Neighbors and the Evolution of the Comparative Advantage of Nations: Evidence of International Knowledge Diffusion?

Dany Bahar, Ricardo Hausmann and Cesar A. Hidalgo

CID Working Paper No. 235
April 2012
Updated July 2013

Working Papers
Center for International Development
at Harvard University

Neighbors and the Evolution of the Comparative Advantage of Nations: Evidence of International Knowledge Diffusion?*

Dany Bahar, Ricardo Hausmann and Cesar A. Hidalgo†

July 26, 2013

Keywords: export similarity, innovation, productivity, knowledge, technology, diffusion, spillovers.

JEL Classification Numbers: O31, O33, F10, F62, F63

*We are grateful to the editor Ariel Burstein and two anonymous referees for their recommendations. We also thank Sebastian Bustos, Elhanan Helpman, Juan Ariel Jimenez, Robert Lawrence, Marc Melitz, Dani Rodrik, Rodrigo Wagner, Muhammed Yildirim, Andrés Zahler and Richard Zeckhauser for their thoughtful comments, a 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.

†Bahar and Hausmann: Harvard Kennedy School and Center for International Development at Harvard University; Hidalgo: Macro Connections Group at MIT Media Lab.

Abstract

The literature on knowledge diffusion shows that it 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, having a neighbor with comparative advantage in them is associated with a growth of exports that is higher by 1.5 percent per annum. While these results could be driven by a common third factor that escapes our controls, they are what would be expected from the localized character of knowledge diffusion.

1 Introduction

Knowledge has become central to modern theories of growth. Knowledge can be embodied in goods that can be shipped around at a cost. When these goods are imported, they accelerates productivity growth in the recipient country (e.g. Rivera-Batiz and Romer, 1990; Coe and Helpman, 1993; Coe et. al., 2009). But significant parts of knowledge are disembodied or tacit (Polanyi, 1962) and its diffusion requires more direct forms of human interaction, which 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 has 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 a 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 it diffuses at short distances its 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, while they would 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, after 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 compete in global markets. In addition, in order to increase their market share, they 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, which we explore both in a static and a dynamic setting. We study both the intensive and the extensive margin 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 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 & 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 from trade. The greater intensity of trade at short distances would force specialization and differentiation, whether -as pointed out by Feenstra, Markusen and Rose (2001)- the differences that cause the 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, cell phones 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 of 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). Also, the emerging consensus in the literature of knowledge diffusion is that it occurs predominantly at a fairly short range (e.g. Jaffe et al. 1993; Branstetter, 2001; Keller, 2002; Bottazzi & Peri, 2003), an observation that is attributed to 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

[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.

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.

Up to 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 & Peri, 2003; Branstetter, 2006) and changes in total factor productivity (e.g. Coe & Helpman, 1995; Keller, 2002; Keller & 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 extended until 2008 using data from the UN COMTRADE website by Hausmann et. al. (2011). It 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 & 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

[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.

the dynamic analysis, the list of countries is reduced to 100, given the exclusion from the sample of countries with no geographic neighbors and those that belonged to the Former Soviet Union (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 share of global trade in that product. 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:

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

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 choose a log form to prevent the correlation to be 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 at the same time preventing the correlation to be 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 & 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

[4]We test that our results are not driven by the choice of ε. See section A.1.1 for robustness checks of this exercise using different values of ε for constructing $S_{c,c'}$. Also, section A.1.2 present robust results with an alternative $S_{c,c'}$ that does not require a log-transformation.

Table 1: Summary Statistics (Year 2000)

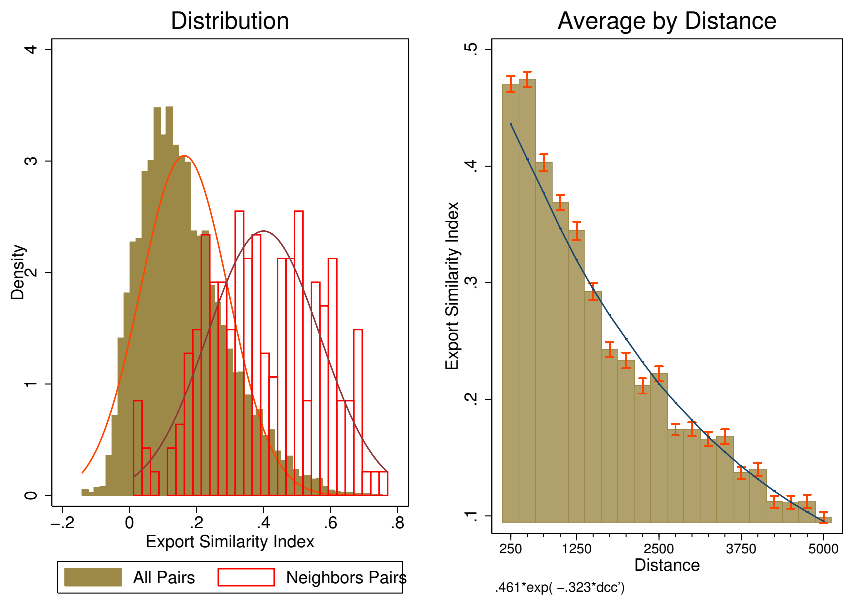
Variable	N	Mean	sd
Similarity Index	7503	0.169	0.137
Similarity Index (NPRB)	7503	0.148	0.132
Simple Distance (Km)	7503	7338.655	4389.738
Ln Simple Distance (Km)	7503	8.649	0.817
Share a Border	7503	0.025	0.155
Same Language	7503	0.103	0.305
Have/Had Colonial Relationship	7503	0.015	0.123
Common Colonizer	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 Pysical 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

markets. Nevertheless, our analysis is robust to using the F&K similarity index and other variations of our own similarity index (see section A.1.3).

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 Figure 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 distribution functions for the two samples based on the histograms. The figure

Figure 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 Neighbors 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.

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 from the sample products that are pinned

[5] This difference in means is statistically significant, with t=−24.16.

Table 2: Lall Classification

Lall Classification	# 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

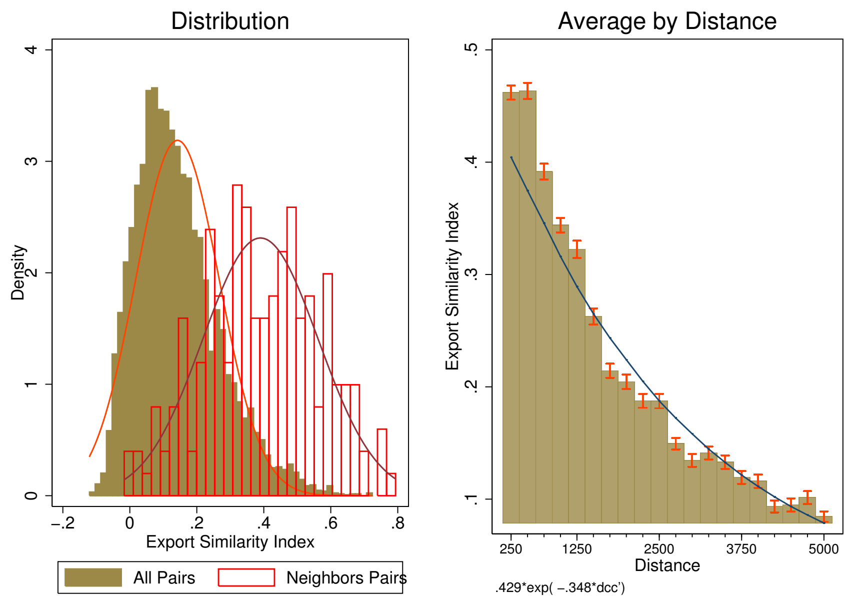
down by geography. We do this by using the technological classification suggested by Lall (2000) that divides products in the categories presented in Table 2.

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 as PRB products those that are classified as Gold, Primary Products and Resource Based Manufactures (categories 1 thru 4 in Table 2), whereas NPRB products are the ones contained in all other categories.

Figure 2 reproduces Figure 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 relation between export similarity and geographical distance is equally strong, suggesting that the observed export similarity among neigh-

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

Figure 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 Neighbors 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.

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

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

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 are 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 interpretation 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

Table 3: Correlates of the Export Similarity Index (Year 2000)

	All	All	All	NPRB	NPRB	NPRB
Ln Simple Distance (Km)	-0.5563	-0.3233	-0.3156	-0.5901	-0.3673	-0.3477
	(0.017)***	(0.022)***	(0.023)***	(0.018)***	(0.022)***	(0.025)***
Share a Border		0.8023	0.6500		0.9037	0.7740
		(0.084)***	(0.084)***		(0.090)***	(0.094)***
Same Region		0.4162	0.1223		0.3551	0.0990
		(0.038)***	(0.043)***		(0.041)***	(0.048)**
Same Language			0.0825			0.0696
			(0.042)*			(0.046)
Have/Had Colonial Relationship			0.0156			-0.0233
			(0.084)			(0.081)
Common Colonizer			0.0334			0.0418
			(0.052)			(0.058)
Abs. Dif. Ln GDP Per Capita (PPP)			-0.2915			-0.2442
			(0.027)***			(0.029)***
Abs. Dif. Ln Population			-0.0940			-0.1121
			(0.011)***			(0.012)***
Log Total Bilateral Trade (Imp + Exp)			-0.0312			-0.0250
			(0.002)***			(0.002)***
Abs. Dif. Ln Pysical Capital Per Worker			-0.0773			-0.0773
			(0.024)***			(0.026)***
Abs. Dif. Ln Human Capital Per Worker			-0.4105			-0.4340
			(0.048)***			(0.050)***
Abs. Dif. Ln Land Per Worker			-0.2206			-0.2330
			(0.032)***			(0.033)***
N	7503	7503	5460	7503	7503	5460
r2	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.

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

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 endowments 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 are expected to have a similarity index that is 0.55 standard deviations 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 deviations from the mean. This result is always robust to the several tests we run in section A.1. 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, higher by 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 from neighboring countries in the same region to have, on average, a similarity index higher by roughly 1.2 standard deviations above the mean relative to non-neighbors from different regions for all goods and 1.25 for NPRB products, not counting the fact that neighbors are at a shorter distance than the average pair of countries. These variables are always strongly significant

[7] $\frac{123 \times 122}{2}$.

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 by about a third the coefficient on the three distance variables but these remain strongly significant in all robustness checks. Coefficients for same language, and colonial relationship are not statistical significant when including other controls. 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.

