



Contents lists available at ScienceDirect

Journal of International Economics

journal homepage: www.elsevier.com/locate/jie

Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion?

Dany Bahar^{a,b,*}, Ricardo Hausmann^{a,b}, Cesar A. Hidalgo^c

^a Harvard Kennedy School, United States

^b Center for International Development at Harvard University, United States

^c Macro Connections Group at MIT Media Lab, United States

ARTICLE INFO

Article history:

Received 26 March 2012

Received in revised form 31 October 2013

Accepted 1 November 2013

Available online xxx

JEL classification:

O31

O33

F10

F62

F63

Keywords:

Export similarity

Productivity

Knowledge

Technology

Diffusion

Spillovers

ABSTRACT

The literature on knowledge diffusion shows that knowledge decays strongly with distance. In this paper we document that the probability that a product is added to a country's export basket is, on average, 65% larger if a neighboring country is a successful exporter of that same product. For existing products, growth of exports in a country is 1.5% higher per annum if it has a neighbor with comparative advantage in these products. While these results could be driven by a common third factor that escapes our controls, they align with our expectations of the localized character of knowledge diffusion.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Knowledge has become central to modern theories of growth. Knowledge is embodied in goods that are then shipped around at a cost. When these goods are imported, they accelerate productivity growth in the recipient country (e.g. Rivera-Batiz and Romer, 1990; Coe and Helpman, 1995; Coe et al., 2009). However, significant parts of knowledge are disembodied or tacit (Polanyi, 1962) and its diffusion requires more direct forms of human interaction, which inevitably limits its scope to more localized or idiosyncratic settings (Arrow, 1969).

Previous research has documented the rapid decay of knowledge diffusion with geographic distance. This literature looked at the impact of distance on the patterns of patent citation (e.g. Jaffe et al., 1993), of R&D and patent output (e.g. Branstetter, 2001; Bottazzi and Peri, 2003), of R&D and productivity (Keller, 2002), and on the sales of subsidiaries of multinational corporations (Keller and Yeaple, 2013). Keller (2002, 2004) has shown that foreign sources of technology

account for up to 90% of domestic productivity growth and that the impact is highly localized.

What are the implications of rapid geographic decay of knowledge diffusion for the patterns of comparative advantage of countries? Ricardian models of trade argue that trade patterns are the reflection of productivity differences: countries export the goods in which they are relatively more productive—i.e. goods in which they exhibit comparative advantage. In this framework, countries become exporters of new goods or increase their market share in existing goods because they become more productive in them. If knowledge drives productivity and diffuses at short distances, then telltale signs should be observable in the geographic patterns of comparative advantage both statically and dynamically. In particular, neighboring countries should share more knowledge and hence have more similar static patterns of comparative advantage, in which case they should exhibit a geographically correlated pattern of product adoption and export growth.

In this paper, we use a novel setting to explore the diffusion of industry-specific productivity increases: the export baskets of countries. The key assumption is that, controlling for product-specific shifts in global demand, firms in a country will be able to incorporate a new good into their export basket only after they have become productive enough to

* Corresponding author. Harvard University, 79 JFK St, Cambridge MA 02138, United States.

E-mail address: dbahar@fas.harvard.edu (D. Bahar).

compete in global markets. Additionally, in order to increase their market share, firms will also need to become more productive. If knowledge diffusion decays strongly with distance, countries with the relevant knowledge should induce shifts in productivity in their neighbors—we explore this in both a static and a dynamic setting. We study both the intensive and the extensive margins of exports, exploring whether neighbors matter in affecting the ability of a country to gain market share or to become productive enough to export a product for the first time. As has been shown, the extensive margin accounts for a significant fraction of the growth of global trade in the last decades (Zahler, 2007; Kehoe and Ruhl, 2013). We therefore also explore the intensive margin, looking at the impact of neighbors in the evolution of a country's market share.

From a static perspective, we find that the export baskets of neighbors are remarkably similar, even after controlling for similarity in size, level of development, culture, institutional setting and factor endowments, among other controls: sharing a border and a region makes countries two standard deviations more similar than the average. From a dynamic perspective, we find that—after controlling for all time-varying sources of aggregate similarity between pairs of countries, for time varying product characteristics and for a country's own predisposition to adopt a product—countries are 65% more likely to start exporting a product which was being exported with comparative advantage by one of its geographic neighbors at the beginning of the period.

This result is not obvious. After all, gravity models have shown that, *ceteris paribus*, trade is more intense at short distances (Tinbergen, 1963; Bergstrand, 1985; Leamer and Levinshohn, 1995; Frankel, 1997). Hence, we should expect neighbors to specialize in different industries in order to exploit their comparative advantage and benefit from the gains of trade. The higher intensity of trade at short distances should force specialization and differentiation, whether—as pointed out by Feenstra, Markusen and Rose (2001)—the differences causing specialization arise as a result of an Armington structure of demand (e.g. Anderson, 1979; Bergstrand, 1985; Deardorff, 1998), economies of scale (e.g. Helpman and Krugman, 1985; Bergstrand, 1989), technological differences across countries (e.g. Davis, 1995; Eaton and Kortum, 1997), differences in factor endowments (e.g. Deardorff, 1998); or whether they arise from reciprocal dumping in models of homogeneous goods, imperfect competition and segmented markets (e.g. Brander, 1981; Brander and Krugman, 1983; Venables, 1985).

We can understand our results in the context of an endogenous Ricardian framework, where comparative advantage evolves with the progressive acquisition of knowledge or technologies which diffuse geographically.¹ However, under such a Ricardian framework, a reasonable question to ask is, what aspects of technology have limited tradability so that geography could be a defining factor in its diffusion pattern? Clearly, the technology that is embodied in machines and tradable goods and services should diffuse more broadly: after all, cellphones are available everywhere. However, tacit knowledge (Polanyi, 1962)—knowledge that is disembodied and hard to codify and teach because it cannot be captured by blueprints or instruction manuals—should diffuse with more difficulty. How does tacit knowledge diffuse? As mentioned above, Kenneth Arrow argued that knowledge diffusion requires more direct forms of human interaction, which limits its scope to more localized or idiosyncratic settings (Arrow, 1969). Furthermore, the emerging consensus in the literature of knowledge diffusion is that diffusion occurs predominantly within a fairly short range (e.g. Jaffe et al., 1993; Branstetter, 2001; Keller, 2002; Bottazzi and Peri, 2003), an observation that is attributed to the characteristics of tacit knowledge. Hence, if indeed knowledge

diffusion translates into productivity shifts that can shape the export basket of countries, then, in a world in which knowledge diffuses preferentially at short ranges, a country's export basket—as well as its evolution—will be shaped by the knowledge available in its neighborhood.

The localized nature of knowledge diffusion should generate the observables that we document in this paper. In particular, if knowledge has been homogenized preferentially at shorter distances, a snapshot view of the export basket of countries (a realization of their comparative advantage) should resemble that of their neighbors. Dynamically, we should also observe a geographically correlated pattern of adoption of new export goods and of changes in market shares. In this interpretation, there is a causal link between the presence of productive knowledge in a country and its diffusion to a neighbor. However, there is always the possibility that these correlated events may be caused by a third factor that is common to neighboring countries and that explains both the static similarity and the time-lapsed pattern of adoption without there being a causal link between the two. We will try to control, as best we can, for these alternative channels but we do not claim to have ruled them out completely. We discuss this more in detail in the body of the paper.

Until now, the burgeoning literature on international knowledge diffusion has relied on three main indicators to measure knowledge acquisition: patent citations (e.g. Jaffe et al., 1993), patent output (e.g. Bottazzi and Peri, 2003; Branstetter, 2006) and changes in total factor productivity (e.g. Coe and Helpman, 1995; Keller, 2002; Keller and Yeaple, 2009). One contribution of this paper consists in bringing to the literature a more tangible measure of knowledge acquisition: the ability of a country to achieve or improve its comparative advantage in the export of goods.

This paper is organized as follows. In the next section we discuss our sample and present a set of stylized facts based on the static export similarity between countries. In Section 3 we study the dynamics of this process. Section 4 discusses the results and Section 5 presents concluding remarks.

2. Data and stylized facts

2.1. Data

Data on exports in the period 1962–2000 comes from the World Trade Flows (WTF) Dataset (Feenstra et al., 2005) and was extended until 2008 using data from the UN COMTRADE website by Hausmann et al. (2011). This data contains the total export value for 1005 products using the SITC 4-digit (rev. 2) classification.

We exclude countries with less than 1.2 million citizens and with total trade below USD \$1 billion in 2008. Also excluded are countries with poor data on exports such as Iraq, Chad and Macau. This cut of the data accounts for 99% of World trade, 97% of World total GDP and 95% of World population (Hausmann et al., 2011). We use time varying national variables from the World Development Indicators (World Bank, 2010). In addition, we use data on conventionally measured factors of production (stock of physical capital, human capital and land) from UNCTAD (Shirotori et al., 2010). Bilateral data, such as distance between the most populated cities, common continent or region, territorial contiguity, common colonizer and colonizer–colony relationship, are from CEPII's GeoDist dataset (Mayer and Zignago, 2011).

In the static analysis, for which we use a cross-country data of the year 2000,² the base sample consists of 123 countries (7503 country pairs).³ For the dynamic analysis, the list of countries is reduced to 100, given the exclusion of countries with no geographic neighbors from the sample and those that belonged to the Former Soviet Union

¹ Alvarez et al. (2012) provides a useful framework to think about this. In their model, technology diffuses through the interaction of domestic and foreign business partners and competitors. Although they do not discuss the geographic implications of this assumption, one could expect this effect to be stronger at short distances as suggested by Keller and Yeaple (2013) in the context of multinational corporations and their foreign subsidiaries.

