but not significant (columns (5)). Again, this is further evidence on the critical role of high-skilled diasporas for worker-to-worker co-inventorship, and their less important effect for applicant-to-inventor relations.

4.3. Sectoral heterogeneity

The use of patent data allows me to identify sector-specific particularities in the relationship between international co-patenting and diaspora networks. By so doing, it is possible to gauge whether the former relationship holds at the industry level, but also, and more important, whether substantial differences across sectors emerge. 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.

Column (1) of Table 5 regresses bilateral co-inventorship between countries i and j in sector s – out of 5 aggregated sectors – on the presence of migrants from country i in country j in sector s – and includes sector FE. The estimated coefficient of inventor diasporas is slightly diminished as compared to column (3) of Table 2. However, it remains strongly significant and economically meaningful.

Columns (2) through (6) split collaborations and diaspora data into the five sectors and re-estimate the baseline model. The coefficient on the inventor diaspora is positive and statistically significant at 1-percent level in all the specifications. Moreover, and contrary to other variables, differences across sectors are not large, witnessing the importance of networks regardless of the technology being analyzed.

[Table 5 about here]

4.4. Are China and India different after all?

Next, I look at the robustness of my 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, i.e. the US, and its main providers of skilled talent – i.e., China, India, and other Asian economies. Thus, for instance, Agrawal's et al. (2011) provide results for the case of the Indian inventor diaspora in the US. Kerr (2008) extends the analysis to only nine ethnicities – Chinese, English, European, Hispanic, Indian, Japanese, Korean, Russian, and Vietnamese – finding that only the Chinese inventor diaspora successfully diffuse knowledge back to its homeland.

Saxenian (2006, 2002, 1999) studies the Indian and Chinese migrant entrepreneurs. On top of that, these and related studies generally focus only on one destination country, the US, whilst migration and international business networks are multi-country phenomena (Breschi et al., 2013). Other diaspora studies (trade, FDI…) centre only on the US too, or look at one single emerging country – e.g.., China. In light of this, some scholars argue that lessons from case studies of China and India cannot extrapolate to other migrant communities – that is, it is difficult to say whether high-skilled emigrants systematically engage in business networks and knowledge transfers with their homelands or rather the Indian and Chinese diasporas are so famous for being the exception rather than the rule (Gibson and McKenzie, 2012).

In order to explore this issue, Table 6 repeats the preferred estimations – with and without country-specific time dummies – but removing from the sample either the BRICS countries (Brazil, Russia, India, China and South Africa), the US, or both. Contrary to the arguments posit 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 as compared to the previous estimates.

[Table 6 about here]

4.5. Robustness analysis

To further check the robustness of my results, Table 7 runs the baseline specification using different estimation methods as well as a large number of FE. Column (1) runs zero-inflated Poisson (ZIP), in order to account for the excess of zeros of the dependent variable, including country and time FE. The estimated coefficient is slightly larger as compared to Table 2, but it is fairly comparable. OLS estimates including also individual country FE interacted with time FE deliver similar results concerning my focal variable (column (2)). Column (3) estimates the baseline model by means of linear panel data methods – with country-pair FE. The estimated coefficient of interest remains positive and strongly significant, although slightly diminished. Column (4) further includes country-pair FE and estimates the main model by means of PPML. The diaspora coefficient remains virtually unaltered. Finally, column (5) estimates again the model by means of PPML and introduces pairwise FE as well as country-specific time dummies. With the inclusion of these FE, this estimation controls for as much as confounding factors as possible and explains a large proportion of the variation – only three explanatory variables remain in the model when including these FE. Results show a larger coefficient for the focal variable as compared to the baseline estimations. The estimated difference goes in the same direction as the instrumental variables regressions. Therefore, I take the baseline regression as lower bound estimates.[14]

[Table 7 about here]

5. Conclusion

This paper examines the impact of high-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. In order to study this relationship, I make use of a unique dataset on inventors with migratory background. I claim that the use of these new data entails two main novelties with respect to the diaspora literature reviewed in section 2. First, inventor data capture diasporas at the upper tail of the skills distribution, which also constitutes a more homogeneous class of workers than the tertiary educated labor force as a whole. Second, as the data come from registered patents, they are available for a large number of sending and receiving countries, and on a longitudinal basis – as opposed to census-based data, which is only available every 10 years. Making use of these data, I study the relationship between inventor diasporas, on the one hand, and international co-inventorship and R&D offshoring, on the other. To my knowledge, there have been no 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 high-skilled diasporas and the internationalization of inventive activity between developed and developing countries. The effect is statistically and economically significant: a 10-percent increase

[14] Parsons (2012) estimates the relationship between migration and trade and, contrary to the present analysis, he finds no effects of the former over the latter once country-pair fixed effects are included to account for unobserved bilateral factors

.

in the inventor diaspora abroad is associated with a 1.5 to 2.2 percent increase in international patent collaborations at the level of inventors. The effect found is robust to the inclusion of a bunch of controls and FE, including individual country FE interacted with time FE. Given the variables included, the econometric approach – including instrumental variables estimates, and the robustness checks performed, I am fairly confident that my focal regressors do not pick up any confounding effect that might bias their point estimates. These findings do not suffice to conclude that a 'brain gain' that outreaches the loss of high-skilled human capital of sending economies exists, although they are undeniable necessary elements.

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, i.e., the US. The results also suggest that high-skilled diaspora effects are weakened 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 face-to-face relations and trust building are critical to explain co-inventorship – where contracts are usually more tacit and contract enforcement is difficult, but less important to explain more formal and hierarchical relationships – where probably explicit, written contracts are the rule. Further research, possibly at the firm and inventor levels, will shed more light on this particular issue.