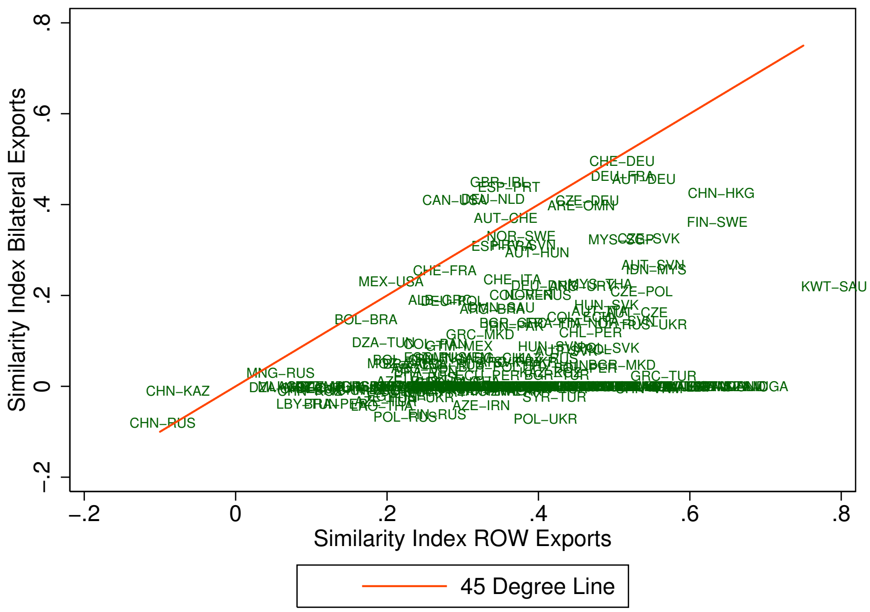
These figures suggest that geographic neighbors have similar export baskets, even when accounting for country fixed effects, common characteristics on culture and institutions (through the inclusion of data on colonial history and language), trade between them and differences in their income, populations 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 Hecksher-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 (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 & Krugman, 1985). Also, in a world with integrated supply chains, the similarity in exports could be a result of neighboring countries trading inputs that might be 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 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 the bilateral export among each pair of countries, and a similarity export based on exports of each pair of countries to the rest of the world (excluding the bilateral exports). Figure 3 plots the two measures using data from 2000, and neighboring country pairs only. As it is clear in the figure, neighbors are remarkably more similar in terms of what they export to the rest of the world than what they trade among 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, sharing a border and being in the same region holds when looking only at exports to the rest of the world as the basis for

Figure 3: Neighbors Similarity (on bilateral exports vis-à-vis ROW exports)



This figure uses data from year 2000. It shows a scatterplot, where every observation is a country-pair. On the horizontal axis, it 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). On the vertical axis, it 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).

similarity between all pairs of countries. Moreover, the last column on this table includes as a regressor the bilateral similarity index. Its inclusion does not qualitatively change the results. Section A.3 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 similarity in exports. This is a puzzle not easily explained by traditional frameworks, that would 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 Exports 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 that is presently active?

To explore this issue, we present a dynamic analysis that studies the role of neighbors in the ability of countries to add a particular good to their export basket or to expand their comparative advantage in it. 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

Table 4: Correlates of the ROW Export Similarity Index (Year 2000)

	ROW	ROW	ROW	ROW
Ln Simple Distance (Km)	-0.5863	-0.3664	-0.3593	-0.3325
	(0.018)***	(0.023)***	(0.025)***	(0.025)***
Share a Border		0.9325	0.8089	0.6868
		(0.093)***	(0.096)***	(0.091)***
Same Region		0.3390	0.0906	0.0949
		(0.041)***	(0.048)*	(0.046)**
Same Language			0.0755	0.0869
			(0.046)	(0.045)*
Have/Had Colonial Relationship			0.1095	0.0929
			(0.097)	(0.089)
Common Colonizer			-0.0422	-0.0251
			(0.054)	(0.053)
Abs. Dif. Ln GDP Per Capita (PPP)			-0.3214	-0.2985
			(0.028)***	(0.027)***
Abs. Dif. Ln Population			-0.0655	-0.0730
			(0.012)***	(0.012)***
Log Total Bilateral Trade (Imp + Exp)			-0.0215	-0.0216
			(0.002)***	(0.002)***
Abs. Dif. Ln Pysical Capital Per Worker			-0.0253	-0.0284
			(0.025)	(0.025)
Abs. Dif. Ln Human Capital Per Worker			-0.3472	-0.3400
			(0.050)***	(0.049)***
Abs. Dif. Ln Land Per Worker			-0.2991	-0.2831
			(0.036)***	(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.
$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

dataset described in section 2.1, with 100 countries[8]. We divide our sample in four periods: 1970-1980, 1980-1990, 1990-2000 and 2001-2008[9]. In 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 $RCA_{c,p} \leq 0.1$ to $RCA_{c,p} \geq 1$ in 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 in the share in world trade of that product[11].

To avoid noise in our definition, we restrict jumps to two conditions. First, a jump needs to keep an 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 intend to rule

[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 A.5 we replicate the results using different thresholds to test robustness. We present results by defining jumps as achieving an $RCA_{c,p} \geq 2$ and an $RCA_{c,p} \geq 5$.

[12]In section A.5 we present robustness checks that limit the sample to observations for

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. It 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 in 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 present 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, besides Germany, all countries in the list are developing countries mostly 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, repeats many countries from the top list, but includes also many African countries. The presence of Germany in the list calls our attention, given that is the only developed country in the list. However, its high ranking in the period 1990-2000 could well be due classification errors associated with

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.

Table 5: Frequency of Jumps by Product

SITC4	Product Name	Period	O	B	P
NPRB Products					
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
PRB Products					
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
NPRB Products				
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
PRB Products				
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

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:

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

where $J_{c,p,t\to T}$ 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 vari-

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

able of interest, $ln(RCA_{c_N,p,t})$, is the natural logarithm of the RCA of the neighbor with the the largest RCA in product p for country of 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]. In order 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 as a control 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 (2006) and used in Hidalgo et. al. (2007). It measures the intensity with which a country exports products that are strongly co-exported in other countries with the 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

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

product to its export basket (Hausmann & Klinger, 2007; C. A. Hidalgo et al. 2007). We use it 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, using the neighbor that has 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, Trajtenberg and Henderson (1993), we created a control group to our sample in order to test for 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 each product across all iterations. We compare the results of our dataset with

[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 exlcude that variable from our controls.

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 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 be using interchangeably. First, the continuous variable "Ln Maximum RCA [of] Neighbors", which is the natural logarithm of $RCA_{c_N,p,t}$ (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 $RCA_{c_N,p,t} \geq 1$ 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

Table 7: Summary Statistics Dynamics of Export Similarity (1970-2008)

	All			**NPRB**		
Variable	N	Mean	sd	N	Mean	sd
New Product (10 Years)	173433	0.015	0.123	90811	0.016	0.125
Baseline Ln RCA	173433	-2.227	0.157	90811	-2.224	0.159
Baseline Density	173433	0.087	0.087	90811	0.073	0.077
Growth Rate RCA	173433	3.105	9.873	90811	3.171	9.408
Zero RCA	173433	0.650	0.477	90811	0.631	0.483
Baseline bilateral imports (p)	173433	0.827	2.215	90811	0.829	2.190
Max RCA Neighbors	173433	2.290	31.057	90811	0.814	4.643
Ln Max RCA Neighbors	173433	-1.256	1.436	90811	-1.397	1.228
Neighbor Exports	173433	0.172	0.377	90811	0.137	0.344

using 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 points 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%). For Panel B of table 8 estimates that if a country has one neighbor that 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 for the average product an increase of roughly 65% in the probability of "jumping" (from 1.5% to 2.5%) [18].

[18] See section A.5 for a number of robustness checks of these results, varying the defi-

Table 8: Dynamics of Exports Similarity

Panel A: Continous Independent Variable				
	Real		Control	
	All	NPRB	All	NPRB
Ln Max RCA Neighbors	0.0037	0.0040	-0.0019	-0.0060
	(0.000)***	(0.001)***	(0.003)	(0.007)
Baseline Ln RCA	0.0073	-0.0035	0.0091	-0.0042
	(0.004)**	(0.005)	(0.004)**	(0.006)
Baseline Density	0.1302	0.2266	0.1557	0.2536
	(0.034)***	(0.076)***	(0.033)***	(0.075)***
Growth Rate RCA (t-1)	-0.0006	-0.0005	-0.0006	-0.0005
	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Zero RCA (t-1)	0.0062	0.0078	0.0056	0.0073
	(0.001)***	(0.002)***	(0.001)***	(0.002)***
N	173433	90811	173433	90811
r2	0.08	0.11	0.08	0.11
Panel B: Binary Independent Variable				
	Real		Control	
	All	NPRB	All	NPRB
Neighbor Exports	0.0106	0.0103	0.0008	-0.0002
	(0.001)***	(0.002)***	(0.001)	(0.002)
Baseline Ln RCA	0.0082	-0.0029	0.0091	-0.0042
	(0.004)**	(0.005)	(0.004)**	(0.006)
Baseline Density	0.1389	0.2343	0.1555	0.2523
	(0.034)***	(0.076)***	(0.033)***	(0.075)***
Growth Rate RCA (t-1)	-0.0006	-0.0005	-0.0006	-0.0005
	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Zero RCA (t-1)	0.0060	0.0077	0.0056	0.0072
	(0.001)***	(0.002)***	(0.001)***	(0.002)***
N	173433	90811	173433	90811
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 have 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 ammount 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.

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

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 uses as its dependent variable the compound average annual growth rate of RCA for the same time periods as before. We use all the observations in the dataset without the low RCA restriction we used for the extensive margin. The table shows 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. As before, we do the analysis using both a continuous and a binary independent variable. 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 A8 in section A.5 shows that this result is robust to the use of export growth rather than RCA growth as the dependent variable.

We repeat this analysis for different regions, periods and types of products (Table 10). Each row in the table presents results for a different cut of the data. The left panel 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

nition of the LHS variable, the method, the sample used and the dataset. All tests show full robustness with the results presented here.

Table 9: Dynamics of Exports Similarity (RCA Growth)

Panel A: Continous Independent Variable				
	Real		Control	
	All	NPRB	All	NPRB
Ln Max RCA Neighbors	0.7334	0.7113	-0.7788	-0.4180
	(0.045)***	(0.069)***	(0.383)**	(0.586)
Baseline Ln RCA	-3.9685	-5.0351	-3.8515	-4.9510
	(0.113)***	(0.186)***	(0.119)***	(0.191)***
Baseline Density	23.0227	29.3487	28.1252	33.7179
	(2.444)***	(3.740)***	(2.731)***	(3.817)***
Growth Rate RCA (t-1)	-0.0381	-0.0002	-0.0397	-0.0010
	(0.007)***	(0.012)	(0.008)***	(0.013)
Zero RCA (t-1)	-1.0230	-0.7455	-1.1853	-0.8229
	(0.133)***	(0.183)***	(0.142)***	(0.197)***
N	262017	136929	262017	136929
r2	0.20	0.26	0.20	0.25
Panel B: Binary Independent Variable				
	Real		Control	
	All	NPRB	All	NPRB
Neighbor Exports	1.7851	1.5242	0.0092	0.1153
	(0.119)***	(0.160)***	(0.112)	(0.139)
Baseline Ln RCA	-3.8980	-4.9979	-3.8311	-4.9397
	(0.112)***	(0.186)***	(0.116)***	(0.189)***
Baseline Density	24.2354	30.3715	28.0300	33.6793
	(2.451)***	(3.715)***	(2.739)***	(3.819)***
Growth Rate RCA (t-1)	-0.0387	0.0001	-0.0395	-0.0009
	(0.007)***	(0.012)	(0.008)***	(0.013)
Zero RCA (t-1)	-1.1038	-0.7782	-1.1845	-0.8211
	(0.132)***	(0.183)***	(0.142)***	(0.197)***
N	262017	136929	262017	136929
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 have 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 ammount 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.
$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

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 each different row recalculates the coefficient for each cut of the data.

From table 10 we find that our results are, in fact, driven mostly by developing countries, given that, both in the left and right panel, the estimator for β is statistically significant only for the non OECD countries, as opposed to OECD countries. When we divide the world in 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 the different periods of time, the confidence intervals in both panels show that the estimated coefficients are significant and stable across all periods.

Finally, we divide the sample in 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, minerals fuels and several manufacturing categories.