² We limit the analysis to one period (year 2000) in order to avoid artificially low standard errors given that most variables that will be used in the static analysis are fixed in time.

³ When we include data on factor endowments in our analysis, the dataset is limited to 105 countries.

(FSU). We exclude FSU countries from the dynamic analysis given that their data is non-existent prior to 1990 and sparse and scattered until 1995.

2.2. Exploring static similarity

As a descriptive exercise, we first study the correlation between geographic proximity and the similarity in exports of countries. To do so, we measure the intensity with which a country exports each product by computing its revealed comparative advantage (RCA) (Balassa, 1965). The RCA that a country has in a product is defined as the ratio between the share of total exports that the product represents in the country's export basket and the product's share of global trade. For example, in the year 2000, "aircrafts (between 2 and 15 tons)" represented 4.5% of Brazil's exports, but accounted only for 0.23% of total world trade. Hence, Brazil's RCA in aircrafts for that year was $RCA_{BRA,Aircrafts} = 4.5/0.23 = 19.56$, indicating that aircrafts are about 20 times more prevalent in Brazil's export basket than in that of the world. A product is over-represented in a country's export basket if its RCA is above 1. Formally, if $exp_{c,p}$ is equal to the dollar exports of country c in product p , then the RCA of country c in product p is defined as:

$$RCA_{c,p} \equiv \frac{exp_{c,p} / \sum_p exp_{c,p}}{\sum_c exp_{c,p} / \sum_c \sum_p exp_{c,p}} \tag{1}$$

To create a measure of similarity in the export structure of a pair of countries c and c' we define the *Export Similarity Index* ($S_{c,c'}$) as the Pearson correlation between the logarithm of the RCA vectors of the two countries. The *Export Similarity Index* is defined as:

$$S_{c,c'} \equiv \frac{\sum_p (r_{c,p} - \bar{r}_c) \sum_p (r_{c',p} - \bar{r}_{c'})}{\sqrt{\sum_p (r_{c,p} - \bar{r}_c)^2 \sum_p (r_{c',p} - \bar{r}_{c'})^2}} \tag{2}$$

where $r_{c,p} = \ln(RCA_{c,p} + \varepsilon)$ and \bar{r}_c is the average of $r_{c,p}$ over all products for country c . We chose a log form to prevent the correlation from being driven by the few products that countries export with very high RCA and we add ε , defined as 0.1, to assign a value to the zeroes, while also preventing the correlation being driven by similarities in the RCA of products that countries export very little of or not at all.⁴

$S_{c,c'}$ is larger than zero for pairs of countries that tend to export a similar set of goods with similar intensities, and negative for pairs of countries exporting different sets of goods. This feature of our index differs from the Finger & Kreinin (F&K) Export Similarity Index (Finger and Kreinin, 1979), which is calculated as the sum of the minimums of the export shares of each pair of countries. We prefer our measure as it distinguishes between products that are exported by one country and not the other from those that are exported by neither. Also, we use RCA, which gives equal weights to all products, while the F&K measure privileges products with large global markets. Nevertheless, our analysis is robust to using the F&K Similarity Index and other variations of our own similarity index (see Section B1.3 of the web appendix).

Table 1 presents summary statistics for our base sample which contains bilateral country-level data for the year 2000. Note that data on factor endowments is limited to fewer countries.

The left panel of Fig. 1 contains histograms for the Export Similarity ($S_{c,c'}$) in year 2000 for neighboring countries (unfilled) to all other country pairs (filled). The continuous lines are empirically fitted probability

Table 1
Summary statistics (year 2000).

Variable	N	Mean	sd
Similarity Index	7503	0.169	0.137
Similarity Index (NPRB)	7503	0.148	0.132
Simple Distance (km)	7503	7338.655	4389.738
Ln Simple Distance (km)	7503	8.649	0.817
Share A Border	7503	0.025	0.155
Same Language	7503	0.103	0.305
Have/Had Colonial Relationship	7503	0.015	0.123
Common Colonizer	7503	0.062	0.241
Log Total Bilateral Trade (Imp + Exp)	7503	8.854	8.580
Abs. Dif. Ln GDP per Capita (PPP)	7503	1.424	1.006
Abs. Dif. Ln Population	7503	1.572	1.211
Abs. Dif. Ln Physical Capital per Worker	5460	1.649	1.214
Abs. Dif. Ln Human Capital per Worker	5460	0.446	0.369
Abs. Dif. Ln Land per Worker	5460	0.609	0.728
	N	Mean	Mean within same region
Same Region	7503	0.1501	–
East Asia	7503	0.0160	0.1066
Eastern Europe	7503	0.0400	0.2664
Western and Central Europe	7503	0.0181	0.1208
Latin America and Caribbean	7503	0.0253	0.1687
Middle East and North Africa	7503	0.0160	0.1066
North America	7503	0.0001	0.0009
South Asia	7503	0.0008	0.0053
Sub-Saharan Africa	7503	0.0337	0.2247

distribution functions for the two samples based on the histograms. The figure shows that countries sharing a border have export baskets that are, on average, twice as similar as pairs of countries that do not share a border. The average $S_{c,c'}$ for border sharing geographic neighbors (i.e. share a border) is 0.40, compared to 0.16 for non-neighbors.⁵ In the right panel of the same figure, we show that export similarity decays exponentially with distance.

Export similarity, however, can be the consequence of shared geology or climate, which is more likely to be the case for geographic neighbors. To control for this fact, we exclude products from the sample that are confined by geography. We do this by using the technological classification suggested by Lall (2000) that divides products into the categories presented in Table 2.

Lall's classification is used to create two categories of products: primary and resource based (PRB) products and non-primary or non-resource based (NPRB) products. We consider PRB products as those that are classified as Gold, Primary Products and Resource Based Manufactures (categories 1 thru 4 in Table 2), whereas NPRB products are those contained in all other categories.

Fig. 2 reproduces Fig. 1 using NPRB products only. In this case, the mean Export Similarity Index of neighboring country-pairs is also significantly larger than in the non-neighbors sample of country-pairs,⁶ and the negative relationship between export similarity and geographical distance is equally strong, suggesting that the observed export similarity among neighbors is not driven solely by primary and resource based products. We include more controls in this analysis next.

2.3. The correlates of Export Similarity

The fact that, beyond geology and climate, export similarity decays with distance could be due to a number of different reasons. We study the correlates of the Export Similarity Index through an adapted "gravity model" (Zipf, 1946; Tinbergen, 1963). We do so in order to

⁴ We test that our results are not driven by the choice of ε . See Section 1 of the web appendix for robustness checks of this exercise using different values of ε for constructing $S_{c,c'}$. Also, Section 2 of the web appendix presents robust results with an alternative $S_{c,c'}$ that does not require a log-transformation.

⁵ This difference in means is statistically significant, with $t = -24.16$.

⁶ The difference in means between neighbors and non-neighbors is statistically different with $t = -26.38$.

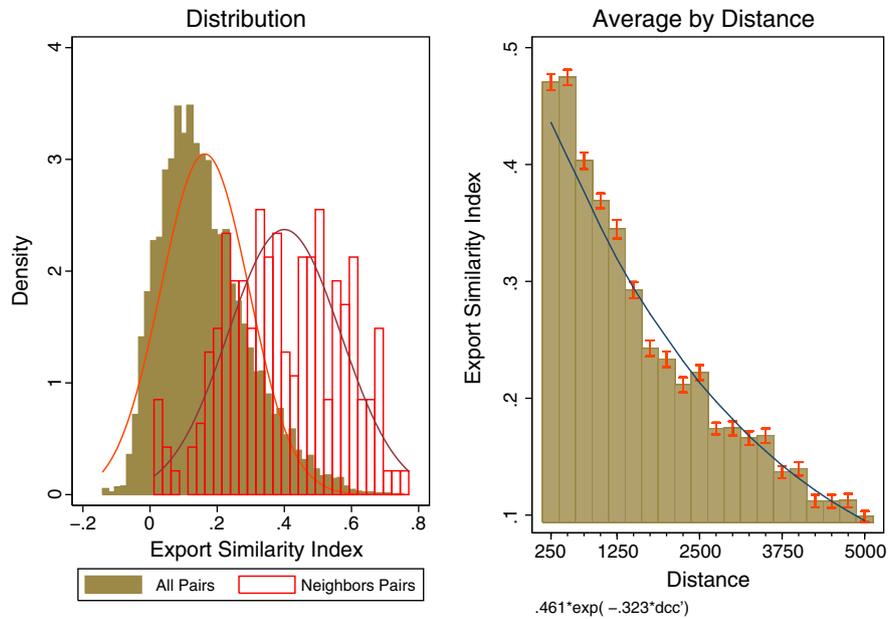


Fig. 1. Export Similarity Index (year 2000). The left panel of the figure shows the histogram, with a fitted pdf, of the Export Similarity Index in year 2000 for All (not neighbors) Country Pairs, and for Neighbor Pairs only. The right panel shows the average Export Similarity Index for country pairs in each bracket of distance between 250 km. to 5000 km.

understand whether the role of geographic proximity is actually driven by similarity in other dimensions such as income, size, factor endowments, institutions and culture, among others. Our adapted gravity model follows the functional form:

$$S_{c,c'} = \alpha + \beta \times d_{c,c'} + z_{c,c'}\gamma + l_{c,c'}\theta + b_{c,c'}\delta + \mu_c + \mu_{c'} + \varepsilon_{c,c'} \quad (3)$$

where $d_{c,c'}$ is the distance between countries c and c' (in logs), $z_{c,c'}$ is a set of two binary variables related to geographical closeness between c and c' : sharing a border and being in the same geographical region (i.e. continent). $l_{c,c'}$ is a set of binary variables representing cultural and institutional closeness between c and c' , which includes speaking a common official language, having had the same colonizer or having had a colony–colonizer relationship. $b_{c,c'}$ is a set of continuous regressors which measure differentials in quantifiable attributes between countries c and c' such as gaps in income per capita, population and factor endowments. $b_{c,c'}$ also includes total bilateral trade (imports plus exports) between each pair of countries. Finally, μ_c and $\mu_{c'}$ are country dummies capturing any individual country characteristic for countries c and c' respectively (analogous to the multilateral resistance dummies from Anderson and Van Wincoop (2001)). $\varepsilon_{c,c'}$ represents the error term. The results of this regression are presented in Table 3. For easier interpretative purposes, we use a normalized version of $S_{c,c'}$ as the dependent variable, with mean zero and unit standard deviation.