References

Adam, C., Cobham, D., 2007. Modelling multilateral trade resistance in a gravity model with exchange rate regimes (CDMA Conference Paper Series No. 0702). Centre for Dynamic Macroeconomic Analysis.

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

Aleksynska, M., Peri, G., 2013. Isolating the Network Effect of Immigrants on Trade. The World Economy n/a–n/a.

Alnuaimi, T., Opsahl, T., George, G., 2012. Innovating in the periphery: The impact of local and foreign inventor mobility on the value of Indian patents. Research Policy 41, 1534–1543.

Beine, M., Docquier, F., Özden, Ç., 2011. Diasporas. Journal of Development Economics 95, 30–41.

Beine, M., Docquier, F., Rapoport, H., 2007. Measuring International Skilled Migration: A New Database Controlling for Age of Entry. World Bank Econ Rev 21, 249–254.

Breschi, S., Lissoni, F., Tarasconi, G., 2013. Inventor Data for Research on Migration & Innovation: A Survey and a Pilot.

Burger, M., van Oort, F., Linders, G.-J., 2009. On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation. Spatial Economic Analysis 4, 167–190.

Cameron, A.C., Trivedi, P.K., 1998. The analysis of count data. Cambridge University Press, Cambridge.

Carrington, W., Detragiache, E., 1998. How Big Is the Brain Drain? International Monetary Fund.

Chander, A., 2001. Diaspora Bonds'(2001). New York University Law Review 76, 1005–1043.

De Rassenfosse, G., Schoen, A., Wastyn, A., 2013. Selection bias in innovation studies: A simple test. Technological Forecasting and Social Change.

Docquier, F., Lodigiani, E., 2010. Skilled Migration and Business Networks. Open Econ Rev 21, 565–588.

Docquier, F., Lowell, B.L., Marfouk, A., 2009. A Gendered Assessment of Highly Skilled Emigration. Population and Development Review 35, 297–321.

Docquier, F., Marfouk, A., 2006. International Migration by Education Attainment (1990-2000)-Release 1.1. working paper.

Egger, P.H., von Ehrlich, M., Nelson, D.R., 2012. Migration and Trade. The World Economy 35, 216–241.

Feenstra, R.C., 2004. Advanced international trade: theory and evidence. Princeton University Press, Princeton, N.J.

Foley, C.F., Kerr, W.R., 2013. Ethnic Innovation and U.S. Multinational Firm Activity. Management Science.

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 109–133.

Gibson, J., McKenzie, D., 2012. The Economic Consequences of "Brain Drain" of the Best and Brightest: Microeconomic Evidence from Five Countries*. The Economic Journal 122, 339–375.

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, 302–316.

Guellec, D., van Pottelsberghe de la Potterie, B., 2001. The internationalisation of technology analysed with patent data. Research Policy 30, 1253–1266.

Guellec, D., Van Pottelsberghe de la Potterie, B., 2002. The Value of Patents and Patenting Strategies: Countries and Technology Areas Patterns. Economics of Innovation and New Technology 11, 133–148.

Hall, B.H., 2011. The internationalization of R&D (UNU-MERIT Working Paper Series No. 049). United Nations University, Maastricht Economic and social Research and training centre on Innovation and Technology.

Harhoff, D., Mueller, E., Reenen, J.V., 2013. What are the Channels for Technology Sourcing? Panel Data Evidence from German Companies (CEP Discussion Paper No. dp1193). Centre for Economic Performance, LSE.

Head, K., Ries, J., 1998. Immigration and Trade Creation: Econometric Evidence from Canada. The Canadian Journal of Economics / Revue canadienne d'Economique 31, 47–62.

Helliwell, J.F., 1998. How Much Do National Borders Matter? Brookings Institution Press.

Hoekman, J., Frenken, K., Tijssen, R.J.W., 2010. Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. Research Policy 39, 662–673.

Hunt, J., Gauthier-Loiselle, M., 2008. How Much Does Immigration Boost Innovation? (Working Paper No. 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, 984–1001.

Javorcik, B.S., Özden, Ç., Spatareanu, M., Neagu, C., 2011. Migrant networks and foreign direct investment. Journal of Development Economics 94, 231–241.

Jensen, P.H., Thomson, R., Yong, J., 2011. Estimating the patent premium: Evidence from the Australian Inventor Survey. Strategic Management Journal 32, 1128–1138.

Jonkers, K., Cruz-Castro, L., 2013. Research upon return: The effect of international mobility on scientific ties, production and impact. Research Policy 42, 1366–1377.

Kapur, D., McHale, J., 2005. Give us your best and brightest: The global hunt for talent and its impact on the developing world. Brookings Inst Press.

Katz, J.S., Martin, B.R., 1997. What is research collaboration? Research Policy 26, 1–18.

Kerr, W.R., 2008. Ethnic Scientific Communities and International Technology Diffusion. Review of Economics and Statistics 90, 518–537.

Kugler, M., Levintal, O., Rapoport, H., 2013. Migration and Cross-Border Financial Flows (CReAM Discussion Paper Series No. 1317). Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London.

Kugler, M., Rapoport, H., 2005. Skilled Emigration, Business Networks and Foreign Direct Investment (CESifo Working Paper Series No. 1455). CESifo Group Munich.

Kugler, M., Rapoport, H., 2007. International labor and capital flows: Complements or substitutes? Economics Letters 94, 155–162.

Mayer, T., Zignago, S., 2011. Notes on CEPII's distances measures: the GeoDist database [WWW Document]. URL http://mpra.ub.uni-muenchen.de/36347/ (accessed 9.2.13).

McCallum, J., 1995. National Borders Matter: Canada-U.S. Regional Trade Patterns. American Economic Review 85, 615–23.