Table 10: Odds Ratio

		RCA Neighbor				**Neighbor Exporter**			
	N	β 95% C.I.	Top 25%	Bottom 75%	Ratio	β 95% C.I.	1	0	Ratio
All Observations	173433	(0.003, 0.005)	2.52%	1.20%	**2.10**	(0.008, 0.014)	2.88%	1.25%	**2.31**
Non OECD	147081	(0.003, 0.004)	2.40%	0.93%	**2.58**	(0.007, 0.014)	2.85%	1.00%	**2.84**
OECD	26352	(-0.001, 0.004)	2.96%	2.77%	**1.07**	(-0.003, 0.008)	3.01%	2.76%	**1.09**
East Asia & Pacific	20636	(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	33918	(0.001, 0.005)	2.35%	1.01%	**2.31**	(0.003, 0.013)	2.45%	1.01%	**2.42**
Middle East & N. Africa	34394	(-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	54751	(0.000, 0.003)	1.89%	0.59%	**3.21**	(0.002, 0.013)	2.98%	0.67%	**4.43**
Western Europe	15635	(0.000, 0.005)	3.50%	3.22%	**1.09**	(-0.005, 0.011)	3.52%	3.21%	**1.10**
Period 1970-1980	31462	(0.000, 0.003)	1.35%	0.54%	**2.50**	(0.001, 0.011)	1.63%	0.59%	**2.77**
Period 1980-1990	50133	(0.004, 0.009)	3.12%	2.09%	**1.49**	(0.010, 0.025)	3.70%	2.14%	**1.73**
Period 1990-2000	48012	(0.002, 0.004)	2.41%	0.55%	**4.36**	(0.007, 0.016)	2.96%	0.62%	**4.78**
Period 2001-2008	43826	(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.	20291	(-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	25478	(0.001, 0.003)	2.42%	1.15%	**2.10**	(0.003, 0.013)	2.53%	1.18%	**2.14**
Food & live animals	20740	(0.002, 0.006)	2.62%	1.34%	**1.96**	(0.002, 0.014)	2.68%	1.34%	**2.00**
Machinery & transport equipment	34628	(-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	41490	(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	18039	(0.002, 0.011)	4.34%	1.80%	**2.42**	(0.001, 0.023)	5.59%	1.91%	**2.92**

4 Interpretation of the results

As we argued in the introduction, the literature on knowledge diffusion has documented its rapid decline with distance. If this assertion is true, neighboring countries should share more 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 the similarity in knowledge among neighbors should map into a 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 looking only at goods that are not pinned down by geology or climate (NPRB) and after taking account of the similarity 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 would have the implication that knowledge acquisition would occur preferentially in countries that have 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 happens 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 in which knowledge diffusion decays strongly with distance, our results could be driven by factors other than knowledge diffusion. The documented similar dynamics in the evolution of export baskets among 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, countries may be on a similar development trajectory, moving –for instance– from agriculture to light manufactures and into more complex products but one country is ahead of the other. As a consequence, 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 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 among themselves and 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 sure 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 would be expected 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 be a reflection of the similarity in factor endowments. But after taking account of 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 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 as the greater intensity of trade at short distances should have incentivized 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 has left 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 at the aggregate level that imports-weighted foreign R&D investment are correlated with total factor productivity growth. But these results per se are not enough to account for our observations: we require a 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 cause knowledge spillovers. If this is so it would be translated in a coevolution in trends of comparative advantage 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. Borensztein, De Gregorio and Lee (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 & 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 shown what is the contribution that FDI plays in 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, 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 through 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, the limited geographic knowledge diffusion is an important observation in its own right. It may lurk behind 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 definitely not an economic insignificant phenomenon. It is more than a side effect. It can shape the evolution of the comparative advantage of nations.

References

Aitken, B.J., and A.E. Harrison. “Do domestic firms benefit from direct foreign investment? Evidence from Venezuela.” *American Economic Review* 89, 3: (1999) 605–618.

Alvarez, Fernando, F Buera, and RE Lucas. “Idea Flows, Economic Growth, and Trade.”, 2012.

Andersen, Thomas Barnebeck, and Carl-Johan Dalgaard. “Flows of people, flows of ideas, and the inequality of nations.” *Journal of Economic Growth* 16, 1: (2011) 1–32.

Anderson, JE. “A Theoretical Foundation for the Gravity Equation.” *American Economic Review* 69, 1: (1979) 106–116.

Anderson, JE, and Eric Van Wincoop. “Gravity with gravitas: A solution to the border puzzle.” *American Economic Review* 93, 1: (2001) 170–191.

Arrow, Kenneth J. “Classificatory Notes on the Production and Transmission of Technologcal Knowledge.” *The American Economic Review* 59, 2: (1969) 29–35.

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

Bergstrand, JH. “The Gravity Equation in International Trade: some Microeconomic Foundation and Empirical Evidence.” *The review of economics and statistics* 67, 3: (1985) 474–481.

———. “The generalized gravity equation, monopolistic competition, and the factor-proportions theory in international trade.” *The review of economics and statistics* 71, 1: (1989) 143–153.

Borensztein, E., J. De Gregorio, and J-W. Lee. “How does foreign direct investment affect economic growth?” *Journal of International Economics* 45, 1: (1998) 115–135.

Bottazzi, Laura, and Giovanni Peri. “Innovation and spillovers in regions : Evidence from European patent data.” *European Economic Review* 47: (2003) 687–710.

Brander, J, and P Krugman. “A reciprocal dumping model of international trade.” *Journal of International Economics* 15, 1983: (1983) 313–321.

Brander, JA. “Intra-Industry trade in identical commodities.” *Journal of International Economics* 11, 1981: (1981) 1–14.

Branstetter, Lee G. “Are knowledge spillovers international or intranational in scope ? Microeconometric evidence from the U.S. and Japan.” *Journal of International Economics* 53: (2001) 53–79.

———. “Is foreign direct investment a channel of knowledge spillovers? Evidence from Japan’s FDI in the United States.” *Journal of International Economics* 68, 2: (2006) 325–344.

Coe, David T, and Elhanan Helpman. “International R&D spillovers.” *European Economic Review* 2921, 94.

Coe, David T., Elhanan Helpman, and Alexander W. Hoffmaister. "International R&D spillovers and institutions." *European Economic Review* 53, 7: (2009) 723–741.

Davis, Donald R. "Intra-industry trade: A Heckscher-Ohlin-Ricardo approach." *Journal of International Economics* 39, 3-4: (1995) 201–226.

Deardorff, Alan. "Determinants of bilateral trade: does gravity work in a neoclassical world?" In *The Regionalization of the World Economy*, edited by Jeffrey Frankel, Chicago: University of Chicago Press, 1998.

Eaton, Jonathan, and Samuel Kortum. "Technology and bilateral trade." *NBER Working Paper Series* , 6253.

Feenstra, Robert C., Robert E. Lipsey, Haiyan Deng, Alyson C. Ma, and Hengyong Mo. "World Trade Flows: 1962-2000." *NBER Working Paper No. 11040* 1962–2000.

Feenstra, Robert C., James R. Markusen, and Andrew K. Rose. "Using the gravity equation to differentiate among alternative theories of trade." *Canadian Journal of Economics/Revue Canadienne d'Economique* 34, 2: (2001) 430–447.

Finger, JM, and ME Kreinin. "A Measure of 'Export Similarity' and Its Possible Uses." *The Economic Journal* 89, 356: (1979) 905–912.

Frankel, Jeffrey Alexander. *Regional Trading Blocs in the World Economic System.* Peterson Institute for International Economics, 1997.

Giroud, Xavier. "Proximity and Investment: Evidence from Plant-Level Data." *The Quarterly Journal of Economics* 128, 2: (2012) 861–915.

Haskel, JE, SC Pereira, and MJ Slaughter. "Does inward foreign direct investment boost the productivity of domestic firms?" *The Review of Economics and Statistics* 89, 3: (2007) 482–496.

Hausmann, Ricardo, César A Hidalgo, Sebastián Bustos, Michele Coscia, Sarah Chung, Juan Jímenez, Alexander Simoes, and Muhammed A. Yildirim. *The Atlas of Economic Complexity: Mapping Paths to Prosperity*. Cambridge, MA, 2011.

Hausmann, Ricardo, and Bailey Klinger. "The Structure of the Product Space and the Evolution of Comparative Advantage." *CID Working Paper Series* , 146.

Helpman, Elhanan, and Paul Krugman. *Market Structure and International Trade*. Cambridge: MIT Press, 1985.

Hidalgo, César A, Bailey Klinger, AL Barabási, and Ricardo Hausmann. "The product space conditions the development of nations." *Science (New York, N.Y.)* 317, 5837: (2007) 482–7.

Jaffe, A.B., M. Trajtenberg, and R. Henderson. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *The Quarterly Journal of Economics* 108, 3: (1993) 577.

Javorcik, Beata S. "Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages." *The American Economic Review* 94, 3: (2004) 605–627.

Kehoe, Timothy J., and Kim J. Ruhl. "How Important Is the New Goods Margin in International Trade?" *Journal of Political Economy* 121, 2: (2013) 358–392.

Keller, Wolfgang. "Geographic localization of international technology diffusion." *American Economic Review* 92, 1: (2002) 120–142.

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

Keller, Wolfgang, and Stephen R Yeaple. "Multinational enterprises, international trade, and productivity growth: firm-level evidence from the United States." *The Review of Economics and Statistics* 91, November: (2009) 821–831.

Keller, Wolfgang, and Stephen Ross Yeaple. "The Gravity of Knowledge." *American Economic Review* 103, 4: (2013) 1414–1444.

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

Lall, Sanjaya. "The Technological Structure and Performance of Developing Country Manufactured Exports, 1985-98." *Oxford Development Studies* 28, 3: (2000) 337–369.

Leamer, EE, and James Levinsohn. "International Trade Theory: The Evidence." In *Handbook of International economics*, edited by Gene M.

Grossman, and K Rogoff, Elsevier B.V., 1995, volume III, chapter 26, 1341–1375.

Linder, Staffan Burenstam. *An essay on trade and transformation.* Stockholm: Almqvist & Wicksell, 1961.

Loungani, Prakash, Ashoka Mody, and Assaf Razin. “The global disconnect: the rle of transactional distance and scale economies in gravity equations.” *Scottish Journal of Political Economy* 49, 5: (2002) 526–543.

Mayer, Thierry, and Soledad Zignago. “Notes on CEPII distances measures : The GeoDist database.” *CEPII Working Paper* , 25.

Polanyi, M. *Personal knowledge: Towards a post-critical philosophy.* London, UK: Routledge, 1962.

Portes, Richard, and Hélène Rey. “The determinants of cross-border equity flows.” *Journal of International Economics* 65, 2: (2005) 269–296.

Rivera-Batiz, L.A., and P.M. Romer. “Economic integration and endogenous growth.” *The Quarterly Journal of Economics* 106, 2: (1990) 531–555.

Shirotori, M, B Tumurchudur, and O Cadot. “Revealed Factor Intensity Indices at the Product Level.” *Policy Issues in International Trade and Commodities* 2010, 44.

Stein, Ernesto, and Christian Daude. “Longitude matters: Time zones and the location of foreign direct investment.” *Journal of International Economics* 71, 1: (2007) 96–112.

Tinbergen, J. "Shaping the world economy." *The International Executive* 5, 1: (1963) 27–30.

United Nations. "COMTRADE database.", 2010. http://comtrade.un.org/.

Venables, Anthony J. "Trade and trade policy with imperfect competition: The case of identical products and free entry." *Journal of International Economics* 19, 1-2: (1985) 1–19.

World Bank. "World Development Indicators Online.", 2010. http://data.worldbank.org/.

Zahler, Andres. "Decomposing world export growth and the relevance of new destinations." *CID Working Paper Series* 2007, 20.

Zipf, George Kingsley. "The P1 P2/D Hypothesis: On the Intercity Movement of Persons." *American Sociological Review* 11, 6: (1946) pp. 677–686.