The first three columns of Table 3 correspond to the results with the (normalized) Export Similarity Index computed with all products, while the last three columns uses a version of the Export Similarity Index computed with NPRB products only. The base dataset contains 123 countries, which sum up to 7503 unique country pairs in year 2000.⁷ Columns 3 and 6 include factor endowment data, which reduces the sample to 5460 unique country pairs (105 countries).

Column 1 shows a negative correlation between similarity in exports and distance: the estimated coefficient implies that a pair of countries separated by twice the average distance is expected to have a similarity index that is 0.55 standard deviation below the mean. Column 4 repeats the same equation using only NPRB products and finds a slightly higher

coefficient, with similarity declining in 0.59 standard deviation from the mean. This result is always robust to the several tests we run in Section B1 of the web appendix. Columns 2 and 5 include two variables that represent alternative measures of geographic proximity and are highly correlated with distance: sharing a border and being in the same region. Sharing a border is associated with an Export Similarity Index that is, on average, 0.8 standard deviations above the mean for all products and 0.9 for NPRB goods. We can add to this another 0.4 or 0.35 standard deviations respectively if the two countries are in the same geographical region. This means that we could expect neighboring countries in the same region to have, on average, a similarity index roughly 1.2 standard deviations above the mean relative to non-neighbors from different regions for all goods and 1.25 for NPRB products—this does not take into account the fact that neighbors are a shorter distance apart than the average pair of countries. These variables are always strongly significant in all our robustness checks. This motivates our use of neighboring countries in our dynamic analysis in the next section.

In columns 3 and 6 we include a full set of other controls. These reduce the coefficient on the three distance variables by about a third, although they remain strongly significant in all robustness checks. Coefficients for the same language, and colonial relationship are not statistically significant when other controls are included. The table also shows that, as expected, differences in income levels, size and factor endowments are associated with lower levels of similarity in exports. Total bilateral trade is also negatively associated with similarity in exports: countries that trade more among themselves are less similar in their export baskets, as would be expected.

These figures suggest that geographic neighbors have similar export baskets, even when accounting for country fixed effects, common characteristics on culture and institutions (through the inclusion of data on colonial history and language), trade between them and differences in their income, populations and factor endowments. The measures of difference in factor endowments (physical capital, human capital and land) have the negative coefficient that would be expected from a Heckscher–Ohlin (HO) model, but they do not crowd out the economic or statistical significance of the geography regressors.

The similarity of the results between the three first columns and the last three columns of Table 3, which use as the dependent variable the

⁷ 123*122.

Table 2
Lall classification.

Lall classification	Number of Products
Gold	1
Primary Products	193
Resource Based Manufactures 1 (agro-based products)	130
Resource Based Manufactures 2 (others non-agro based products)	108
Low Technology Manufacture 1 (textiles, garments and footwear)	100
Low Technology Manufacture 2 (others)	97
Medium Technology Manufacture 1 (automotive)	15
Medium Technology Manufacture 2 (process)	109
Medium Technology Manufacture 3 (engineering)	135
High Technology Manufacture 1 (electronic and electrical)	49
High Technology Manufacture (others)	34
Special	12
Unclassified	22

(normalized) NPRB Export Similarity Index, suggests that climate and geology are not the central players in the impact of geographic proximity on export similarity.

However, the similarity in the composition of NPRB exports among neighbors might be driven by other factors, such as similarity in preferences. Following the Linder Trade Hypothesis (Linder, 1961), countries with similar preferences and hence demand structure, are likely to trade more, which in a Helpman–Krugman interpretation is due to the fact that they enjoy different varieties of similar products (Helpman and Krugman, 1985). Moreover, in a world with integrated supply chains, the similarity in exports could be a result of neighboring countries trading inputs that are classified in the data in the same category as the outputs themselves. Since neighbors trade more intensively, then similarity in bilateral trade may be driving our results. We check for this by comparing the similarity index $S_{c,c'}$ of the bilateral exports of neighbors with the similarity index of their exports to the rest of the world. To do this, we construct a similarity index for each pair of countries based on bilateral exports between each pair of countries, and a similarity index based on each pair's exports compared to the rest of the world (excluding the bilateral exports). Fig. 3 plots the two measures using data from 2000 and neighboring country pairs only. As Fig. 3 shows, neighbors are remarkably more similar in terms of what they export to the rest of the world than what they trade between themselves. This implies that export similarity is not driven by the composition of bilateral trade between neighbors.

An alternative exercise to explore this point consists of repeating the estimation of model (3), using the similarity index of their exports to the rest of the world as the dependent variable. The result is presented in Table 4. The relationship with distance—taken to mean sharing a border and being in the same region—holds when considering only exports to the rest of the world as the basis for similarity between all pairs of countries. Moreover, the last column on this table includes the bilateral similarity index as a regressor, although its inclusion does not qualitatively change the results. Section A.2 in the Appendix presents further analysis.

In sum, even when we look at NPRB products and we exclude bilateral trade, geographic proximity plays a role in explaining export similarity. This is a puzzle not easily explained by traditional frameworks that predict greater differentiation among countries that face lower transportation costs (i.e. shorter distances). In the next section, we turn our attention to the dynamics underlying this process.

3. Dynamics of Export Similarity

The previous section established that neighbors have more similar NPRB export baskets, even after controlling for similarities in size, income levels, cultural and institutional measures, factor endowments and taste. Is this a static bequest of history or the consequence of a dynamic process presently active?

To explore this issue, we use a dynamic analysis to study the role neighbors play in the ability of countries to add a particular good to their export basket or to expand their comparative advantage in a product. We start by discussing the extensive margin. More specifically, we study the probability that a country will add a product to its export basket in period T (i.e. “jump” to the product) if it has at least one neighbor that is already exporting that product in period t (with $T > t$). For this task, we use the dataset described in Section 2.1, with 100 countries.⁸ We divide our sample into four periods: 1970–1980, 1980–1990, 1990–2000 and 2001–2008.⁹ For each period, we eliminate all products that were not exported by any country and all countries that did not export any product. The total number of countries in the dataset is 100, and the total number of products is 777.

We define a “jump” as a tenfold or more increase in the RCA of country c in product p , from $RCA_{cp} \leq 0.1$ to $RCA_{cp} \geq 0.1$ within a ten year period.¹⁰ This setting allows us to explore the extensive margin of exports. We are interested in studying the probability of a product being exported in the next period, given that it was not being exported (or exported only in very small quantities) at the beginning of the current period. Furthermore, we are interested in products that achieve an RCA above 1, implying significant gains in comparative advantage and increases in its share of world trade.¹¹

To avoid noise, we restrict jumps to two conditions: first, a jump needs to keep RCA above 1 for four years after the end of the period, year T (the forward condition); second, we restrict jumps to products that had an RCA below 0.1 for two years before the beginning of the period (the backward condition).¹² These two conditions help rule out the possibility of “temporary jumps” in the data driven by noise, errors, shocks in commodity prices or other exogenous reasons.¹³

Table 5 presents the ten NPRB and PRB products with the largest frequency of “jumps” in our dataset. For instance, the NPRB product with the largest number of appearances in the data (i.e. with RCA going from less than 0.1 to above 1 in ten years) is SITC 8441 (men's undershirts), in the period 1980–1990. SITC 8441 had 6 occurrences (denoted by O) out of the 74 countries that had an RCA < 0.1 in 1980 (denoted by B). This means that 8% of the eligible countries acquired RCA > 1 for SITC 8441 over that period (denoted by P). Seven out of the top ten products for the NPRB categories are garments and textiles in the period 1980–1990.

Table 6 presents the ten countries with the largest number of product appearances in our dataset, classified by NPRB and PRB products. When looking at the ranking based on NPRB products, all countries in the list are developing countries, besides Germany, mostly located in Southeast Asia. China, at the top of the list, added 17 NPRB products to its export basket in the period 1980–1990, or 7% of the 234 products that, at the time, were being exported with an RCA below 0.1. The bottom list, based on PRB products, shows the repeated appearance of many countries from the top list, but also includes many African countries. The presence of Germany in the list is surprising, given that it is the only developed country that appears. However, its high ranking in the period

⁸ Since our main focus will be on geographic neighbors, we eliminate all islands. Also, given that this is a dynamic setting, we eliminate all Former Soviet Union countries, because their export data is non-existent prior to 1990 and sparse and scattered until 1995.

⁹ Since the original Feenstra data runs up to year 2000, and since 2001 and on was extended by the authors, we prefer to start the last period in 2001 to avoid discrepancies in the data.

¹⁰ With the exception of our last period which is seven years long (2001–2008).

¹¹ In Section B3 of the web appendix we replicate the results using different thresholds to test robustness. We present results by defining jumps as achieving an $RCA_{cp} \geq 2$ and an $RCA_{cp} \geq 5$.