Miguelez, E., Fink, C., 2013. Measuring the International Mobility of Inventors: A New Database (WIPO Economic Research Working Paper No. 8). World Intellectual Property Organization - Economics and Statistics Division.

Montobbio, F., Sterzi, V., 2013. The Globalization of Technology in Emerging Markets: A Gravity Model on the Determinants of International Patent Collaborations. World Development 44, 281–299.

Nanda, R., Khanna, T., 2010. Diasporas and Domestic Entrepreneurs: Evidence from the Indian Software Industry. Journal of Economics & Management Strategy 19, 991–1012.

Özden, Ç., Parsons, C.R., Schiff, M., Walmsley, T.L., 2011. Where on Earth is Everybody? World Bank Economic Review 25, 12–56.

Parsons, C.R., 2012. Do migrants really foster trade ? the trade-migration nexus, a panel approach 1960-2000 (Policy Research Working Paper Series No. 6034). The World Bank.

Patel, P., Pavitt, K., 1991. Large Firms in the Production of the World's Technology: An Important Case of "Non-Globalisation". Journal of International Business Studies 22, 1–21.

Patel, P., Vega, M., 1999. Patterns of internationalisation of corporate technology: location vs. home country advantages. Research Policy 28, 145–155.

Picci, L., 2010. The internationalization of inventive activity: A gravity model using patent data. Research Policy 39, 1070–1081.

Rauch, J., 2003. Diasporas and Development: Theory, Evidence, and Programmatic Implications. Department of Economics, University of California at San Diego.

Rauch, J.E., 2001. Business and Social Networks in International Trade. Journal of Economic Literature 39, 1177–1203.

Rauch, J.E., Trindade, V., 2002. Ethnic Chinese Networks in International Trade. Review of Economics and Statistics 84, 116–130.

Regets, M., 2001. Research and Policy Issues in High-Skilled International Migration: A Perspective with Data from the United States (IZA Discussion Paper No. 366). Institute for the Study of Labor (IZA).

Santos Silva, J.M.C., Tenreyro, S., 2006. The Log of Gravity. Review of Economics and Statistics 88, 641–658.

Santos Silva, J.M.C., Tenreyro, S., 2010. On the existence of the maximum likelihood estimates in Poisson regression. Economics Letters 107, 310–312.

Saxenian, A., 1999. Silicon Valley's New Immigrant Entrepreneurs.

Saxenian, A., 2002. Silicon Valley's New Immigrant High-Growth Entrepreneurs. Economic Development Quarterly 16, 20–31.

Saxenian, A., 2006. The New Argonauts: Regional Advantage in a Global Economy. Harvard University Press.

Saxenian, A., Motoyama, Y., Quan, X., 2002. Local and global networks of immigrant professionals in Silicon Valley. Public Policy Institute of California, San Francisco, CA.

Saxenian, A., Sabel, C., 2008. Roepke Lecture in Economic Geography Venture Capital in the "Periphery": The New Argonauts, Global Search, and Local Institution Building. Economic Geography 84, 379–394.

Scellato, G., Franzoni, C., Stephan, P., 2012. Mobile Scientists and International Networks (Working Paper No. 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., de Rassenfosse, G., Webster, E., 2013. Quantifying the effect of R&D offshoring on industrial productivity at home. Mimeo, University of Melbourne.

Thursby, J.G., Thursby, M.C., 2006. Here or There? A Survey of Factors in Multinational R&D Location: Report to the Government/University/Industry Research Roundtable (SSRN Scholarly Paper No. ID 1456460). Social Science Research Network, Rochester, NY.

Van Zeebroeck, N., van Pottelsberghe de la Potterie, B., 2011. Filing strategies and patent value. Economics of Innovation and New Technology 20, 539–561.

Windmeijer, F.A.G., Silva, J.M.C.S., 1997. Endogeneity in Count Data Models: An Application to Demand for Health Care. Journal of Applied Econometrics 12, 281–94.

Appendix 1.

List of developed countries

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

List of developing/emerging/transition countries

Armenia, Angola, Argentina, Azerbaijan, Bangladesh, Burkina Faso, Bulgaria, Bolivia, Brazil, Botswana, Belarus, Democratic Republic of the Congo, Central African Republic, Congo, Cote d'Ivoire, Chile, Cameroon, China, Colombia, Costa Rica, Czech Republic, Algeria, Ecuador, Estonia, Egypt, Eritrea, Ethiopia, Georgia, Ghana, Guinea, Greece, Guatemala, Hong Kong (SAR China), Honduras, Hungary, Indonesia, Israel, India, Iraq, Iran, Jordan, Kenya, Kyrgyzstan, Cambodia, Kazakhstan, Lao People's Democratic Republic, Lebanon, Liberia, Lithuania, Latvia, Libyan Arab Jamahiriya, Morocco, Moldova, Madagascar, Mali, Myanmar, Mongolia, Macau (SAR China), Mauritania, Malawi, Mexico, Malaysia, Mozambique, Namibia, Niger, Nigeria, Nicaragua, Panama, Peru, Philippines, Pakistan, Poland, Paraguay, Romania, Russia, Saudi Arabia, Singapore, Slovenia, Slovakia, Senegal, El Salvador, Syrian Arab Republic, Chad, Thailand, Tajikistan, Turkmenistan, Tunisia, Turkey, United Republic of Tanzania, Ukraine, Uganda, Uruguay, Uzbekistan, Venezuela, Viet Nam, South Africa, Zambia, and Zimbabwe.

Appendix 2.