A Appendix

A.1 Robustness of the Stylized Facts

A.1.1 Variations in ε

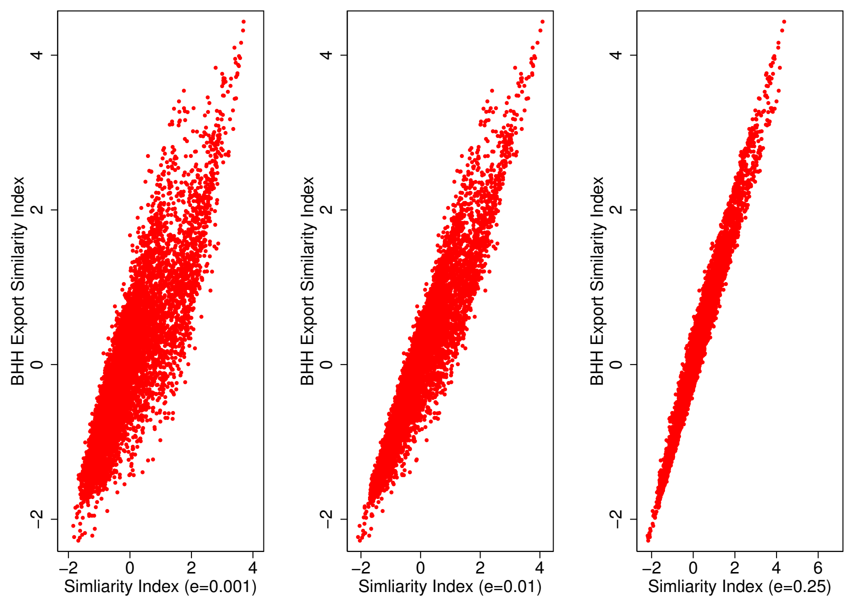
In this section we address robustness concerns with regard to our choice of $\varepsilon = 0.1$ in the calculation of $S_{c,c'}$ based on equation (2). Our original choice of $\varepsilon = 0.1$ allow us to deal with values of $RCA_{c,p} = 0$ in the log-transformation. However, this raises concerns that the choice of ε might drive the results we presented in table 3. Therefore, we recalculated $S_{c,c'}$ defining ε as 0.001, 0.01 and 0.25. Figure A1 shows the correlation of these new measures and the original $S_{c,c'}$ (using $\varepsilon = 0.1$). As can be seen, the different choices of ε are highly correlated with our original choice.

To convince the readers of the robustness of our choice of ε, we reproduce figures 1 and 2, along with table 3, using this time $S_{c,c'}$ defined with each of the new ε values. Results are presented in figures A2-A4, and tables A1-A3. For all variations of ε, the results are qualitatively robust to our original measure.

A.1.2 Indexed RCA Export Similarity Index

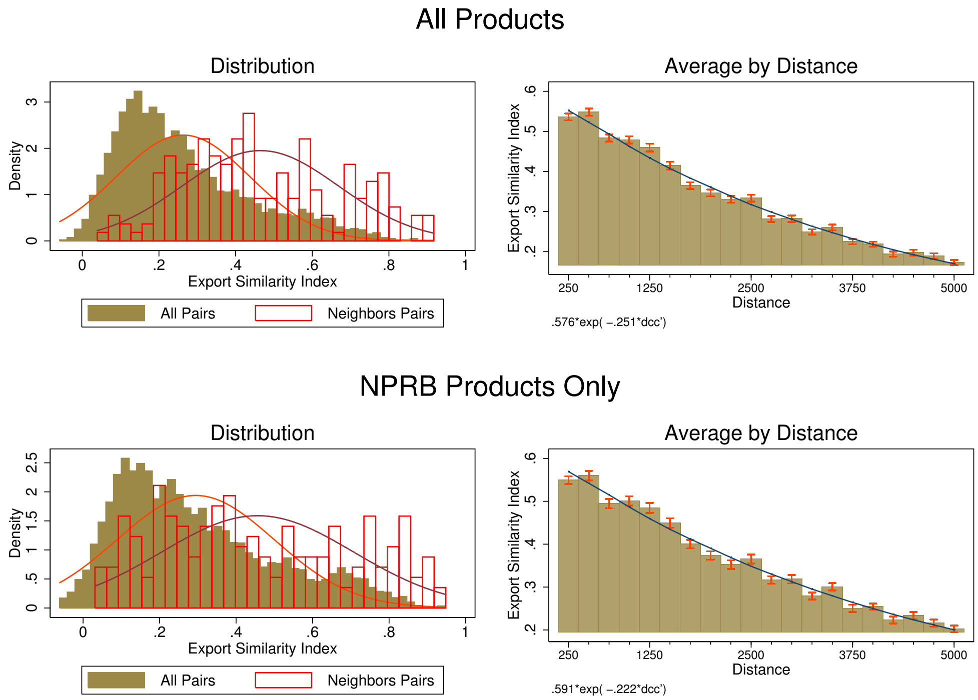
To erase concerns regarding the log-transformation of the RCA vectors in computing $S_{c,c'}$ based on equation (2), we constructed a variation of our similarity index, which we call the Indexed RCA Export Similarity Index, by substituting $r_{c,p}$ in equation (2), by:

Figure A1: Scatter of $S_{c,c'}$ (with $\varepsilon = \{0.001, 0.01, 0.25\}$) vs. $S_{c,c'}$ (with $\varepsilon = 0.1$)



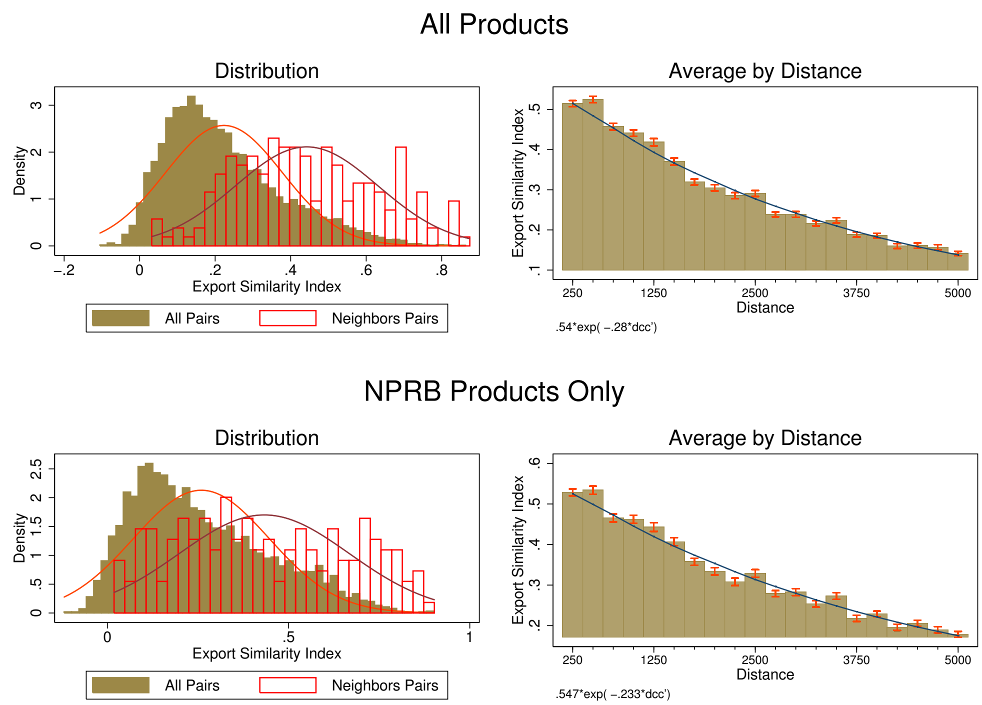
The figure contains three scatterplots comparing our original (BHH) Export Similarity Index and the recalculations of the Similarity Index using different values of ε (0.001, 0.01 and 0.25 from left to right), for all country pairs, in year 2000.

Figure A2: Stylized Facts Similarity Index ($\varepsilon = 0.001$)



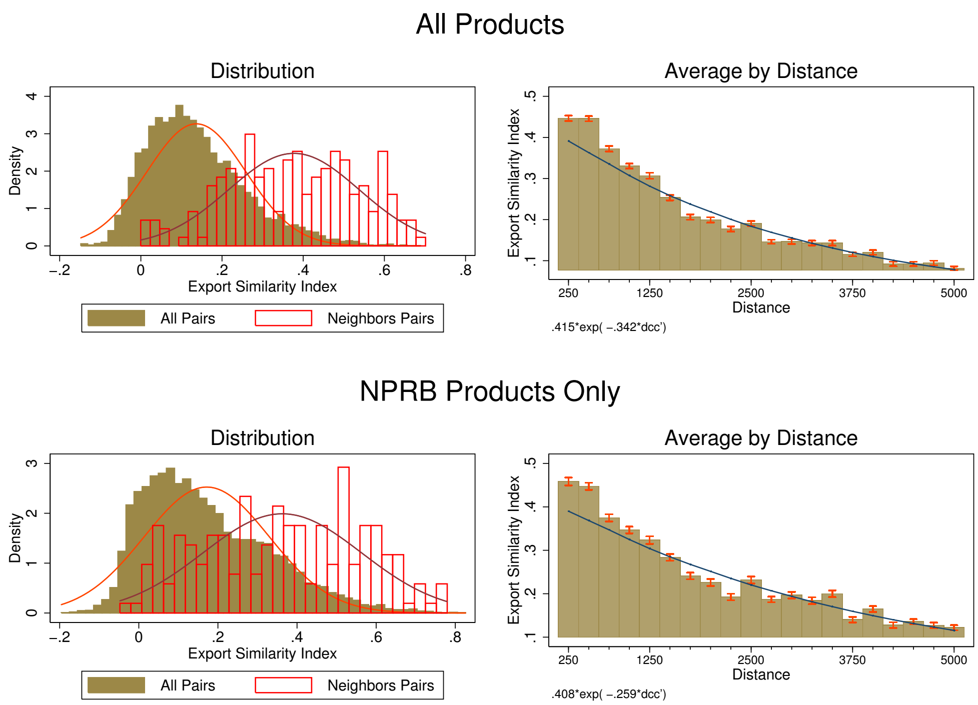
The left panel of the figure shows the distributions (in year 2000) of the Export Similarity Index (using $\varepsilon = 0.001$) for All (not neighbors) Country Pairs, and for Neighbors Pairs only. The right panel shows the average Export Similarity Index (using $\varepsilon = 0.001$) for country pairs in each bracket of distance between 250 km. to 5000 km. The upper figures use the Export Similarity Index (using $\varepsilon = 0.001$) for all products, and the lower figures use the Export Similarity Index (using $\varepsilon = 0.001$) for NPRB products only.

Figure A3: Stylized Facts Similarity Index ($\varepsilon = 0.01$)



The left panel of the figure shows the distributions (in year 2000) of the Export Similarity Index (using $\varepsilon = 0.01$) for All (not neighbors) Country Pairs, and for Neighbors Pairs only. The right panel shows the average Export Similarity Index (using $\varepsilon = 0.01$) for country pairs in each bracket of distance between 250 km. to 5000 km. The upper figures use the Export Similarity Index (using $\varepsilon = 0.01$) for all products, and the lower figures use the Export Similarity Index (using $\varepsilon = 0.01$) for NPRB products only.

Figure A4: Stylized Facts Similarity Index ($\varepsilon = 0.25$)



The left panel of the figure shows the distributions (in year 2000) of the Export Similarity Index (using $\varepsilon = 0.25$) for All (not neighbors) Country Pairs, and for Neighbors Pairs only. The right panel shows the average Export Similarity Index (using $\varepsilon = 0.25$) for country pairs in each bracket of distance between 250 km. to 5000 km. The upper figures use the Export Similarity Index (using $\varepsilon = 0.25$) for all products, and the lower figures use the Export Similarity Index (using $\varepsilon = 0.25$) for NPRB products only.