¹² In Section B3 of the web appendix we present robustness checks that limit the sample to observations for which RCA is equal to zero at the beginning of the period.

¹³ For the last period (2001–2008) we eliminate the forward condition due to data limitations.

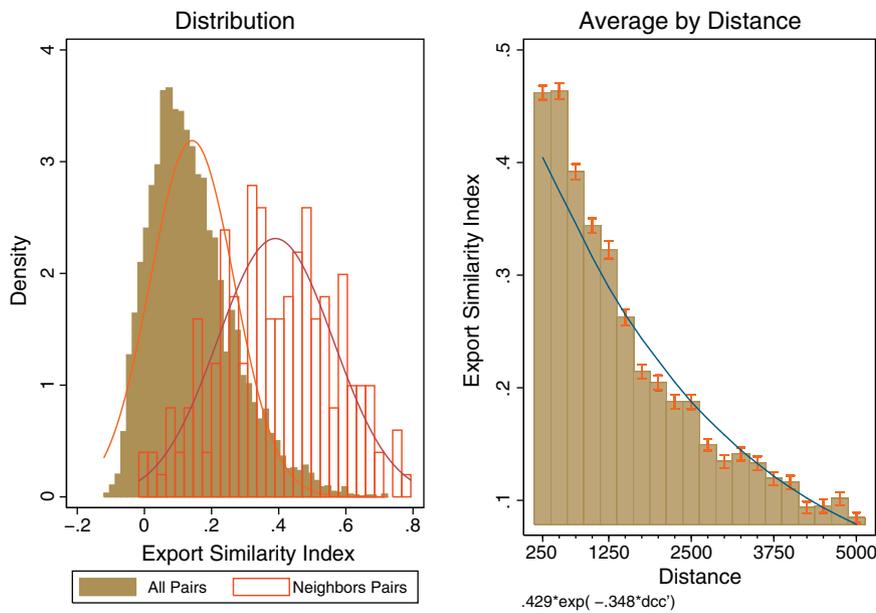


Fig. 2. Export Similarity Index NPRB products (year 2000). The left panel of the figure shows the distribution (in year 2000) of the Export Similarity Index for All (not neighbors) Country Pairs, and for Neighbor Pairs only. The right panel shows the average Export Similarity Index for country pairs in each bracket of distance between 250 km to 5000 km. This figure uses the Export Similarity Index for NPRB products only.

1990–2000 might well be due to classification errors associated with the reunification of the country.¹⁴

To test our hypothesis regarding the importance of the RCA of neighbors in the evolution of the extensive margin of exports, we estimate the following empirical specification:

$$J_{c,p,t \rightarrow T} = \alpha + \beta \ln(RCA_{c_N,p,t}) + \text{controls}_{c,p,t} + \varphi_{p,t} + \mu_{c,c_N,t} + \varepsilon_{c,p,t} \quad (4)$$

where $J_{c,p,t \rightarrow T}$ is a binary variable that takes the value of 1 when there was a “jump” between year t and T in product p and country c . The variable of interest, $\ln(RCA_{c_N,p,t})$, is the natural logarithm of the RCA of the neighbor with the largest RCA in product p for country c (we name this neighbor c_N). We also include a set of control variables at the country-product level. This includes the baseline RCA of country c in product p to account for differences in the probability of future exports for products that were larger at the beginning of the period. We also include the average annual growth rate of the RCA in the previous ten year period in order to control for parallel trends in comparative advantage for neighboring countries.¹⁵ To correct for undefined growth rates caused by zeros in the denominator, we compute the growth rate using $RCA + 0.1$ for all observations, thus pairing down the rate of growth for very low RCA products. To control for our own correction, we also add a dummy variable indicating whether the RCA was zero at the initial year of the computed growth rate used in the right hand side of the specifications, which are the observations more likely to be distorted. We also control for the “density” of the country in the product at the beginning of the period. The variable “density”, which distributes between 0 and 1, was developed by Hausmann and Klinger (2007) and used in Hidalgo et al. (2007). It measures the intensity with which a country exports products that are strongly co-exported by other countries who also export the product under consideration. In other words, the density of a product proxies for the existence of other exports that share similar technologies or inputs (as measured by their co-occurrence across countries). Density strongly affects the likelihood that a country adds the product to its export

¹⁴ To avoid this classification problem we have removed all former Soviet Union countries from the data.

¹⁵ For the first period 1970–1980 we used the previous eight year average annual growth rate (1962–1970) due to data limitations.

basket (Hausmann and Klinger, 2007; Hidalgo et al., 2007). We use density to control for the likelihood that a country would jump to a product given the initial composition of its export basket.¹⁶ $\varphi_{p,t}$ are product-year fixed effects which control for any time-varying product characteristic such as global demand, price or productivity shocks, particular to product p . $\mu_{c,c_N,t}$ are country-neighbor-year fixed effects, which use the neighbor with the largest RCA in that product. By adding $\mu_{c,c_N,t}$ we control for time-varying country-neighbor aggregate characteristics such as similarity in institutions, geography, climate, culture, history, productivity, economic development, population, initial factor endowments, inflation, bilateral exchange rates, etc.¹⁷

Following the seminal work of Jaffe et al. (1993), we created a control group for our sample in order to test the economic significance of our results. In the control dataset we replace a country’s real neighbors with an equal number of randomly chosen countries. For instance, if South Africa has four neighbors: Botswana, Mozambique, Namibia and Zimbabwe, in our randomization, South Africa will still have four neighbors, but these are chosen randomly. We iterate this randomization 500 times, and average the largest RCA in the neighborhood of each country for every product across all iterations. We compare the results of our dataset with those achieved using the control dataset. We expect that, if neighbors play a role in determining the ability of a country to become more productive in a good, the magnitude of β will be larger in the estimation using the real dataset than when using the control dataset. Our randomizations yield similar means for the RCA of neighbors in the overall sample.

Table 7 shows the summary statistics of the data used for this exercise, in which each observation is at the country-product-period level. Our sample includes only observations which are “eligible to jump”, that is, all observations in our dataset for which $RCA_{c,p,t} \leq 0.1$ at the

¹⁶ All results are robust to the exclusion of this variable. In fact, the inclusion of this variable reduces the size of our estimator of interest.

¹⁷ In robustness tests we added as a control the total bilateral imports of product p from country c' at time t , to study whether the likelihood of jumping is partly explained by importing that same good. The variable added very little to the specifications, and in most cases was not significant (though with a negative sign: the more you import from that good the less likely you are to export it). Given its poor performance, and the fact that determining the channels behind the results is out of the scope of this paper, we decided to exclude that variable from our controls.

Table 3
Correlates of the Export Similarity Index (year 2000).

	All	All	All	NPRB	NPRB	NPRB
Ln Simple Distance (km)	−0.5563 (0.017)***	−0.3233 (0.022)***	−0.3156 (0.023)***	−0.5901 (0.018)***	−0.3673 (0.022)***	−0.3477 (0.025)***
Share A Border		0.8023 (0.084)***	0.6500 (0.084)***		0.9037 (0.090)***	0.7740 (0.094)***
Same Region		0.4162 (0.038)***	0.1223 (0.043)***		0.3551 (0.041)***	0.0990 (0.048)**
Same Language			0.0825 (0.042)*			0.0696 (0.046)
Have/Had Colonial Relationship			0.0156 (0.084)			−0.0233 (0.081)
Common Colonizer			0.0334 (0.052)			0.0418 (0.058)
Abs. Dif. Ln GDP per Capita (PPP)			−0.2915 (0.027)***			−0.2442 (0.029)***
Abs. Dif. Ln Population			−0.0940 (0.011)***			−0.1121 (0.012)***
Log Total Bilateral Trade (Imp + Exp)			−0.0312 (0.002)***			−0.0250 (0.002)***
Abs. Dif. Ln Physical Capital per Worker			−0.0773 (0.024)***			−0.0773 (0.026)***
Abs. Dif. Ln Human Capital per Worker			−0.4105 (0.048)***			−0.4340 (0.050)***
Abs. Dif. Ln Land per Worker			−0.2206 (0.032)***			−0.2330 (0.033)***
N	7503	7503	5460	7503	7503	5460
r ²	0.37	0.39	0.57	0.33	0.35	0.52

The dependent variable in this table is the normalized Export Similarity Index, with mean zero and unit standard deviation. Columns 1–3 estimates model (3) with the Export Similarity Index computed using all products, while columns 4–6 uses the Export Similarity Index computed using NPRB products only. All regressions include country dummies. Standard errors are clustered at the country-pair level.

* $p < 0.10$. ** $p < 0.05$. *** $p < 0.01$.

beginning of the period. Our sample has almost 175,000 observations when using all products, and around 90,000 when restricting the sample to NPRB products only. The left-hand side variable in our specifications is “New Product (10 years)”, which has a mean value of 0.015 in the overall sample (or 0.016 in the sample restricted to NPRB products). That is, the unconditional probability of “jumping” is 1.5% (or 1.6% for NPRB products only). In the right-hand side, there are two variables of interest that we will use interchangeably. First, the continuous variable “Ln Maximum RCA [of] Neighbors”, which is the natural logarithm of $RCA_{cN,p,t}$ (being c_N the neighbor of c with the largest RCA for each product p in time t). Second, the binary variable “Neighbor Exports”, which is a dummy variable that takes the value of 1 if the country has a neighbor with $RCA_{cN,p,t} \geq 1$ in that product.