Table A.2.1. Summary statistics

	Observations	Mean	St. Dev	Min.	Max.
Collab. inv_i- inv_j	38720	0.83	8.69	0	678
Collab. app_i- inv_j	38720	1.15	12.25	0	708
Diapora size	38720	14.72	367.69	0	26661
Diapora share	38720	0.00	0.00	0	0.05
Distance	38720	7453.09	4077.03	59.62	19629.50
Contiguity	38720	0.00	0.07	0	1
Common language	38720	0.11	0.31	0	1
Colonial links	38720	0.04	0.19	0	1
EXP+IMP	38720	0.01	0.02	0	0.41
Tech.distance	38720	0.62	0.30	0.02	1
# patents_i	38720	543.66	2903.53	0	64990
# patents_j	38720	42919.68	99161.05	52	692364
GDP p.c._i	37540	6324.47	7167.66	140.02	49876.90
GDP p.c._j	38720	29369.81	6177.65	11382.60	48799.70

Notes: '_i' and '_j' stand for, respectively, developing country and developed country.

Table A.2.2. Correlation matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1													
2	0.93	1												
3	0.38	0.37	1											
4	0.29	0.29	0.84	1										
5	0.00	0.01	-0.17	-0.21	1									
6	0.04	0.02	0.11	0.13	-0.23	1								
7	0.05	0.06	0.08	0.11	0.12	-0.02	1							
8	0.01	0.02	0.12	0.11	-0.03	0.14	0.38	1						
9	0.10	0.10	0.34	0.26	-0.15	0.08	0.04	0.11	1					
10	-0.16	-0.17	-0.48	-0.49	0.20	-0.09	0.02	-0.01	-0.34	1				
11	0.20	0.21	0.53	0.54	-0.22	0.11	-0.07	-0.01	0.38	-0.79	1			
12	0.14	0.14	0.43	0.17	0.00	0.01	-0.02	0.05	0.35	-0.16	0.19	1		
13	0.07	0.07	0.26	0.27	-0.19	0.09	-0.14	-0.02	0.30	-0.53	0.63	0.07	1	
14	0.09	0.09	0.27	0.17	-0.11	0.03	0.01	-0.05	0.14	-0.17	0.19	0.52	0.07	1

Notes: 1. Collab. inv_i- inv_j; 2. Collab. app_i- inv_j; 3. ln(Diapora size); 4. ln(Diapora share); 5. ln(Distance); 6. Contiguity; 7. Common language; 8. Colonial links; 9. ln(EXP+IMP); 10. ln(tech. distance); 11. ln(# patents)_i; 12. ln(# patents)_j; 13. ln(GDP p.c.)_i; 14. ln(GDP p.c.)_j

Figure 1. Share of immigrant inventors over time, 1985-2010, by selected countries