Table A1: Correlates of the Similarity Index, using $\varepsilon = 0.001$

	All	All	All	NPRB	NPRB	NPRB
Ln Simple Distance (Km)	-0.3691	-0.1737	-0.1686	-0.2458	-0.0777	-0.0899
	(0.013)***	(0.017)***	(0.016)***	(0.012)***	(0.016)***	(0.016)***
Share a Border		0.4396	0.3029		0.2830	0.1455
		(0.062)***	(0.059)***		(0.055)***	(0.054)***
Same Region		0.4227	0.1510		0.3934	0.1601
		(0.029)***	(0.030)***		(0.026)***	(0.028)***
Same Language			0.0619			0.0698
			(0.031)**			(0.029)**
Have/Had Colonial Relationship			-0.0379			-0.0151
			(0.058)			(0.053)
Common Colonizer			0.0874			0.1061
			(0.041)**			(0.039)***
Abs. Dif. Ln GDP Per Capita (PPP)			-0.2030			-0.1994
			(0.021)***			(0.021)***
Abs. Dif. Ln Population			-0.1063			-0.0850
			(0.009)***			(0.008)***
Log Total Bilateral Trade (Imp + Exp)			-0.0280			-0.0265
			(0.002)***			(0.002)***
Abs. Dif. Ln Pysical Capital Per Worker			-0.1341			-0.0959
			(0.019)***			(0.019)***
Abs. Dif. Ln Human Capital Per Worker			-0.3121			-0.1752
			(0.037)***			(0.035)***
Abs. Dif. Ln Land Per Worker			-0.1149			-0.0552
			(0.022)***			(0.021)***
N	7503	7503	5460	7503	7503	5460
r2	0.58	0.60	0.76	0.64	0.65	0.77

This table uses a normalized version of the Similarity Index using $\varepsilon = 0.001$ (with mean zero and unit standard deviation) as the dependent variable. Columns 1-3 estimates model (3) using the Similarity Index using $\varepsilon = 0.001$ with all products, while columns 4-6 uses the Similarity Index using $\varepsilon = 0.001$ computed with NPRB products only. All regressions include country dummies. Standard errors are clustered at the country-pair level.

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

Table A2: Correlates of the Similarity Index, using $\varepsilon = 0.01$

	All	All	All	NPRB	NPRB	NPRB
Ln Simple Distance (Km)	-0.4473	-0.2346	-0.2276	-0.2856	-0.1046	-0.1154
	(0.014)***	(0.019)***	(0.018)***	(0.013)***	(0.017)***	(0.018)***
Share a Border		0.5747	0.4277		0.3719	0.2234
		(0.070)***	(0.068)***		(0.061)***	(0.061)***
Same Region		0.4297	0.1410		0.4026	0.1528
		(0.033)***	(0.035)***		(0.028)***	(0.031)***
Same Language			0.0738			0.0769
			(0.035)**			(0.032)**
Have/Had Colonial Relationship			-0.0172			0.0286
			(0.067)			(0.061)
Common Colonizer			0.0706			0.1039
			(0.045)			(0.044)**
Abs. Dif. Ln GDP Per Capita (PPP)			-0.2434			-0.2556
			(0.024)***			(0.024)***
Abs. Dif. Ln Population			-0.1043			-0.0800
			(0.010)***			(0.009)***
Log Total Bilateral Trade (Imp + Exp)			-0.0301			-0.0266
			(0.002)***			(0.002)***
Abs. Dif. Ln Pysical Capital Per Worker			-0.1152			-0.0634
			(0.021)***			(0.021)***
Abs. Dif. Ln Human Capital Per Worker			-0.3692			-0.1779
			(0.041)***			(0.039)***
Abs. Dif. Ln Land Per Worker			-0.1606			-0.0856
			(0.026)***			(0.024)***
N	7503	7503	5460	7503	7503	5460
r2	0.50	0.52	0.70	0.57	0.58	0.71

This table uses a normalized version of the Similarity Index using $\varepsilon = 0.01$ (with mean zero and unit standard deviation) as the dependent variable. Columns 1-3 estimates model (3) using the Similarity Index using $\varepsilon = 0.01$ with all products, while columns 4-6 uses the Similarity Index using $\varepsilon = 0.01$ computed with NPRB products only. All regressions include country dummies. Standard errors are clustered at the country-pair level.

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

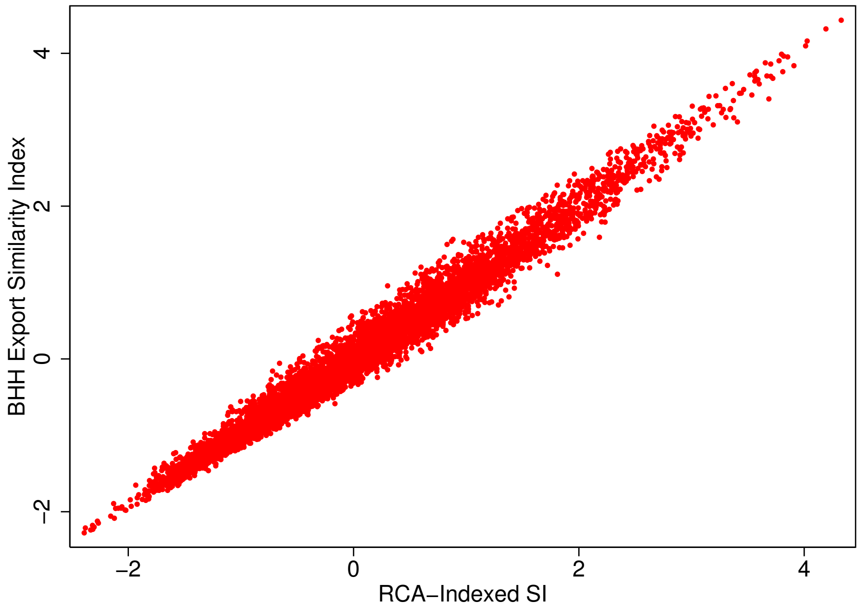
Table A3: Correlates of the Similarity Index, using $\varepsilon = 0.25$

	All	All	All	NPRB	NPRB	NPRB
Ln Simple Distance (Km)	-0.5855	-0.3498	-0.3417	-0.3613	-0.1659	-0.1720
	(0.018)***	(0.023)***	(0.024)***	(0.017)***	(0.022)***	(0.024)***
Share a Border		0.8969	0.7489		0.5880	0.4407
		(0.089)***	(0.090)***		(0.078)***	(0.082)***
Same Region		0.3943	0.1134		0.3759	0.1197
		(0.040)***	(0.046)**		(0.035)***	(0.040)***
Same Language			0.0804			0.0819
			(0.045)*			(0.042)*
Have/Had Colonial Relationship			0.0318			0.1553
			(0.089)			(0.088)*
Common Colonizer			0.0098			0.0623
			(0.055)			(0.055)
Abs. Dif. Ln GDP Per Capita (PPP)			-0.2977			-0.3413
			(0.028)***			(0.030)***
Abs. Dif. Ln Population			-0.0842			-0.0606
			(0.012)***			(0.011)***
Log Total Bilateral Trade (Imp + Exp)			-0.0294			-0.0247
			(0.002)***			(0.002)***
Abs. Dif. Ln Pysical Capital Per Worker			-0.0560			0.0185
			(0.025)**			(0.026)
Abs. Dif. Ln Human Capital Per Worker			-0.3995			-0.1755
			(0.049)***			(0.050)***
Abs. Dif. Ln Land Per Worker			-0.2374			-0.1335
			(0.034)***			(0.033)***
N	7503	7503	5460	7503	7503	5460
r2	0.32	0.35	0.51	0.37	0.38	0.48

This table uses a normalized version of the Similarity Index using $\varepsilon = 0.25$ (with mean zero and unit standard deviation) as the dependent variable. Columns 1-3 estimates model (3) using the Similarity Index using $\varepsilon = 0.25$ with all products, while columns 4-6 uses the Similarity Index using $\varepsilon = 0.25$ computed with NPRB products only. All regressions include country dummies. Standard errors are clustered at the country-pair level.

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

Figure A5: Scatter of Indexed RCA Similarity Index vs. BHH Similarity Index



The figure is a scatterplot comparing our original (BHH) Export Similarity Index and the Indexed RCA Export Similarity Index for all country pairs, in year 2000.

$$r_{c,p} = \frac{RCA_{c,p} - 1}{RCA_{c,p} + 1}$$

Under this definition, $r_{c,p} = 1$ if $RCA_{c,p} \to \infty$, and $r_{c,p} = -1$ if $RCA_{c,p} = 0$. This transformation also deals with fat tails in the original distribution of $RCA_{c,p}$ and hence eliminates the need to do a log-transformation.

Figure A5 shows the high correlation between the original $S_{c,c'}$ and the Indexed RCA Export Similarity Index.

Figure A6 uses the Indexed-RCA $S_{c,c'}$ to replicate figures 1 and 2. Our original $S_{c,c'}$ is robust to this new transformation in terms of its correlation with distance.

We also estimated model (3) using the Indexed-RCA $S_{c,c'}$ (standardized with mean zero and unit standard deviation). The results are presented in table A4. Our static analysis is robust to using this other methodology of measuring similarity in export baskets, in terms of the signs and explanatory power of the variables.

A.1.3 The Finger & Kreinin Export Similarity

We replicate the results shown in the main body of the paper using the Finger & Kreinin (F&K) Export Similarity Index (Finger & Kreinin, 1979). The F&K Similarity Index is constructed using the formula:

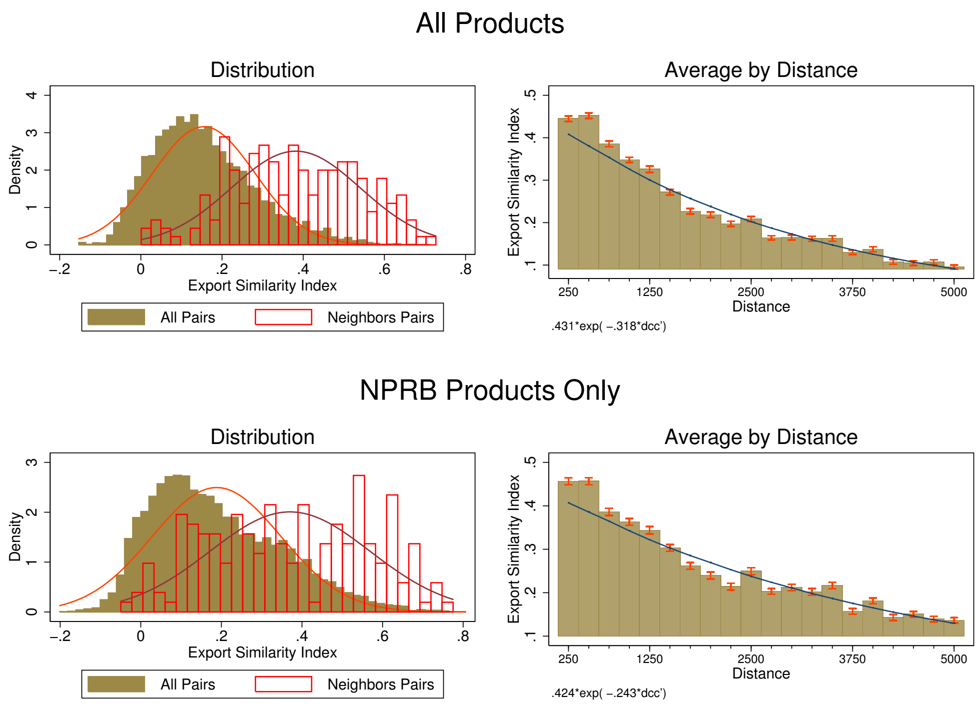
$$S_{c,c'}^{F\&K} = \sum_{p} min(s_p^c, s_p^{c'})$$

where p represents products, c and c' represent any two countries and s_p^c is the share of product p exported by country c out of the total export baskets for country c. Hence, two countries c and c' that export the exact same products in the exact same proportion would have $S_{c,c'}^{F\&K} = 1$.

Figure A7 shows the scatter of both export similarity indices – our own named BBH Export Similarity Index and F&K's one – showing a strong positive correlation between them ($\rho = 0.65$), implying that both indexes capture much of the same information.