Our results are presented in Table 8. Panel A estimates the model (4) with the “Ln Maximum RCA [of] Neighbors” variable as the regressor of interest, while panel B estimates the same model with the “Neighbor Exports” binary variable. The first two columns in both panels present the results from our original sample. The last two columns in both panels use the control sample with randomly assigned neighbors—as previously explained. Our variables of interests, in both their continuous and binary form, are economically and statistically significant in columns 1 and 2 (which estimate the model with the real sample) and neither economically nor statistically significant in columns 3 and 4 (using the control sample). The economic significance of this result is the following: a doubling in the export intensity of a product by a geographic neighbor (i.e. RCA) at the beginning of the period is associated, on average, with a 0.4 percentage point increase in the likelihood of a country adding that product to its export basket. This is roughly a 25% increase (based on the unconditional probability of “jumping” of 1.5%). Panel B of Table 8 estimates that if a country has one neighbor who already exports product p with an RCA above 1 at the beginning of the period, then the chance that the country “jumps” to that product increases by 1 percentage point. This represents an increase of roughly

65% for the average product in the probability of “jumping” (from 1.5% to 2.5%).¹⁸

We look now at the intensive margin of trade, asking whether having neighbors with higher RCA in the initial year is associated with faster growth in RCA in the next period. Table 9 replaces the dependent variable of specification (4) with the compound average annual growth rate of RCA for the same time periods as before. In this exercise we use all the observations in the dataset without the low RCA restriction we used for the extensive margin. The intention is to estimate future growth in exports for a particular product in which a geographic neighbor has revealed comparative advantage, instead of focusing on new appearances. The main difference between the two approaches is that, by looking at the intensive margin, we include products that are already being exported by the country under consideration, and do not limit the sample to non-exported products.

The results in Table 9 are organized in the same way as those in Table 8. The upper panel estimates the model using the “Ln Maximum RCA [of] Neighbors” variable as the regressor of interest, while the lower panel estimates the same model with the “Neighbor Exports” binary variable. The results in panel A show a strong positive association between a country’s increase in future product RCA growth and the highest RCA of a neighboring country in that product at the beginning of the period. Panel B shows that having a neighbor with $RCA > 1$ is associated with a future annual growth of RCA of 1.5%–1.7% for that product in the next ten year period (or 16–18% cumulative). Table B7 in Section B3 of the web appendix shows that this result is robust to the use of export growth as the dependent variable, rather than RCA growth.

¹⁸ See Section B3 of the web appendix for a number of robustness checks of these results, varying the definition of the LHS variable, the method, the sample used and the dataset. All tests show full robustness with the results presented here.

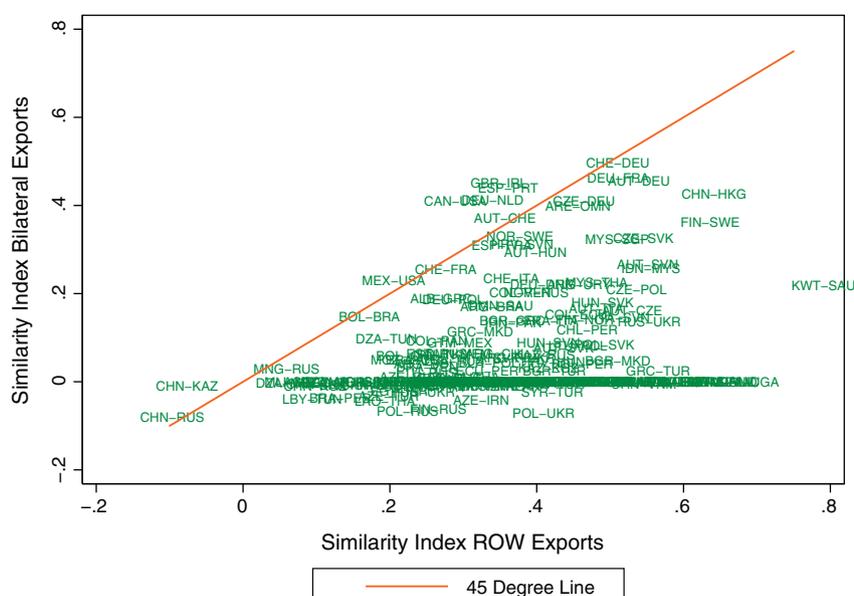


Fig. 3. Neighbor Similarity (on bilateral exports vis-à-vis ROW exports). This figure uses data from year 2000. It shows a scatter plot, where every observation is a country-pair. The horizontal axis measures the Similarity Index on Rest-of-the-World Exports (a measure of how similar a pair of countries is in terms of their exports to the rest of the world, excluding bilateral exports). The vertical axis measures the Similarity Index on Bilateral Exports (a measure of how similar a pair of countries is in terms of their bilateral exports to each other).

We repeat this analysis for different regions, periods and types of products, to understand whether there are differential effects across any of these dimensions. That is, we are interested in understanding which set of observations in the sample are driving the observed results. Table 10 summarizes this exercise by presenting results for a different cut of the data in each row.

The left panel of Table 10 uses the maximum log RCA of neighboring countries, and the right panel uses the dummy variable which takes the value 1 if a country has a neighbor with RCA > 1 in that product (at the beginning of the period). The left and right panels are analogous to the upper and lower panels of Tables 8–9 respectively. For instance, the first row considers all observations eligible to “jump” (i.e. with a baseline RCA below 0.1). Of these, 2.52% achieved an RCA above 1 in the following ten years if they had a neighbor with an RCA in that same product in the top 25% of the distribution. The same number drops to 1.2% if the best neighboring exporter had an RCA in the bottom 75% of the distribution. The ratio of these two numbers indicates that the first group was 2.1 times more likely to “jump”. The table also presents the 95% confidence interval for the estimate of the coefficient on the neighbor RCA variable in model (4). The first row is analogous to the results presented in Table 8, but every row recalculates the coefficient for each cut of the data.

From Table 10 we find that our results are in fact dominated mostly by developing countries, given that, both in the left and right panels, the estimator for β is statistically significant only for non-OECD countries. When we divide the world into regions we see the same pattern. The “neighbor effect” is statistically significant for East Asia & the Pacific, Latin America and the Caribbean and Sub-Saharan Africa, in both specifications.

When we look at different time periods, the confidence intervals of both panels show that the estimated coefficients are significant and stable across all periods.

Finally, we divide the sample into ten product groups based on the first digit SITC code. For all product categories the odds ratios are above 1.5 and often above 2, but the 95% confidence intervals for β are statistically significant in crude materials, food and live animals, mineral fuels and several manufacturing categories.

4. Interpretation of the results

As was argued in the Introduction section, literature on knowledge diffusion documents the rapid deterioration of knowledge with

distance. If this assertion is true, neighboring countries should share knowledge than more distant countries. If product-specific knowledge is a fundamental component of product-level productivity, then a Ricardian model of trade would predict that knowledge similarity between neighbors should correlate with similarity in the patterns of comparative advantage, and that this similarity should decay with distance. Our results are compatible with this logic. In fact, our results are what the literature on knowledge diffusion would predict regarding the geographic evolution of both the extensive and the intensive margins of trade. In order to become globally competitive in a new product, or to improve its productivity in an existing product, a country's firms would have to acquire the relevant knowledge. If there are significant obstacles to the geographic spread of that knowledge, products whose technology exists nearby will be favored.

Our static results show just this: neighboring countries have very similar export baskets, even when only looking only at goods not pinned down by geology or climate (NPRB) and after taking into account similarities in income, factor endowments, common language and history and a set of other controls. The estimated effects are large: considering only NPRB products, sharing a region and a border makes a pair of countries between 1.3 and 2 standard deviations more similar. These results are not driven by bilateral trade, limiting the explanatory power of interpretations based on similarity of demand.

By the same token, the diffusion of knowledge over time implies that knowledge acquisition would occur preferentially in countries with neighbors in possession of that knowledge. Our dynamic product-level results document that countries preferentially become good at the products that their neighbors are already good at, both in the extensive as well as in the intensive margin. This occurs even after controlling for product-year fixed effects, which capture any product specific global demand or supply shock, and after controlling for country–neighbor–year fixed effects, which control for any time-varying similarity in aggregate bilateral characteristics.

While our observations are what would be expected in a world where knowledge diffusion decays strongly with distance, our results could be driven by factors other than knowledge diffusion. The documented similarity of dynamics in the evolution of export baskets between neighbors could be influenced by a common third factor that expresses itself in the region, albeit not simultaneously—there could be both supply or demand stories. On the supply side, for instance,

Table 4
Correlates of the ROW Export Similarity Index (year 2000).

	ROW	ROW	ROW	ROW
Ln Simple Distance (km)	−0.5863 (0.018)***	−0.3664 (0.023)***	−0.3593 (0.025)***	−0.3325 (0.025)***
Share A Border		0.9325 (0.093)***	0.8089 (0.096)***	0.6868 (0.091)***
Same Region		0.3390 (0.041)***	0.0906 (0.048)*	0.0949 (0.046)**
Same Language			0.0755 (0.046)	0.0869 (0.045)*
Have/Had Colonial Relationship			0.1095 (0.097)	0.0929 (0.089)
Common Colonizer			−0.0422 (0.054)	−0.0251 (0.053)
Abs. Dif. Ln GDP per Capita (PPP)			−0.3214 (0.028)***	−0.2985 (0.027)***
Abs. Dif. Ln Population			−0.0655 (0.012)***	−0.0730 (0.012)***
Log Total Bilateral Trade (Imp + Exp)			−0.0215 (0.002)***	−0.0216 (0.002)***
Abs. Dif. Ln Physical Capital per Worker			−0.0253 (0.025)	−0.0284 (0.025)
Abs. Dif. Ln Human Capital per Worker			−0.3472 (0.050)***	−0.3400 (0.049)***
Abs. Dif. Ln Land per Worker			−0.2991 (0.036)***	−0.2831 (0.035)***
Bilateral Exp. Sim. Index (standardized)			0.1397 (0.016)***	
N	7260	7260	5356	5356
r2	0.33	0.35	0.47	0.49

The dependent variable in this table is the normalized Export Similarity Index, based on exports to the rest of the world, with mean zero and unit standard deviation. Columns 1–3 estimates model (3) with the Export Similarity Index computed using all products. * $p < 0.10$. ** $p < 0.05$. *** $p < 0.01$.

countries may be on a similar development trajectory, moving—from agriculture to light manufactures and into more complex products, although one country is ahead of the other. Consequently, neighboring countries become good at the same products, but with a time lag. We try to control for this with the highly significant density variable—which captures a country's own predisposition to move into that product—and by the lagged growth rate of the product's exports in the country.