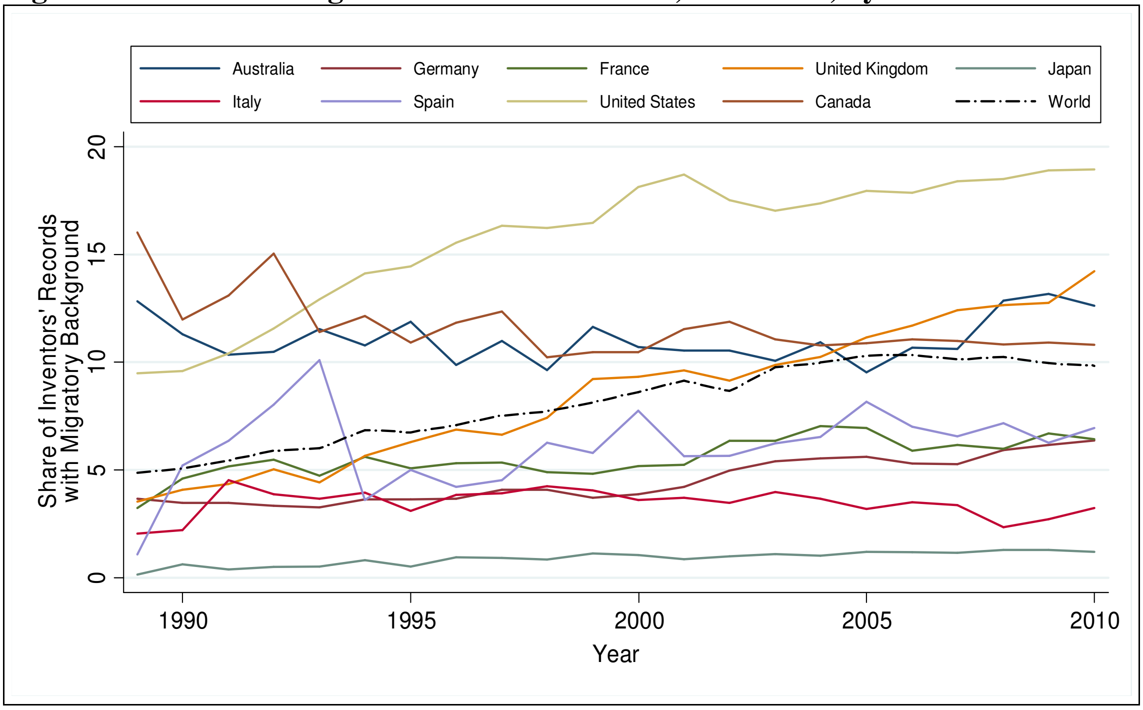

Figure 2. Immigration rates of inventors, 2001-2010, receiving countries

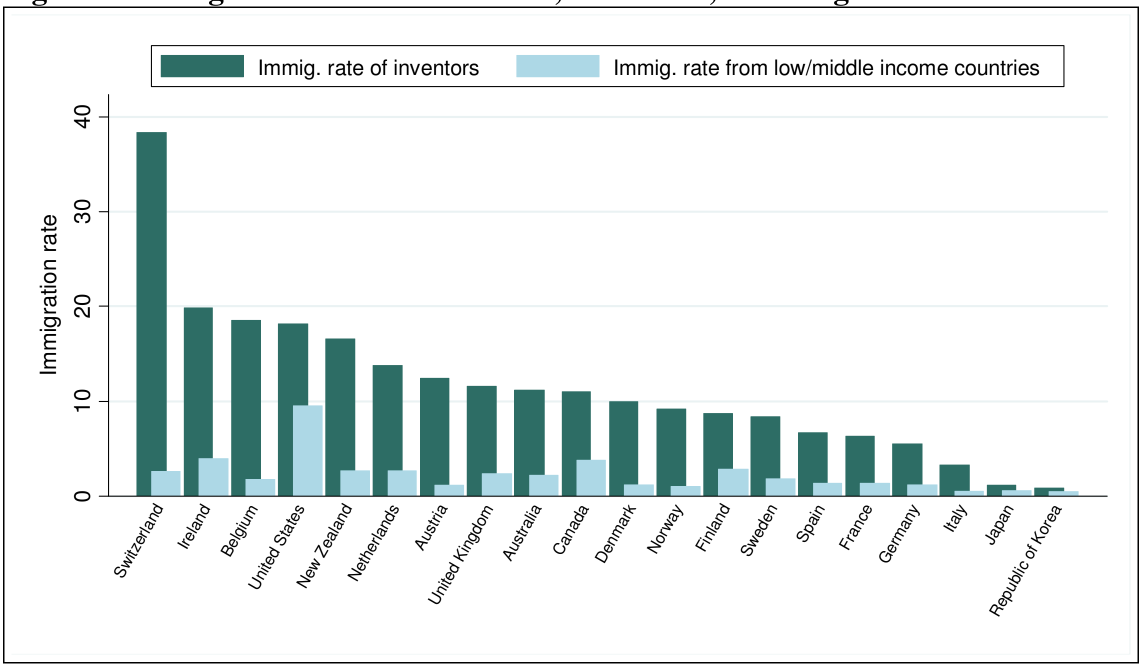

Figure 3. Bilateral corridors: shares across world areas, 1990-2010

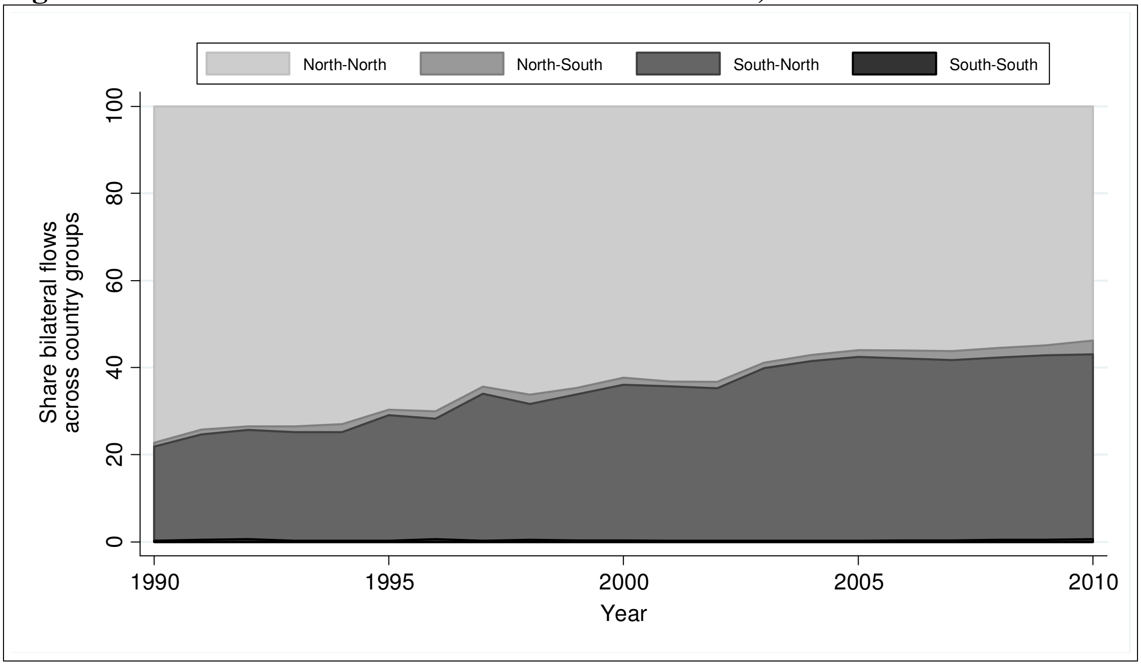

Figure 4. International Collaborations of Inventors in PCT patents, 1990-2010

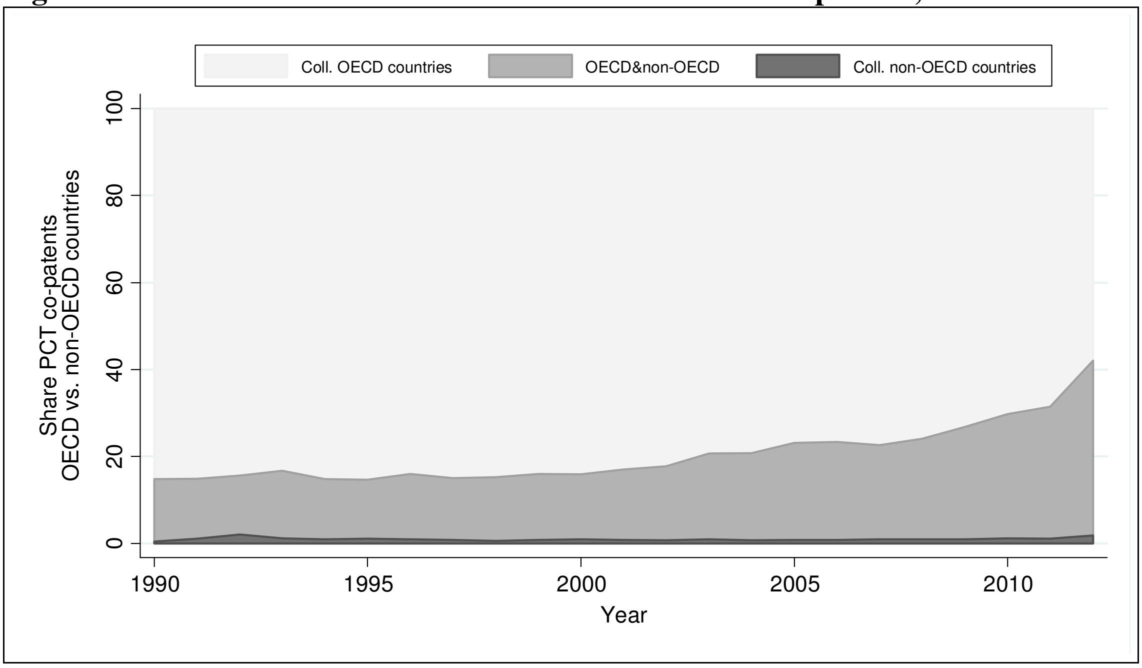

Table 1.

Panel a. Share of immigrant inventors over total immigrants, 2001-2010

Total migrants			South-North migrants		
Country code	Immigrants	Share total immigrants	Country code	Immigrants	Share total immigrants
United States	194,609	57.17	United States	105,336	74.87
Germany	25,341	7.44	Germany	6,031	4.29
Switzerland	20,416	6.00	Singapore	4,375	3.11
U.K.	15,758	4.63	Japan	3,927	2.79
Netherlands	9,665	2.84	U.K.	3,729	2.65
France	9,540	2.80	Canada	2,503	1.78
Canada	7,257	2.13	France	2,230	1.59
Singapore	6,720	1.97	Netherlands	2,128	1.51
Japan	6,715	1.97	Switzerland	1,451	1.03
Belgium	5,042	1.48	Finland	1,265	0.90

Panel b. Top-20 most populated corridors, 2001-2010

Largest inventor migration corridors			Largest inventor migration corridors, limited to non-OECD sending countries		
Origin	Destination	Counts	Origin	Destination	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

Table 2. PPML baseline specifications

	(1)	(2)	(3)	(4)	(5)	(6)
	Inventor-to-inventor co-patents				Applicant-to-inventor co-patents	
ln(Diaspora)	0.325***		0.181***		0.0858**	
	(0.0307)		(0.0248)		(0.0402)	
ln(Diaspora share)		0.415***		0.286***		0.170***
		(0.0340)		(0.0268)		(0.0493)
ln(Distance)			-0.275***	-0.239***	-0.0977	-0.0684