Figure A8 shows that our analysis presented in the main body of this paper is robust to using the F&K Similarity Index. The upper panel of the figure presents the distribution of the index for geographical neighbors and

Figure A6: Stylized Facts Indexed RCA Similarity Index



The left panel of the figure shows the distributions (in year 2000) of the Indexed-RCA Export Similarity Index for All (not neighbors) Country Pairs, and for Neighbors Pairs only. The right panel shows the average Indexed-RCA Export Similarity Index for country pairs in each bracket of distance between 250 km. to 5000 km. The upper figures use the Indexed-RCA Similarity Index for all products, and the lower figures use the Indexed-RCA Similarity Index for NPRB products only.

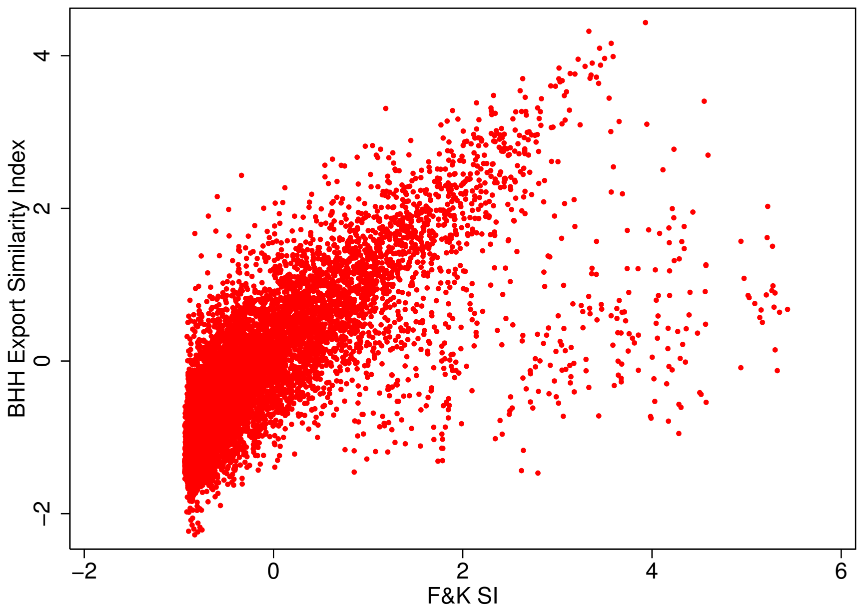
Table A4: Correlates of the Indexed RCA Similarity Index

	All	All	All	NPRB	NPRB	NPRB
Ln Simple Distance (Km)	-0.5587	-0.3299	-0.3210	-0.3407	-0.1493	-0.1582
	(0.017)***	(0.022)***	(0.023)***	(0.016)***	(0.021)***	(0.023)***
Share a Border		0.8009	0.6650		0.5462	0.4052
		(0.086)***	(0.087)***		(0.075)***	(0.078)***
Same Region		0.4046	0.1150		0.3773	0.1289
		(0.039)***	(0.044)***		(0.034)***	(0.039)***
Same Language			0.0964			0.0914
			(0.044)**			(0.041)**
Have/Had Colonial Relationship			0.0255			0.1377
			(0.090)			(0.087)
Common Colonizer			0.0270			0.0770
			(0.053)			(0.054)
Abs. Dif. Ln GDP Per Capita (PPP)			-0.3065			-0.3465
			(0.028)***			(0.029)***
Abs. Dif. Ln Population			-0.0887			-0.0602
			(0.012)***			(0.011)***
Log Total Bilateral Trade (Imp + Exp)			-0.0311			-0.0250
			(0.002)***			(0.002)***
Abs. Dif. Ln Pysical Capital Per Worker			-0.0563			0.0260
			(0.024)**			(0.025)
Abs. Dif. Ln Human Capital Per Worker			-0.3992			-0.1719
			(0.049)***			(0.049)***
Abs. Dif. Ln Land Per Worker			-0.2224			-0.1316
			(0.033)***			(0.032)***
N	7503	7503	5460	7503	7503	5460
r2	0.34	0.36	0.53	0.39	0.40	0.51

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

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

Figure A7: Scatter of F&K Similarity Index vs. BHH Similarity Index



The figure is a scatterplot comparing our original (BHH) Export Similarity Index and the F&K Export Similarity Index for all country pairs, in year 2000.

non neighbors, and the declining relationship of the index with distance, for all products. The lower panel replicates the graphs but using the F&K index computed with NPRB products only. We find that the results are robust to the ones presented in figures 1 and 2.

We also replicated the analysis presented in table 3, which estimates model (3). This time, we use as the dependent variable a normalized version of the F&K Similarity Index with mean zero and unit standard deviation. The results are presented in Table A5. Our static analysis is robust to using this other way of measuring similarity in export baskets.

A.1.4 Proximity Weighted Similarity Index

Another possible way to compute a similarity index which takes into account not only the intensity of exports for each product as measured by the RCA, but it is also weighted by the proximity matrix ϕ. In other words, the index would be as follows:

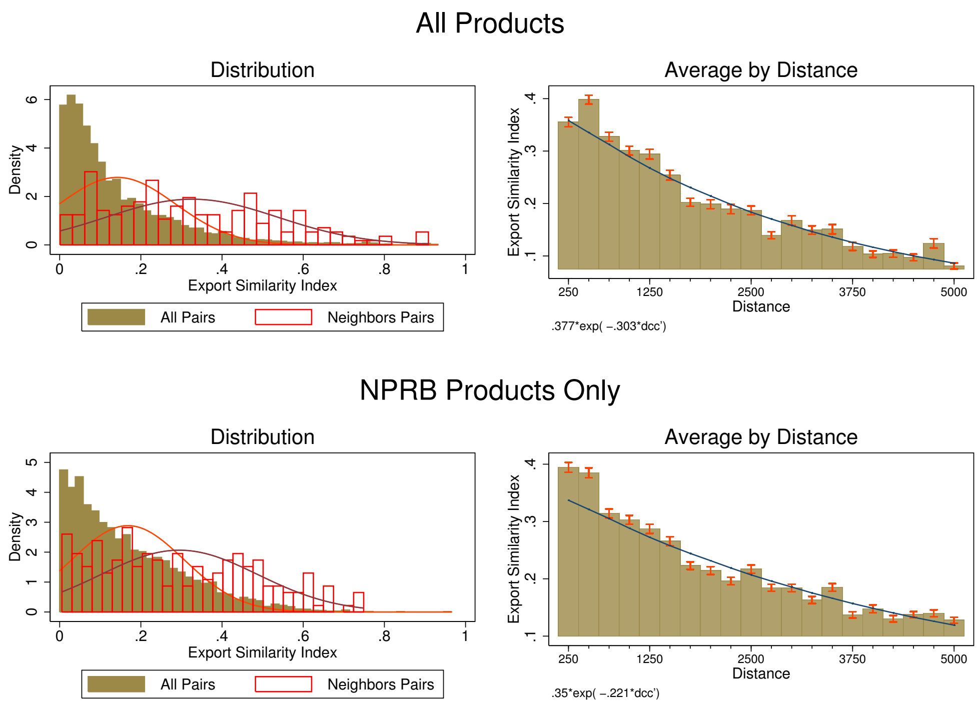
$$S_{c,c'}^{PROX} \equiv \frac{\sum_p (r_{c,p} - \bar{r_c})(r_{c',p} - \bar{r_{c'}})}{\sqrt{\sum_p (r_{c,p} - \bar{r_c})^2 \sum_p (r_{c',p} - \bar{r_{c'}})^2}} \tag{5}$$

where this time $r_{c,p} = ln(RCA_{c,p}^{PROX} + \varepsilon)$ and $\bar{r_c}$ is the average of $r_{c,p}$ over all products for country c. ε is defined as 0.1 in our calculations and,

$$RCA_{c,p}^{PROX} = \frac{\sum_{p'} RCA_{c,p} \times \phi_{p,p'}}{\sum_{p'} \phi_{p,p'}} \tag{6}$$

$RCA_{c,p}^{PROX}$ is basically a proximity-weighted RCA measure (similarly to the "density" measure developed by Hausmann and Klinger, 2007 and Hi-

Figure A8: Stylized Facts F&K Similarity Index



The left panel of the figure shows the distributions (in year 2000) of the F&K Export Similarity Index for All (not neighbors) Country Pairs, and for Neighbors Pairs only. The right panel shows the average F&K Export Similarity Index for country pairs in each bracket of distance between 250 km. to 5000 km. The upper figures use the F&K Similarity Index for all products, and the lower figures use the F&K Similarity Index for NPRB products only.

Table A5: Correlates of the F&K Similarity Index

	All	All	All	NPRB	NPRB	NPRB
Ln Simple Distance (Km)	-0.3690	-0.1920	-0.1834	-0.2712	-0.1168	-0.1236
	(0.018)***	(0.024)***	(0.026)***	(0.018)***	(0.022)***	(0.024)***
Share a Border		0.6239	0.5076		0.3889	0.2230
		(0.102)***	(0.109)***		(0.078)***	(0.084)***
Same Region		0.3118	0.1523		0.3208	0.0989
		(0.045)***	(0.054)***		(0.035)***	(0.041)**
Same Language			0.2358			0.0940
			(0.054)***			(0.042)**
Have/Had Colonial Relationship			-0.1422			0.1361
			(0.084)*			(0.083)
Common Colonizer			-0.0496			0.0891
			(0.064)			(0.056)
Abs. Dif. Ln GDP Per Capita (PPP)			-0.1612			-0.2816
			(0.035)***			(0.030)***
Abs. Dif. Ln Population			-0.0447			-0.0483
			(0.012)***			(0.010)***
Log Total Bilateral Trade (Imp + Exp)			-0.0324			-0.0246
			(0.003)***			(0.002)***
Abs. Dif. Ln Pysical Capital Per Worker			-0.0077			-0.0203
			(0.032)			(0.027)
Abs. Dif. Ln Human Capital Per Worker			-0.4265			-0.1915
			(0.054)***			(0.049)***
Abs. Dif. Ln Land Per Worker			-0.0450			-0.0282
			(0.033)			(0.029)
N	7503	7503	5460	7503	7503	5460
r2	0.27	0.29	0.40	0.44	0.45	0.56

This table uses a normalized version of the F&K Similarity Index (with mean zero and unit standard deviation) as the dependent variable. Columns 1-3 estimates model (3) using the F&K Export Similarity Index with all products, while columns 4-6 uses the F&K Export Similarity Index computed with NPRB products only. All regressions include country dummies. Standard errors are clustered at the country-pair level.

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

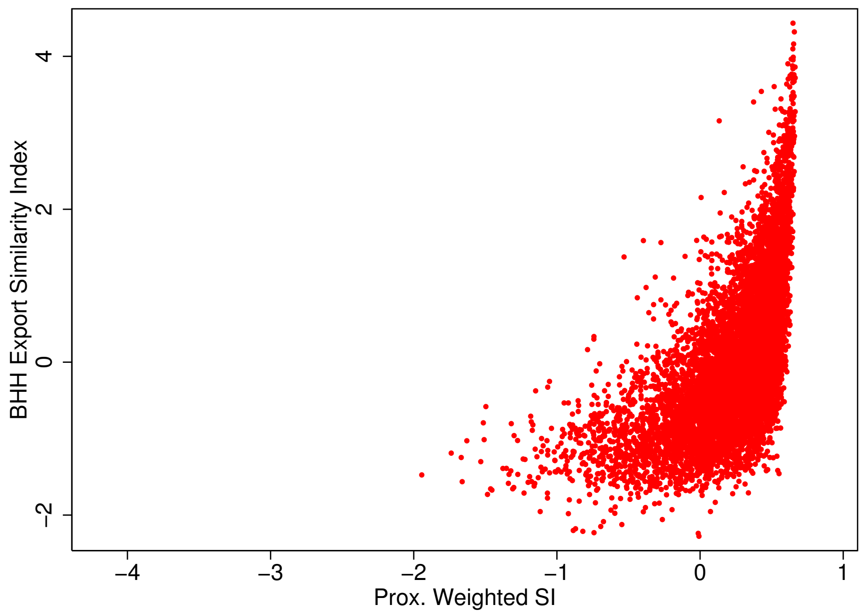
dalgo et. al. 2007). Proximity ($\phi_{p,p'}$), a product-product variable, measures the minimum conditional probability of two products being co-exported by any two countries. Hausmann and Klinger (2007) and Hidalgo et. al. (2007) interpret two products having a high proximity value as requiring similar capabilities or technologies.