On the demand side, countries could have similar preferences, but slightly different levels of income. As they both become richer, they

Table 5
Frequency of jumps by product.

SITC4	Product name	Period	O	B	P
<i>NPRB products</i>					
8441	Men's undershirt	1980–1990	6	74	0.08
6781	Iron pipes	2001–2008	5	60	0.08
8439	Other women outerwear	1980–1990	5	70	0.07
5721	Prepared explosives	2001–2008	4	43	0.09
8442	Men's underwear	1980–1990	4	83	0.05
6521	Unbleached cotton woven fabrics	1980–1990	4	40	0.10
8459	Other knitted outerwear	1980–1990	4	64	0.06
5913	Herbicides	1980–1990	4	83	0.05
8423	Men's trousers	1980–1990	4	65	0.06
8452	Knitted women's suits & dresses	1980–1990	4	72	0.06
<i>PRB products</i>					
812	Bran, sharps & other cereal residues	1990–2000	5	58	0.09
611	Raw sugar beet & cane	1980–1990	4	66	0.06
3344	Fuel oils	1980–1990	4	33	0.12
342	Frozen fish, excluding fillets	1980–1990	4	72	0.06
9710	Gold, non-monetary	2001–2008	4	40	0.10
3415	Coal & water gases	2001–2008	4	73	0.05
344	Frozen fish fillets	1990–2000	3	49	0.06
2927	Flora	1990–2000	3	51	0.06
723	Cocoa butter & paste	1970–1980	3	72	0.04
343	Fresh or chilled fish fillets	1990–2000	3	64	0.05

Table 6
Frequency of jumps by country.

ISO3	Period	O	B	P
<i>NPRB products</i>				
China	1980–1990	17	234	0.07
Germany	1990–2000	13	111	0.12
Syrian Arab Republic	2001–2008	11	315	0.03
Bangladesh	1980–1990	11	382	0.03
Vietnam	1990–2000	11	364	0.03
Malaysia	1980–1990	11	307	0.04
Cambodia	1990–2000	10	422	0.02
Tanzania	2001–2008	10	358	0.03
Austria	1980–1990	9	151	0.06
Lao PDR	1990–2000	9	412	0.02
<i>PRB products</i>				
Germany	1990–2000	25	236	0.11
Syrian Arab Republic	2001–2008	12	251	0.05
Tanzania	2001–2008	11	208	0.05
Mozambique	2001–2008	9	258	0.03
Namibia	2001–2008	8	211	0.04
Malawi	2001–2008	8	268	0.03
Lao PDR	2001–2008	8	287	0.03
Botswana	2001–2008	8	304	0.03
Zimbabwe	1980–1990	7	293	0.02
Turkey	1970–1980	7	199	0.04

would express those preferences in similar goods, but in a time-lagged fashion. This Linder-inspired hypothesis would be more plausible if bilateral trade was an important component of the similarity between countries. However, as we have shown, this is not the case: countries are much more similar in what they export to third countries than in what they trade between themselves, while neighboring countries that trade more intensely are less similar than those that do not.

In spite of our extensive list of controls—density, product–year and country–neighbor–year fixed effects, initial RCA and lagged growth in RCA—it is difficult to be certain that the correlations we document are not caused by some other common third factor that would explain the time-lagged appearance of products in neighboring countries and the dynamic geographic patterns of comparative advantage. Any attempt at control is never perfect. But the results we obtain are what we expect from the hypothesis—amply documented in the literature—that knowledge diffusion decays very rapidly with distance.

5. Concluding remarks

This paper has established that neighboring countries are very similar in their patterns of comparative advantage, a similarity that decays with distance. In a classical Heckscher–Ohlin model, this would reflect the similarity in factor endowments. But, after taking into account a large set of controls, including similarity in incomes, sizes, conventional factor endowments, culture and institutions, among others, and after excluding goods not pinned down by geology or climate, the resemblance in the composition of the export baskets of neighboring countries remains very strong. The factors causing the similarity we

Table 7
Summary statistics dynamics of Export Similarity (1970–2008).

Variable	All			NPRB		
	N	Mean	sd	N	Mean	sd
New Product (10 years)	173,433	0.015	0.123	90,811	0.016	0.125
Baseline Ln RCA	173,433	−2.227	0.157	90,811	−2.224	0.159
Baseline Density	173,433	0.087	0.087	90,811	0.073	0.077
Growth Rate RCA	173,433	3.105	9.873	90,811	3.171	9.408
Zero RCA	173,433	0.650	0.477	90,811	0.631	0.483
Baseline Bilateral Imports (p)	173,433	0.827	2.215	90,811	0.829	2.190
Max RCA Neighbors	173,433	2.290	31.057	90,811	0.814	4.643
Ln Max RCA Neighbors	173,433	−1.256	1.436	90,811	−1.397	1.228
Neighbor Exports	173,433	0.172	0.377	90,811	0.137	0.344

Table 8
Dynamics of Exports Similarity.

Panel A: continuous independent variable				
	Real		Control	
	All	NPRB	All	NPRB
Ln Max RCA Neighbors	0.0037 (0.000)***	0.0040 (0.001)***	−0.0019 (0.003)	−0.0060 (0.007)
Baseline Ln RCA	0.0073 (0.004)**	−0.0035 (0.005)	0.0091 (0.004)**	−0.0042 (0.006)
Baseline Density	0.1302 (0.034)***	0.2266 (0.076)***	0.1557 (0.033)***	0.2536 (0.075)***
Growth Rate RCA (t − 1)	−0.0006 (0.000)***	−0.0005 (0.000)***	−0.0006 (0.000)***	−0.0005 (0.000)***
Zero RCA (t − 1)	0.0062 (0.001)***	0.0078 (0.002)***	0.0056 (0.001)***	0.0073 (0.002)***
N	173,433	90,811	173,433	90,811
r2	0.08	0.11	0.08	0.11
Panel B: binary independent variable				
	Real		Control	
	All	NPRB	All	NPRB
Neighbor Exports	0.0106 (0.001)***	0.0103 (0.002)***	0.0008 (0.001)	−0.0002 (0.002)
Baseline Ln RCA	0.0082 (0.004)**	−0.0029 (0.005)	0.0091 (0.004)**	−0.0042 (0.006)
Baseline Density	0.1389 (0.034)***	0.2343 (0.076)***	0.1555 (0.033)***	0.2523 (0.075)***
Growth Rate RCA (t − 1)	−0.0006 (0.000)***	−0.0005 (0.000)***	−0.0006 (0.000)***	−0.0005 (0.000)***
Zero RCA (t − 1)	0.0060 (0.001)***	0.0077 (0.002)***	0.0056 (0.001)***	0.0072 (0.002)***
N	173,433	90,811	173,433	90,811
r2	0.07	0.11	0.08	0.11

Panel A uses the maximum RCA among all geographic neighbors of a country for a particular product, in natural logarithm, as the independent variable. Panel B uses a dummy variable which takes the value 1 if at least one of the neighbors of a country has an RCA above 1 in the product under consideration. The control group uses a generated dataset in which neighbors are randomly assigned to countries, keeping constant the amount of neighbors per country. All regressions include country–neighbor–by–year and product–by–year fixed effects. Standard errors are clustered at the country–neighbor level.

* $p < 0.10$. ** $p < 0.05$. *** $p < 0.01$.

document go beyond the classical ones: physical capital, human capital, labor and land, including geology and climate.

Moreover, the similarity we document is not obvious—higher intensity trade at short distances should incentivize neighboring countries to specialize in different rather than in similar goods. In fact, our static results show that there is a negative correlation between bilateral trade intensity and export similarity.

To make these observations compatible with a Ricardian model of trade, something must cause a spatial correlation in the patterns of product-level productivity. Knowledge diffusion is a potential candidate, given that previous research has documented its very localized character.

This paper leaves open the question of what are the mechanisms behind the dynamic similarity we document. Future research should be able to elucidate this. Clearly, trade, foreign direct investment and migration are three prime suspects. On the trade front, Coe and Helpman (1995) and Coe et al. (2009) document that imports-weighted foreign R&D investment at the aggregate level are correlated with total factor productivity growth. But these results are not enough per se to account for our observations: we require product-level similarity in productivity, not an aggregate one, and we require an interaction that decays more rapidly with distance than imports, since imports tend to have a much longer diffusion range than knowledge. Alvarez et al. (2012) posit that the human interaction that occurs through trade causes knowledge spillovers. If this is so, knowledge would be translated in a coevolution of comparative advantage trends fueled by the transferability of knowledge from one to the another. Whether this occurs at the product level and what its geographic range is remains to be studied.