			(0.0686)	(0.0674)	(0.0885)	(0.0890)
Contiguity			-0.0248	0.0122	-0.143	-0.103
			(0.125)	(0.122)	(0.220)	(0.224)
Common language			0.534***	0.501***	0.743***	0.715***
			(0.115)	(0.112)	(0.187)	(0.189)
Colonial links			0.166	0.148	0.374**	0.356**
			(0.131)	(0.126)	(0.172)	(0.181)
ln(EXP+IMP)			0.0720***	0.0552***	0.0901***	0.0748**
			(0.0236)	(0.0204)	(0.0305)	(0.0291)
ln(Tech.distance)			-0.0963**	-0.0887**	-0.277***	-0.269***
			(0.0431)	(0.0431)	(0.0567)	(0.0563)
ln(# patents) orig.			0.321***	0.331***	0.344***	0.343***
			(0.0581)	(0.0513)	(0.0734)	(0.0696)
ln(# patents) dest.			0.0297	0.0994	0.368	0.408
			(0.135)	(0.139)	(0.254)	(0.266)
ln(GDP p.c.) orig.			1.224***	1.218***	1.851***	1.834***
			(0.241)	(0.197)	(0.335)	(0.310)
ln(GDP p.c.) dest.			-0.394	-0.933	-0.925	-1.247
			(0.593)	(0.607)	(0.873)	(0.870)
Constant	-1.713***	2.953***	-5.618	1.792	-11.84	-7.464
	(0.284)	(0.369)	(6.510)	(6.674)	(9.716)	(9.745)
Observations	32,382	32,382	31,680	31,680	32,400	32,400
Pseudo R2	0.922	0.916	0.959	0.960	0.915	0.913
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Log Lik	-18625.83	-18387.70	-17121.66	-16909.26	-22510.90	-22357.78

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. The different number of final observations between Table 3 and Table A.2.1 is due to the inclusion of FE in pseudo-maximum likelihood estimations: the PPML method automatically drops the country-specific FE (and their corresponding observations) for which the country has zero recorded co-patents 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 Tenreyro, 2010, for further details).