Hence, this modified similarity index, $S_{c,c'}^{PROX}$, would measure not only the similarity in the intensity of exports of every product for a pair of countries, but also on whether these two countries are similar in the technological bundle that surrounds every product (as measured by the other products they export). This measure will give a higher weight to two countries having the same product with similar surrounding bundles. At the same time it will punish the similarity among two countries when –even if they are exporting the same product– they do not necessarily have the same technological bundle that surrounds such product.

Overall, there is still correlation between the simple and the proximity-weighted similarity index ($\rho = 0.54$) as seen in Figure A9

Figure A10 replicates Figures 1 and 2 using the proximity weighted similarity index. The upper panel uses all products, while the bottom panel uses NPRB products only. In both we can see how the distribution of the proximity weighted similarity index for neighboring countries is shifted to the right. shows that our analysis presented in the main body of this paper is robust to using the Proximity Weighted Similarity Index. The figure presents the distribution of the index for geographical neighbors and non neighbors, and the declining relationship of the index with distance, for all products and NPRB products. The lower panel (NPRB products only) however, shows

Figure A9: Scatter of Proximity Weighted Similarity Index vs. BHH Similarity Index



The figure is a scatterplot comparing our original (BHH) Export Similarity Index and the Proximity Weighted Export Similarity Index for all country pairs, in year 2000.

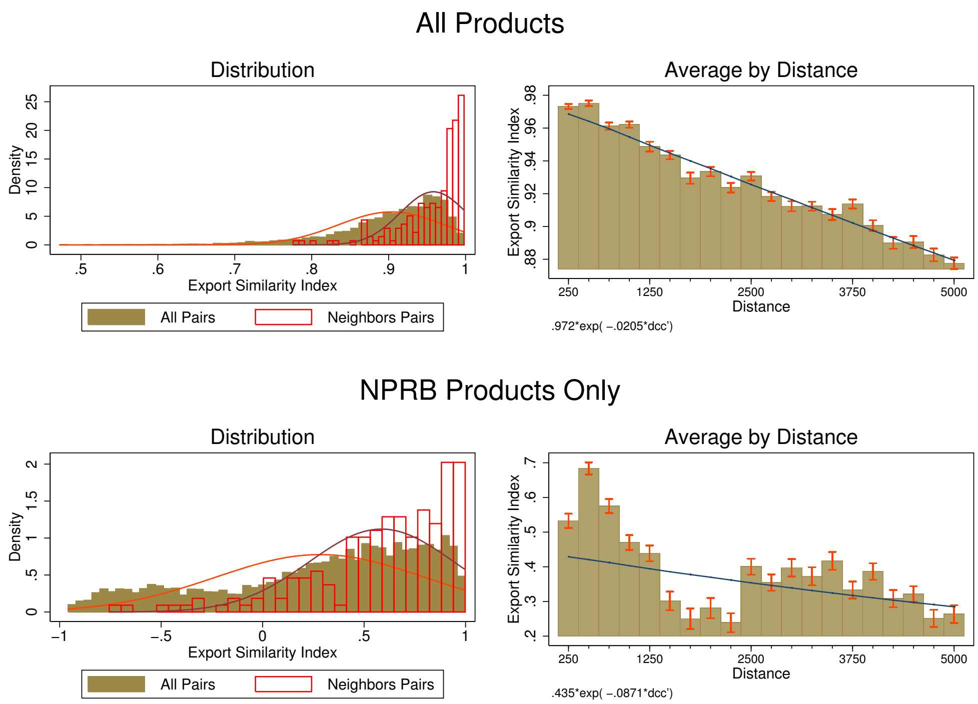
some discontinuity in the declining relation with distance, but it is declining overall.

We turn to study this more in detail by replicating model (3), using on the LHS a normalized version of the Proximity Weighted Similarity Index (with mean zero and unit standard deviation). The results are presented in Table A6. Consistently with the results in the main body of the paper, longer distances are negatively correlated with similarity in exports, while countries sharing a border and in the same region tend to have a larger proximity weighted similarity index, as opposed to non-neighboring countries in different regions.

A.2 The Network of Exports Similarity

There is a puzzling similarity in the export basket of countries that is strongly affected by variables that proxy for distance and that 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. Figure A11 presents the network of export similarity for year 2008 as a graphical network where each node represent 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 the links is proportional to the similarity index and the color of the link indicates whether the

Figure A10: Stylized Facts Proximity Weighted Similarity Index



The left panel of the figure shows the distributions (in year 2000) of the Proximity Weighted Export Similarity Index for All (not neighbors) Country Pairs, and for Neighbors Pairs only. The right panel shows the average Proximity Weighted Export Similarity Index for country pairs in each bracket of distance between 250 km. to 5000 km. The upper figures use the Proximity Weighted Similarity Index for all products, and the lower figures use the Proximity Weighted Similarity Index for NPRB products only.

Table A6: Correlates of the Proximity Weighted Similarity Index

	All	All	All	NPRB	NPRB	NPRB
Ln Simple Distance (Km)	-0.0968	-0.0607	-0.0560	-0.1716	-0.0632	-0.0816
	(0.004)***	(0.006)***	(0.006)***	(0.014)***	(0.016)***	(0.020)***
Share a Border		0.0531	0.0257		0.2404	0.1861
		(0.020)***	(0.022)		(0.068)***	(0.077)**
Same Region		0.0869	0.0114		0.2353	0.1491
		(0.010)***	(0.012)		(0.030)***	(0.041)***
Same Language			0.0154			0.0382
			(0.013)			(0.044)
Have/Had Colonial Relationship			-0.0282			0.1381
			(0.023)			(0.120)
Common Colonizer			0.0075			-0.0580
			(0.016)			(0.042)
Abs. Dif. Ln GDP Per Capita (PPP)			-0.0568			-0.3711
			(0.009)***			(0.028)***
Abs. Dif. Ln Population			-0.0081			0.0103
			(0.003)***			(0.011)
Log Total Bilateral Trade (Imp + Exp)			-0.0100			-0.0174
			(0.001)***			(0.002)***
Abs. Dif. Ln Pysical Capital Per Worker			-0.0109			0.1624
			(0.008)			(0.023)***
Abs. Dif. Ln Human Capital Per Worker			-0.2888			-0.0408
			(0.014)***			(0.044)
Abs. Dif. Ln Land Per Worker			-0.0217			-0.0539
			(0.008)***			(0.031)*
N	7503	7503	5460	7503	7503	5460
r2	0.56	0.56	0.65	0.53	0.53	0.53

This table uses the a normalized version of the Proximity Weighted Similarity Index, with mean zero and unit standard deviation, as the dependent variable. Columns 1-3 estimates the model using the (normalized) Proximity Weighted Similarity Index computed with all products, while columns 4-6 uses the (normalized) Proximity Weighted Similarity Index computed with NPRB products only. All regressions include country dummies. Standard errors are clustered at the country-pair level.

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

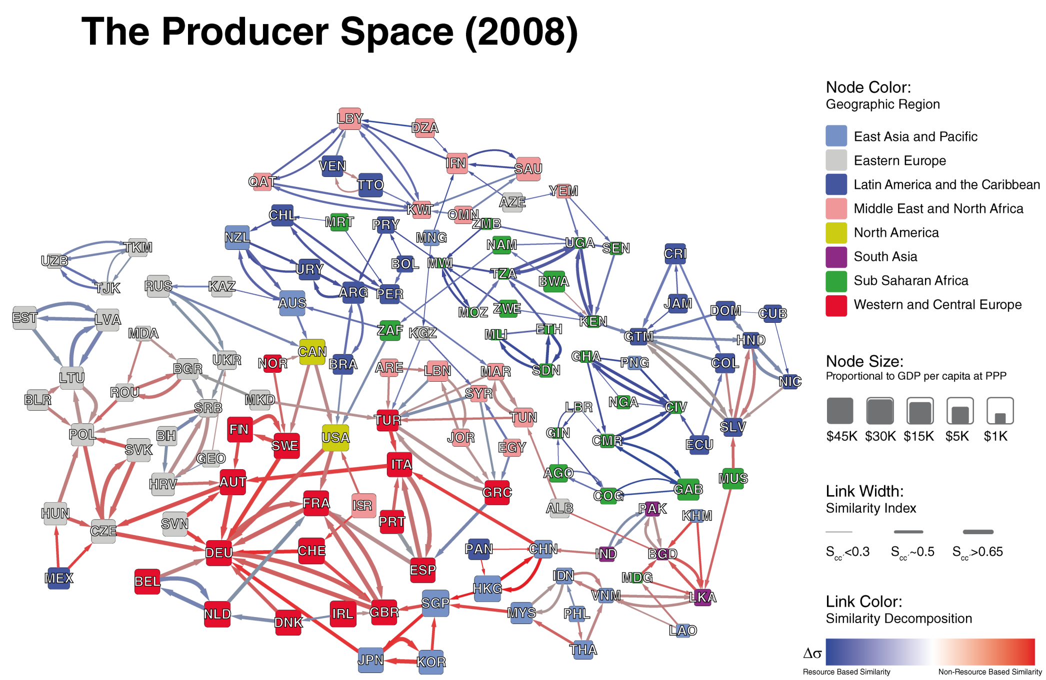
similarity is driven by Primary and Resource Based (PRB) products (blue) or by NPRB products (red) (see section A.4 for more details). We note that, in a large number of cases, the country with the most similar export structure is an immediate neighbor, such as in the case of France, Germany, Austria, 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.3 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 among themselves, as we show in figure 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, and generates a similarity index by computing the Pearson correlation of the RCA vectors, identically to the way we computed the Export Similarity Index $S_{c,c'}$; (2) a Similarity Index on Rest of the World (ROW) Exports , which uses data on exports to the rest of the world excluding bilateral exports for every pair of countries, and similarly, computes the Pearson correlation of the RCA vectors as well.

We use these two measures to show that the variation in the Export Similarity Index ($S_{c,c'}$) is mostly driven by the ROW Export Similarity Index. To support this statement we run a linear regression using the Export Similarity Index as the dependent variable and both decompositions as the

Figure A11: The Network of Exports Similarity (Year 2008)



This figure is a network representation of the Export Similarity matrix in year 2008. In the network each node represents a country. Each country has two outgoing links, which represent the two other countries most similar in terms of their export basket, as measured by our Export Similarity Index $S_{c,c'}$. The color of the nodes represent the geographical region, as defined by the World Bank. The color of the links represent whether NPRB products are driving the similarity (red) or, otherwise, it is being driven by PRB products (blue).

Table A7: Bilateral and ROW Similarity Index, Year 2000

	All	All	Neighbors	Neighbors
Bilateral Exp. Sim. Index	0.5757	0.2758	0.7316	0.4720
	(0.068)***	(0.034)***	(0.063)***	(0.041)***
ROW Exp. Sim. Index		0.8304		0.8462
		(0.007)***		(0.039)***
Constant	0.1617	0.0972	0.3392	0.0923
	(0.002)***	(0.001)***	(0.013)***	(0.014)***
N	7260	7260	179	179
r2	0.07	0.71	0.32	0.87

This table uses the Export Similarity Index (not normalized, for all products) as the dependent variable. Columns 1 and 2 use all the country-pairs in the sample, columns 3 and 4 limits the sample to neighboring country pairs. Standard errors are clustered at the country-pair level.
$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

independent variables for year 2000. The results of such regression are in table A7.