Foreign direct investment is also a potential channel. Borenstein et al. (1998) document aggregate effects of FDI on growth. Aitken and Harrison (1999), using plant level data, find limited spillovers from foreign to domestic firms in the same industry using Venezuelan data. Haskel et al. (2007) find more significant spillovers using data on UK manufacturing plants. Branstetter (2006) finds evidence of spillovers between the foreign direct investment of Japanese firms and US firms. Javorcik (2004) finds evidence of an impact of FDI on the productivity of local upstream suppliers, using Lithuanian data. Keller and Yeaple (2009) find strong evidence of inward FDI on the productivity of US firms, especially in high-tech industries. Moreover, the literature on FDI using gravity equations (Loungani et al., 2002; Portes and Rey, 2005; Stein and Daude, 2007) consistently shows a high elasticity of FDI with respect to distance and a strong additional border effect. However, it remains to be seen what FDI contributes to the evidence on export similarity we document in this paper.

Labor flows or migration could also be a channel for knowledge spillovers. If knowledge resides in brains, it should move with them. If direct human interaction is key to knowledge spillovers, as suggested by much of the literature quoted above, then people could be an important source of knowledge transmission. For instance, Andersen and Dalgaard (2011) show how the ease of travel can explain shifts in aggregate productivity. Other forms of human interaction may also be involved, including the ease of physical and electronic communication, as well as international ethnic/cultural links (e.g. Stein and Daude,

Table 9
Dynamics of Exports Similarity (RCA growth).

Panel A: continuous independent variable				
	Real		Control	
	All	NPRB	All	NPRB
Ln Max RCA Neighbors	0.7334 (0.045)***	0.7113 (0.069)***	−0.7788 (0.383)**	−0.4180 (0.586)
Baseline Ln RCA	−3.9685 (0.113)***	−5.0351 (0.186)***	−3.8515 (0.119)***	−4.9510 (0.191)***
Baseline Density	23.0227 (2.444)***	29.3487 (3.740)***	28.1252 (2.731)***	33.7179 (3.817)***
Growth Rate RCA (t − 1)	−0.0381 (0.007)***	−0.0002 (0.012)	−0.0397 (0.008)***	−0.0010 (0.013)
Zero RCA (t − 1)	−1.0230 (0.133)***	−0.7455 (0.183)***	−1.1853 (0.142)***	−0.8229 (0.197)***
N	262,017	136,929	262,017	136,929
r2	0.20	0.26	0.20	0.25
Panel B: binary independent variable				
	Real		Control	
	All	NPRB	All	NPRB
Neighbor Exports	1.7851 (0.119)***	1.5242 (0.160)***	0.0092 (0.112)	0.1153 (0.139)
Baseline Ln RCA	−3.8980 (0.112)***	−4.9979 (0.186)***	−3.8311 (0.116)***	−4.9397 (0.189)***
Baseline Density	24.2354 (2.451)***	30.3715 (3.715)***	28.0300 (2.739)***	33.6793 (3.819)***
Growth Rate RCA (t − 1)	−0.0387 (0.007)***	0.0001 (0.012)	−0.0395 (0.008)***	−0.0009 (0.013)**
Zero RCA (t − 1)	−1.1038 (0.132)***	−0.7782 (0.183)***	−1.1845 (0.142)***	−0.8211 (0.197)***
N	262,017	136,929	262,017	136,929
r2	0.20	0.25	0.20	0.25

This table presents results using the Compound Average Annual Growth for RCA in the next period as the dependent variable. Panel A uses the maximum RCA among all geographic neighbors of a country for a particular product, in natural logarithm, as the independent variable. Panel B uses a dummy variable which takes the value 1 if at least one of the neighbors of a country has an RCA above 1 in the product under consideration. The control group uses a generated dataset in which neighbors are randomly assigned to countries, keeping constant the amount of neighbors per country. All regressions include country–neighbor–by–year and product–by–year fixed effects. Standard errors are clustered at the country–neighbor level.

* $p < 0.10$. ** $p < 0.05$. *** $p < 0.01$.

Table 10
Intensive and extensive margin comparative statics.

	N	RCA neighbor				Neighbor exporter			
		β 95% C.I.	Top 25%	Bottom 75%	Ratio	β 95% C.I.	1	0	Ratio
All observations	173,433	(0.003, 0.005)	2.52%	1.20%	2.10	(0.008, 0.014)	2.88%	1.25%	2.31
Non OECD	147,081	(0.003, 0.004)	2.40%	0.93%	2.58	(0.007, 0.014)	2.85%	1.00%	2.84
OECD	26,352	(-0.001, 0.004)	2.96%	2.77%	1.07	(-0.003, 0.008)	3.01%	2.76%	1.09
East Asia & Pacific	20,636	(0.002, 0.009)	3.59%	1.43%	2.50	(0.004, 0.028)	3.87%	1.45%	2.67
Eastern Europe	7701	(-0.006, 0.002)	4.15%	3.59%	1.16	(-0.013, 0.011)	4.18%	3.60%	1.16
Latin America & Caribbean	33,918	(0.001, 0.005)	2.35%	1.01%	2.31	(0.003, 0.013)	2.45%	1.01%	2.42
Middle East & N. Africa	34,394	(-0.001, 0.001)	1.37%	0.97%	1.42	(-0.006, 0.004)	1.54%	0.99%	1.55
North America	1146	(-0.003, 0.023)	3.50%	0.81%	4.30	(-0.008, 0.054)	3.18%	0.84%	3.79
South Asia	5252	(-0.016, 0.011)	2.89%	1.65%	1.75	(-0.042, 0.017)	2.79%	1.67%	1.67
Sub-Saharan Africa	54,751	(0.000, 0.003)	1.89%	0.59%	3.21	(0.002, 0.013)	2.98%	0.67%	4.43
Western Europe	15,635	(0.000, 0.005)	3.50%	3.22%	1.09	(-0.005, 0.011)	3.52%	3.21%	1.10
Period 1970–1980	31,462	(0.000, 0.003)	1.35%	0.54%	2.50	(0.001, 0.011)	1.63%	0.59%	2.77
Period 1980–1990	50,133	(0.004, 0.009)	3.12%	2.09%	1.49	(0.010, 0.025)	3.70%	2.14%	1.73
Period 1990–2000	48,012	(0.002, 0.004)	2.41%	0.55%	4.36	(0.007, 0.016)	2.96%	0.62%	4.78
Period 2001–2008	43,826	(0.001, 0.004)	2.76%	1.37%	2.01	(0.001, 0.011)	2.85%	1.37%	2.08
Animal and vegetable oils, fats & waxes	4883	(-0.001, 0.007)	2.54%	0.96%	2.66	(-0.006, 0.027)	2.72%	1.00%	2.72
Beverages & tobacco	2464	(-0.010, 0.005)	2.60%	1.68%	1.55	(-0.019, 0.012)	2.89%	1.62%	1.78
Chemical and related products, n.e.s.	20,291	(-0.000, 0.004)	2.19%	1.40%	1.56	(0.000, 0.012)	2.60%	1.40%	1.86
Commodities & transactions not classified	848	(-0.015, 0.048)	5.19%	1.26%	4.12	(-0.052, 0.123)	5.79%	1.22%	4.76
Crude materials, inedible, except fuels	25,478	(0.001, 0.003)	2.42%	1.15%	2.10	(0.003, 0.013)	2.53%	1.18%	2.14
Food & live animals	20,740	(0.002, 0.006)	2.62%	1.34%	1.96	(0.002, 0.014)	2.68%	1.34%	2.00
Machinery & transport equipment	34,628	(-0.000, 0.005)	1.61%	0.89%	1.81	(0.000, 0.016)	2.41%	0.92%	2.62
Manufactured goods classified by material	41,490	(0.001, 0.004)	2.26%	1.09%	2.07	(0.003, 0.012)	2.57%	1.13%	2.27
Mineral fuels, lubricants & related materials	4572	(0.003, 0.011)	3.41%	1.63%	2.09	(0.004, 0.037)	3.46%	1.71%	2.02
Miscellaneous manufactured articles	18,039	(0.002, 0.011)	4.34%	1.80%	2.42	(0.001, 0.023)	5.59%	1.91%	2.92

2007; Giroud, 2012; Kerr, 2008). Whether these effects significantly contribute to the observed product-level geographic correlations remains to be shown.

In this context, one contribution of this paper is that it proposes a new observable with which to track knowledge diffusion: the export basket of countries. The comparative advantage of countries evolves as they absorb new technologies. Absorption of product-specific knowledge increases the productivity with which a product can be made, inducing more exports. In this paper we use this logic to provide additional evidence of the short range of knowledge diffusion that has been reported using other observables, such as total factor productivity, patent citations or patent productivity. But the use of this observable opens up new areas of research in a field that has been hampered by effective measures. Using export similarity it should be possible to study the impact of trade, FDI, migration, ease of travel and other forms of human interaction on international knowledge diffusion.

However, limited geographic knowledge diffusion is an important observation in its own right. This observation may account for the lack of income convergence at the global level and the fact that rich and poor countries tend to be geographically segregated. It implies that countries are affected by the knowledge that exists in their neighborhood. Knowledge diffusion is unquestionably not an economically insignificant phenomenon. It is more than a side effect. It can shape the evolution of the comparative advantage of nations.

Acknowledgements

We are grateful to the editor Ariel Burstein and two anonymous referees for their recommendations. We also thank Sebastian Bustos, Elhanan Helpman, Sarah Hopkinson, Juan Ariel Jimenez, Robert Lawrence, Marc Melitz, Dani Rodrik, Rodrigo Wagner, Muhammed Yildirim, Andrés Zahler and Richard Zeckhauser for their thoughtful comments, as well as to the participants of the CID Growth Lab Seminar, Harvard International Economics Graduate Students Seminar and Harvard CID Faculty Seminar. All errors are our own.