Table 3. PPML baseline specifications with time-varying multilateral resistance

	(1)	(2)	(3)	(4)
	Inventor-to-inventor co-patents		Applicant-to-inventor co-patents	
ln(Diaspora)	0.227***		0.121***	
	(0.0260)		(0.0396)	
ln(Diaspora share)		0.263***		0.121***
		(0.0256)		(0.0441)
ln(Distance)	-0.0263	-0.0707	0.137*	0.114
	(0.0599)	(0.0608)	(0.0784)	(0.0802)
Contiguity	-0.168	-0.182	-0.299	-0.312
	(0.115)	(0.113)	(0.207)	(0.208)
Common language	0.316***	0.320***	0.540***	0.548***
	(0.0976)	(0.0944)	(0.161)	(0.161)
Colonial links	0.158	0.180*	0.334**	0.348**
	(0.106)	(0.108)	(0.146)	(0.145)
ln(EXP+IMP)	0.257***	0.239***	0.307***	0.306***
	(0.0390)	(0.0397)	(0.0534)	(0.0552)
ln(Tech.distance)	-0.185***	-0.200***	-0.341***	-0.349***
	(0.0501)	(0.0504)	(0.0719)	(0.0723)
Constant	-1.089*	2.117***	-3.214***	-1.655**
	(0.607)	(0.596)	(0.760)	(0.831)
Observations	20,757	20,757	23,300	23,300
Pseudo R2	0.978	0.977	0.953	0.953
Origin FE	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes
Year FE	No	No	No	No
Origin FE*Time FE	Yes	Yes	Yes	Yes
Destination FE*Time FE	Yes	Yes	Yes	Yes
Log Lik	-15623.24	-15579.46	-20011.80	-20027.68

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. The different number of final observations between columns (1) and (2) 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 Tenreyro, 2010, for further details).

Table 4. Cultural proximity and instrumental variables (GMM)

	(1)	(2)	(3)	(4)	(5)
	Inventor-to-inventor co-patents				Applicant-to-inventor
	PPML	PPML	PPML	GMM	GMM
ln(Diaspora)	0.187***	0.174***	0.182***	0.226***	0.106
	(0.0260)	(0.0252)	(0.0260)	(0.0759)	(0.141)
ln(Distance)	-0.249***	-0.257***	-0.252***	-0.207***	-0.152*
	(0.0555)	(0.0546)	(0.0555)	(0.0637)	(0.0886)
Contiguity	0.0141	0.0185	0.0336	0.00230	-0.147
	(0.122)	(0.118)	(0.119)	(0.123)	(0.204)
Common language	0.657***	0.528***	0.639***	0.497***	0.714***
	(0.199)	(0.113)	(0.201)	(0.115)	(0.208)
Colonial links	0.126	0.626**	0.583**	0.169	0.302*
	(0.139)	(0.246)	(0.262)	(0.125)	(0.173)
ln(Diaspora)* Lang.	-0.0216		-0.0189		
	(0.0224)		(0.0228)		
ln(Diaspora)*Colonial		-0.120**	-0.117**		
		(0.0486)	(0.0497)		
ln(EXP+IMP)	0.0713***	0.0728***	0.0715***	0.100***	0.0772*
	(0.0238)	(0.0238)	(0.0237)	(0.0335)	(0.0400)
ln(Tech.distance)	-0.0825*	-0.0825*	-0.0850*	-0.0992**	-0.279***
	(0.0475)	(0.0471)	(0.0470)	(0.0463)	(0.0601)
ln(# patents) orig.	0.328***	0.324***	0.330***	0.259***	0.349***
	(0.0550)	(0.0554)	(0.0552)	(0.0486)	(0.0784)
ln(# patents) dest.	0.0587	0.0383	0.0328	0.0515	0.246
	(0.116)	(0.118)	(0.116)	(0.122)	(0.190)
ln(GDP p.c.) orig.	1.134***	1.149***	1.127***	1.366***	1.728***
	(0.239)	(0.246)	(0.242)	(0.213)	(0.329)
ln(GDP p.c.) dest.	-0.0195	0.0806	0.0994	-0.0976	-1.723**
	(0.575)	(0.569)	(0.571)	(0.603)	(0.845)
Constant	-9.729	-10.48*	-10.58*	-10.69	-0.381
	(6.377)	(6.294)	(6.319)	(6.537)	(8.673)
Observations	31,680	31,680	31,680	37,540	37,540
Pseudo R2	0.958	0.957	0.958		
Origin FE	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
F-stat 1st stage				1,102.94	1,102.94
Shea partial R2				0.499	0.5549
Log Lik	-17948.779	-17933.811	-17930.356		

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.

Table 5. PPML specifications: sectoral level

	(1)	(2)	(3)	(4)	(5)	(6)
	Inventor-to-inventor co-patents					
	Pooled	Electrical engineering	Instruments	Chemistry	Mechanical	Other sectors
ln(Diaspora)	0.151***	0.221***	0.169***	0.161***	0.159***	0.118***
	(0.0229)	(0.0443)	(0.0290)	(0.0238)	(0.0367)	(0.0435)
ln(Distance)	-0.290***	-0.0931	-0.391***	-0.298***	-0.434***	-0.744***
	(0.0615)	(0.0881)	(0.0673)	(0.0695)	(0.0826)	(0.129)
Contiguity	0.00164	0.0308	0.289*	-0.240	0.310	-0.342
	(0.127)	(0.247)	(0.151)	(0.177)	(0.198)	(0.310)
Common language	0.522***	0.255	0.834***	0.455***	0.545***	0.676***
	(0.111)	(0.216)	(0.123)	(0.118)	(0.137)	(0.217)
Colonial links	0.150	0.104	-0.0714	0.266*	0.159	0.473**
	(0.128)	(0.188)	(0.155)	(0.158)	(0.164)	(0.202)
ln(EXP+IMP)	0.0784***	0.166***	0.0628*	0.0719***	0.0234	0.0689
	(0.0291)	(0.0537)	(0.0357)	(0.0265)	(0.0467)	(0.0511)
ln(Tech.distance)	-0.0905*	-0.230***	0.0149	0.0485	-0.0599	-0.418***
	(0.0482)	(0.0719)	(0.0527)	(0.0517)	(0.0696)	(0.132)
ln(# patents) orig.	0.432***	0.590***	0.519***	0.491***	0.526***	0.477***
	(0.0491)	(0.0748)	(0.0574)	(0.0483)	(0.0957)	(0.0905)
ln(# patents) dest.	0.536***	0.00829	0.681***	0.427***	0.501***	0.565
	(0.0767)	(0.190)	(0.199)	(0.121)	(0.184)	(0.371)
ln(GDP p.c.) orig.	0.807***	0.296	-0.0423	0.874***	0.370	0.948**
	(0.183)	(0.398)	(0.285)	(0.191)	(0.299)	(0.454)
ln(GDP p.c.) dest.	-0.119	2.585**	0.414	-1.837***	-1.176	-4.222**
	(0.641)	(1.033)	(1.078)	(0.562)	(1.009)	(1.977)
Constant	-13.26*	-33.21***	-11.72	6.705	4.389	33.93
	(7.080)	(12.04)	(12.14)	(5.902)	(11.49)	(21.91)
Observations	172,000	25,920	25,600	33,200	24,660	20,560
Pseudo R2	0.864	0.933	0.891	0.912	0.808	0.711
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	No	No	No	No	No
Log Lik	-33037.36	-7155.51	-5469.42	-11221.89	-4791.32	-2447.52

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.

Table 6. Are China and India the exception rather than the rule?

	(1)	(2)	(3)	(4)
	Inventor-to-inventor co-patents			
	No BRICS	No US	No BRICS, no US	No BRICS, no US
ln(Diaspora)	0.159***	0.204***	0.188***	0.191***
	(0.0392)	(0.0372)	(0.0433)	(0.0469)
ln(Distance)	-0.381***	-0.323***	-0.465***	-0.265***
	(0.0680)	(0.0617)	(0.0706)	(0.0827)
Contiguity	-0.128	-0.224	-0.284*	-0.332**
	(0.124)	(0.145)	(0.160)	(0.157)
Common language	0.667***	0.285**	0.293*	0.216
	(0.166)	(0.128)	(0.157)	(0.150)
Colonial links	0.139	0.360***	0.407***	0.341**
	(0.160)	(0.120)	(0.148)	(0.142)
ln(EXP+IMP)	0.108***	0.0281	0.0595**	0.226***
	(0.0259)	(0.0211)	(0.0240)	(0.0468)
ln(Tech.distance)	-0.122**	-0.0163	0.0309	-0.0121
	(0.0604)	(0.0667)	(0.0685)	(0.0816)
ln(# patents) orig.	0.0939*	0.313***	0.103*	
	(0.0493)	(0.0584)	(0.0594)	
ln(# patents) dest.	0.133	0.00370	0.0828	
	(0.111)	(0.117)	(0.112)	
ln(GDP p.c.) orig.	0.642*	1.113***	0.470**	
	(0.383)	(0.234)	(0.240)	
ln(GDP p.c.) dest.	0.164	0.00949	0.309	
	(0.699)	(0.620)	(0.719)	
Constant	-7.193	-9.638	-5.097	0.127
	(7.926)	(6.720)	(7.906)	(0.764)
Observations	33,620	32,680	30,799	15,820
Pseudo R2	0.901	0.834	0.664	0.728
Origin FE	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	No
Origin FE*Time FE	No	No	No	Yes
Destination FE*Time FE	No	No	No	Yes
Log Lik	-13615.12	-14899.17	-11180.14	-9835.92

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.

Table 7. Robustness checks. Inventor-to-inventor co-patents

	(1)	(2)	(3)	(4)	(5)
	ZIP	OLS	FE	PPML	PPML
ln(Diaspora)	0.251***	0.233***	0.172***	0.219***	0.261***
	(0.0252)	(0.0159)	(0.0105)	(0.0476)	(0.0591)
ln(Distance)	-0.0846	0.00813			
	(0.0600)	(0.0146)			
Contiguity	0.122	0.318			
	(0.117)	(0.204)			
Common language	0.489***	0.00133			
	(0.111)	(0.0286)			
Colonial links	-0.0459	-0.0151			
	(0.118)	(0.0400)			
ln(EXP+IMP)	0.0694***	0.0129***	-0.00699***	0.0825***	0.316***
	(0.0266)	(0.00310)	(0.00116)	(0.0237)	(0.0561)
ln(Tech.distance)	-0.0663	-0.148***	0.0802***	0.0533	-0.167*
	(0.0424)	(0.0343)	(0.0155)	(0.0531)	(0.0911)
ln(# patents) orig.	0.332***		0.0785***	0.535***	
	(0.0593)		(0.00720)	(0.0396)	
ln(# patents) dest.	-0.220		-0.00631	0.275**	
	(0.135)		(0.00968)	(0.108)	
ln(GDP p.c.) orig.	1.231***		0.155***	0.109	
	(0.245)		(0.0281)	(0.177)	
ln(GDP p.c.) dest.	-0.618		-0.172***	-0.0123	
	(0.709)		(0.0530)	(0.579)	
Constant	-1.765	-4.704***	0.411	-7.563	-1.327**
	(7.613)	(1.319)	(0.556)	(6.314)	(0.597)
Observations	37,540	37,540	37,540	21,301	17,619
Pseudo R2		0.638	0.219	0.963	0.986
Origin FE	Yes	Yes	No	No	No
Destination FE	Yes	Yes	No	No	No
Year FE	Yes	No	Yes	Yes	No
Origin FE*Time FE	No	Yes	No	No	Yes
Dest. FE*Time FE	No	Yes	No	No	Yes
Country-Pair FE	No	No	Yes	Yes	Yes
Log Lik	-17096.75	-11930.79	-6189.99	-16022.92	-13686.51

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.