Appendix A

A.1. The network of Export Similarity

The similarity in the export baskets of countries is strongly affected by variables that proxy for distance and is robust to the inclusion of institutional, income and factor endowment variables. In fact, one way to illustrate the strength of the similarity between neighboring countries is to represent the matrix of export similarity as a network, where each country is connected to the two other countries most similar to it. Fig. A1 presents export similarity for year 2008 as a graphical network where each node represents a country, and each country is connected to the two other countries with the most similar export baskets, as measured by the Export Similarity Index $S_{c,c'}$. countries are colored according to geographic regions, showing that the clusters defined by export similarity correlate strongly with physical distance. The width of links is proportional to the similarity index and the color of the link indicates whether the similarity is driven by primary and resource based (PRB) products (blue) or by NPRB products (red) (see Section B2 of the web appendix for more details). We note that, in a large number of cases, the country with the most similar export structure is an immediate neighbor, such as in the case of France, Germany, Austria, the Czech Republic, Hungary and Slovakia or in the case of India, Pakistan, and Bangladesh. This visualization illustrates the strong association between proximity and export structure that characterizes the world economy.

A.2. Bilateral trade and similarity in exports

We pursue a more analytical approach to show that the similarity Index among neighboring countries is mostly driven by their exports to the rest of the world, and not by the bilateral exports between themselves (see Fig. 3). In order to do so, we decomposed the Export Similarity Index in two measures: (1) a Similarity Index in Bilateral Exports, which uses data on the bilateral exports among each pair of countries,

- Anderson, J.E., 1979. A theoretical foundation for the gravity equation. *Am. Econ. Rev.* 69 (1), 106–116.
- Anderson, J.E., Van Wincoop, Eric, 2001. Gravity with gravitas: a solution to the border puzzle. *Am. Econ. Rev.* 93 (1), 170–191.
- Arrow, Kenneth J., 1969. Classificatory notes on the production and transmission of technological knowledge. *Am. Econ. Rev.* 59 (2), 29–35.
- Balassa, B., 1965. Trade liberalisation and revealed comparative advantage. *Manch. Sch.* 33 (2), 99–123.
- Bergstrand, J.H., 1985. The gravity equation in international trade: some microeconomic foundation and empirical evidence. *Rev. Econ. Stat.* 67 (3), 474–481.
- Bergstrand, J.H., 1989. The generalized gravity equation, monopolistic competition, and the factor-proportions theory in international trade. *Rev. Econ. Stat.* 71 (1), 143–153.
- Borensztein, E., De Gregorio, J., Lee, J.-W., 1998. How does foreign direct investment affect economic growth? *J. Int. Econ.* 45 (1), 115–135.
- Bottazzi, Laura, Peri, Giovanni, 2003. Innovation and spillovers in regions: evidence from European patent data. *Eur. Econ. Rev.* 47, 687–710.
- Brander, J.A., 1981. Intra-Industry trade in identical commodities. *J. Int. Econ.* 11 (1981), 1–14.
- Brander, J., Krugman, P., 1983. A reciprocal dumping model of international trade. *J. Int. Econ.* 15 (1983), 313–321.
- Branstetter, Lee G., 2001. Are knowledge spillovers international or intranational in scope? Microeconomic evidence from the U.S. and Japan. *J. Int. Econ.* 53, 53–79.
- Branstetter, Lee G., 2006. Is foreign direct investment a channel of knowledge spillovers? Evidence from Japan's FDI in the United States. *J. Int. Econ.* 68 (2), 325–344.
- Coe, D.T., Helpman, E., 1995. International R&D spillovers. *Eur. Econ. Rev.* 2921 (94), 859–887.
- Coe, David T., Helpman, Elhanan, Hoffmaister, Alexander W., 2009. International R&D spillovers and institutions. *Eur. Econ. Rev.* 53 (7), 723–741.
- Davis, Donald R., 1995. Intra-industry trade: a Heckscher–Ohlin–Ricardo approach. *J. Int. Econ.* 39 (3–4), 201–226.
- Deardorff, Alan, 1998. Determinants of bilateral trade: does gravity work in a neoclassical world? In: Frankel, Jeffrey (Ed.), *The Regionalization of the World Economy*. University of Chicago Press, Chicago.
- Eaton, Jonathan, Kortum, Samuel, 1997. Technology and bilateral trade. NBER Working Paper Series, 6253.
- Feenstra, Robert C., Markusen, James R., Rose, Andrew K., 2001. Using the gravity equation to differentiate among alternative theories of trade. *Can. J. Econ.* 34 (2), 430–447.
- Feenstra, R.C., Lipsey, R.E., Deng, H., Ma, A.C., Mo, H., 2005. World trade flows: 1962–2000. NBER Working Paper Series.
- Finger, J.M., Kreinin, M.E., 1979. A measure of 'export similarity' and its possible uses. *Econ. J.* 89 (356), 905–912.
- Frankel, Jeffrey Alexander, 1997. *Regional Trading Blocs in the World Economic System*. Peterson Institute for International Economics.
- Giroud, Xavier, 2012. Proximity and investment: evidence from plant-level data. *Q. J. Econ.* 128 (2), 861–915.
- Haskel, J.E., Pereira, S.C., Slaughter, M.J., 2007. Does inward foreign direct investment boost the productivity of domestic firms? *Rev. Econ. Stat.* 89 (3), 482–496.
- Hausmann, Ricardo, Klinger, Bailey, 2007. The structure of the product space and the evolution of comparative advantage. CID Working Paper Series, 146.
- Hausmann, Ricardo, Hidalgo, César A., Bustos, Sebastián, Coscia, Michele, Chung, Sarah, Jiménez, Juan, Simoes, Alexander, Yildirim, Muhammed A., 2011. *The Atlas of Economic Complexity: Mapping Paths to Prosperity* (Cambridge, MA).
- Helpman, Elhanan, Krugman, Paul, 1985. *Market Structure and International Trade*. MIT Press, Cambridge.
- Hidalgo, César A., Klinger, Bailey, Barabási, A.L., Hausmann, Ricardo, 2007. The product space conditions the development of nations. *Science* 317 (5837), 482–487 (New York, N.Y.).
- Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *Q. J. Econ.* 108 (3), 577.
- Javorcik, Beata S., 2004. Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages. *Am. Econ. Rev.* 94 (3), 605–627.
- Kehoe, Timothy J., Ruhl, Kim J., 2013. How important is the new goods margin in international trade? *J. Polit. Econ.* 121 (2), 358–392.
- Keller, Wolfgang, 2002. Geographic localization of international technology diffusion. *Am. Econ. Rev.* 92 (1), 120–142.
- Keller, Wolfgang, 2004. International technology diffusion. *J. Econ. Lit.* XLII, 752–782 (September).
- Keller, Wolfgang, Yeaple, Stephen R., 2009. Multinational enterprises, international trade, and productivity growth: firm-level evidence from the United States. *Rev. Econ. Stat.* 91, 821–831 (November).
- Keller, Wolfgang, Yeaple, Stephen Ross, 2013. The gravity of knowledge. *Am. Econ. Rev.* 103 (4), 1414–1444.
- Kerr, W.R., 2008. Ethnic scientific communities and international technology diffusion. *Rev. Econ. Stat.* 90 (3), 518–537.
- Lall, Sanjaya, 2000. The technological structure and performance of developing country manufactured exports, 1985–98. *Oxf. Dev. Stud.* 28 (3), 337–369.
- Leamer, E.E., Levinsohn, James, 1995. International trade theory: the evidence. In: Grossman, Gene M., Rogoff, K. (Eds.), *Handbook of International Economics*, vol. III. Elsevier B.V., pp. 1341–1375 (chapter 26).
- Linder, Staffan Burenstam, 1961. *An Essay on Trade and Transformation*. Almqvist & Wiksell, Stockholm.
- Loungani, Prakash, Mody, Ashoka, Razin, Assaf, 2002. The global disconnect: the role of transactional distance and scale economies in gravity equations. *Scot. J. Polit. Econ.* 49 (5), 526–543.
- Mayer, Thierry, Zignago, Soledad, 2011. Notes on CEPII distances measures: the GeoDist database. CEPII Working Paper, 25.
- Polanyi, M., 1962. *Personal Knowledge: Towards a Post-critical Philosophy*. Routledge, London, UK.
- Portes, Richard, Rey, Hélène, 2005. The determinants of cross-border equity flows. *J. Int. Econ.* 65 (2), 269–296.
- Rivera-Batiz, L.A., Romer, P.M., 1990. Economic integration and endogenous growth. *Q. J. Econ.* 106 (2), 531–555.
- Shirotori, M., Tumurchudur, B., Cadot, O., 2010. Revealed factor intensity indices at the product level. *Policy Issues in International Trade and Commodities*, 44.
- Stein, Ernesto, Daude, Christian, 2007. Longitude matters: time zones and the location of foreign direct investment. *J. Int. Econ.* 71 (1), 96–112.
- Tinbergen, J., 1963. Shaping the world economy. *Int. Exec.* 5 (1), 27–30.
- United Nations, 2010. COMTRADE Database. <http://comtrade.un.org/>
- Venables, Anthony J., 1985. Trade and trade policy with imperfect competition: the case of identical products and free entry. *J. Int. Econ.* 19 (1–2), 1–19.
- World Bank, 2010. World Development Indicators Online. <http://data.worldbank.org/>
- Zahler, Andres, 2007. Decomposing world export growth and the relevance of new destinations. CID Working Paper Series, 20.
- Zipf, George Kingsley, 1946. The P1 P2/D hypothesis: on the intercity movement of persons. *Am. Sociol. Rev.* 11 (6), 677–686.