The first two columns of table A7 use the dataset for all the country pairs. In terms of explaining the left-hand side variable, the ROW Similarity Index does a much better job, as can be seen in the difference between the R-squared between columns 2 and 1: an increase of 0.65. Also, in terms of magnitude of the coefficients, in column 2, the ROW Similarity Index coefficient is almost three times larger than the bilateral similarity index one.

Columns 3 and 4 of table A7 repeat the exercise but limiting the dataset only to neighboring countries. In fact, in this case, the Bilateral Similarity Index explains a larger portion of $S_{c,c'}$, hinting that neighboring countries do engage in more intra-industry trade, but still, the ROW Similarity Index explains much more. The R-squared is increased by 0.55 from specification 3 to 4, and the magnitude of the ROW Similarity Index estimator is roughly twice as large as the magnitude of the Bilateral Similarity Index coefficient.

In all cases, the similarity index as measured by exports to the rest of the world has a larger explanatory power in the regression. This hints that most of the similarity among countries, and among neighbors, mostly driven by their exports to the rest of the world, and not the bilateral exports among themselves. Hence, the similarity in exports that we find is not driven by intra-industry trade.

A.4 Decomposing Similarity

The observed similarity through the network in Figure A11 is based on the correlated export of resource-based products for some country-pairs (blue links) and by non-resource-based products for others (red links).

We created a measure to determine whether for a pair of countries' similarity is a reflection of the export of primary and resource based (PRB) products or, on the contrary, non primary nor resource based (NPRB) products. The measure is based on decomposing the relative contribution of PRB and NPRB products to export similarity by separating products into these two categories and counting the fraction of PRB and NPRB products that both countries export with an RCA above their respective means. We take the difference between these two fractions as an estimate of the contribution of PRB and NPRB products to export similarity. Formally, we define:

$$\Delta\sigma_{c,c'} = \sigma_{c,c'}^{NPRB} - \sigma_{c,c'}^{PRB} \tag{7}$$

where

$$\sigma_{c,c'}^{NPRB} = \frac{1}{N_{NPRB}} \sum_{p \epsilon NPRB} \delta_{c,c',p} \tag{8}$$

and N_{NPRB} is the total number of NPRB products and

$$\delta_{c,c',p} = \begin{cases} 1 & \text{if } RCA_{c,p} \geq \overline{RCA_c} \text{ and } RCA_{c',p} \geq \overline{RCA_{c'}} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where $\overline{RCA_c}$ is the average RCA of country c over all products.

The definition for $\sigma_{c,c'}^{PRB}$ can be obtained by changing NPRB for PRB in (8).

From equation (7), $\Delta\sigma_{c,c'} > 0$ if the major contributors to the export similarity between c and c' are NPRB products, such as manufactures and chemicals, and negative in the opposite case. As an example, Figure A12 plots Japan's and Korea's RCA in all products in 2008 and shows NRBP products in red and PRB products in blue. The horizontal flat lines represents the average RCA over all products for Korea, while the vertical flat line does so for Japan. In this case $\sigma_{c,c'c}^{NPRB} = 0.6517$, $\sigma_{c,c'}^{PRB} = 0.3471$ and $\Delta\sigma_{c,c'} = 0.3046$, indicating that Japan and Korea export 61.75% of all of their NPRB products with an RCA above their respective means (in the upper right part of the graph), compared to only 34.71% for PRB products. This shows that the similarity between Japan and Korea that we are measuring comes mainly from their correlated export of NPRB products.

By using these measures we are able to document for any pair of countries whether their exports similarity is driven by NPRB or by PRB products. Not all countries similarity is driven by the same kind of products. Figure (A13) summarizes this information by showing, within each region of the World, what proportion of country-pairs are similar mostly due to NPRB products, or PRB products.

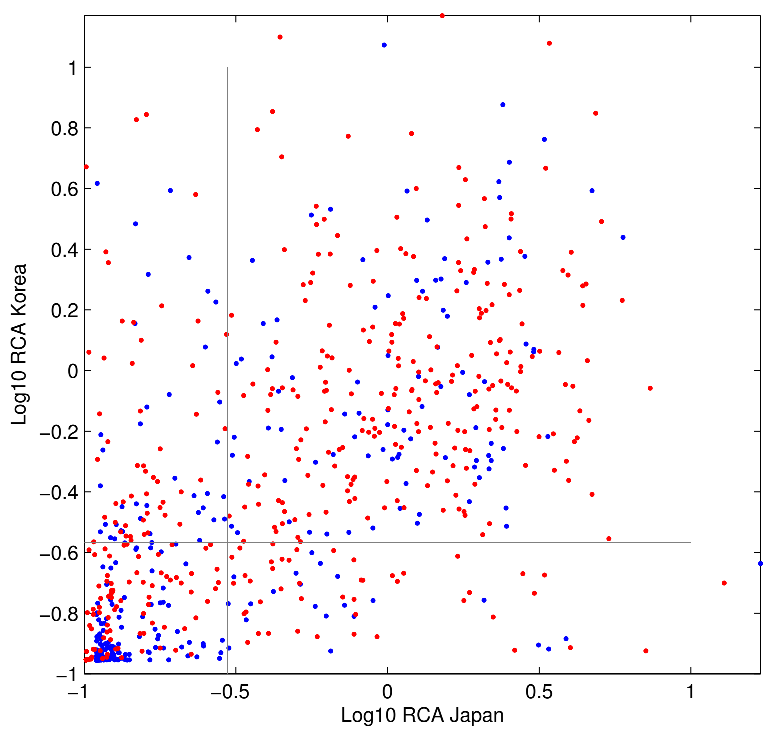
A.5 Robustness Tests: Dynamics of Exports Similarity

The results on the dynamics of export similarity are robust to a number of different specifications of model (4). Our results are robust in two main aspects: the role of geographic neighbors in the likelihood of adding new products to the export basket of a country is always statistically significant; and the coefficient is sharply reduced in magnitude, and it is not statistically different from zero, when using the control sample with a random set of neighbors.

Table A8 is analogous to table 9 in the main body of the paper, but uses growth in export value as the dependent variable, instead of growth in RCA value. Having a neighbor that exports a product is correlated with an increase in the annual growth of exports for that product of 4% to 5%. The results are larger in magnitude (given that in them the export value is nominal) but are qualitatively the same.

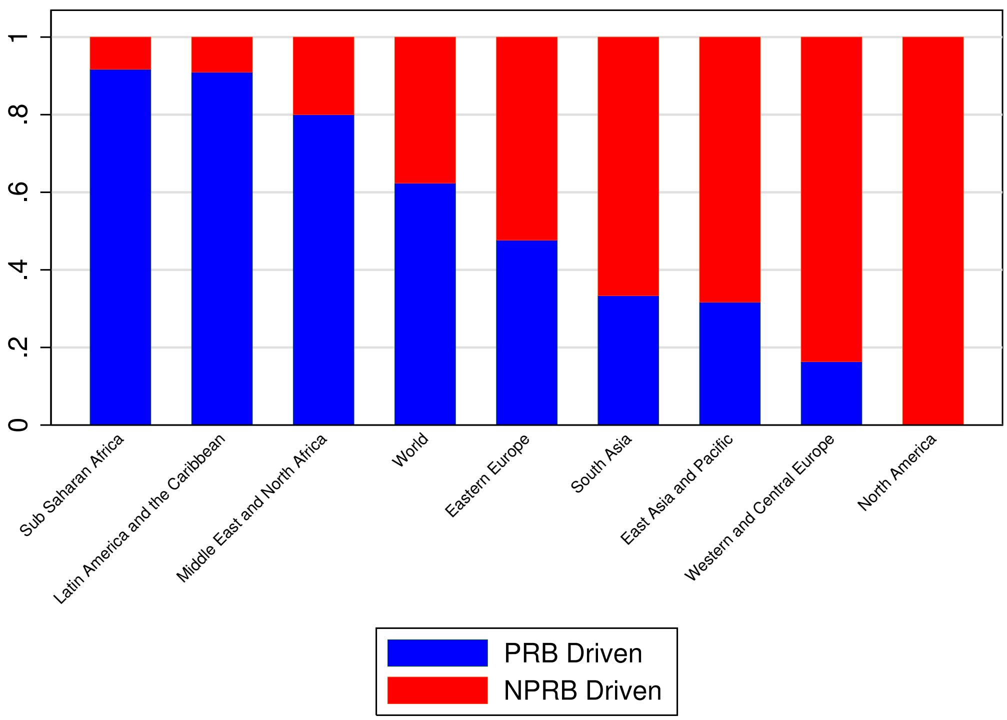
Table A9 replicates the results excluding the period 1980-1990 from the sample, in order to assure our results are robust to the changes in the SITC classification in year 1985.

Figure A12: Decomposition of Similarity Index for Korea and Japan in 2008



The figure shows a scatterplot in where the vertical axis measures the RCA in a product for Korea, and the horizontal axis measures the RCA in a product for Japan. Each dot is a product, and it is red if it is an NPRB product, or red otherwise. The data is from year 2008.

Figure A13: Category of Products Driving Similarities per Region (Year 2008)



This figure shows a bar graph, which represents the share of country pairs, within region, for which their Export Similarity Index is driven by NPRB products (red) or by PRB products (blue).The data is from year 2008.

Our results are also robust to our definition of "jumps": Table A10 presents results limiting the sample to those observations with a baseline RCA equal to zero (as opposed to observations with RCA below 0.1).

Tables A11 and A12 redefine the left-hand side of model (4): here, "jumps" are defined as an increase in the RCA from $RCA_{c,p,t} \leq 0.1$ (at the beginning of the period) to $RCA_{c,p,T} \geq 2$ and to $RCA_{c,p,T} \geq 5$ (by the end of the period), respectively. As expected, given that this definition of "jumps" is stricter, the coefficients for the role of neighbors in the likelihood of adding a new product to the export basket becomes smaller, while still statistically significant.

Our results are also robust to using a logit estimation. Given the computational difficulties of estimating a non-linear model with fixed effects, we pursue this task by limiting our sample to the last period available (2001-2008). Table A13 present the results of this estimation. Also, using non-linear estimation, we are able to considerably improve the (pseudo) R-squared values, and still get consistency in our results.

Finally, we pursue the same analysis using a different dataset, in order to test whether the results are being driven by the way the data is classified. Table A14 uses data from the Harmonized System classification, disaggregated at the 4-digit level. While we do not present results here for NPRB products only, we find that also in this classification we get consistency in our results as compared to the SITC4 dataset: the estimated coefficients are highly similar in their magnitudes and statistical significance, and the coefficients become statistically equal to zero when using the control sample.

Table A8: Dynamics of Exports Similarity (Export Value Growth)

Panel A: Continous Independent Variable				
	Real		Control	
	All	NPRB	All	NPRB
Ln Max RCA Neighbors	2.4806	2.3731	-0.2059	2.5983
	(0.148)***	(0.251)***	(1.186)	(1.686)
Baseline Ln Exports	-3.2959	-3.9931	-3.2105	-3.9462
	(0.061)***	(0.090)***	(0.060)***	(0.089)***
Baseline Density	3.2031	-14.0180	22.8693	3.5617
	(5.993)	(9.677)	(5.626)***	(8.653)
Growth Rate Exports (t-1)	-0.0096	-0.0039	-0.0088	-0.0032
	(0.005)*	(0.007)	(0.006)	(0.007)
Zero Exports (t-1)	-2.5476	0.1431	-3.1926	-0.1543
	(0.434)***	(0.632)	(0.450)***	(0.671)
N	262017	136929	262017	136929
r2	0.37	0.45	0.36	0.45
Panel B: Binary Independent Variable				
	Real		Control	
	All	NPRB	All	NPRB
Neighbor Exports	5.5296	4.2085	-0.1668	0.3416
	(0.374)***	(0.550)***	(0.358)	(0.501)
Baseline Ln Exports	-3.2445	-3.9633	-3.2103	-3.9490
	(0.060)***	(0.089)***	(0.060)***	(0.090)***
Baseline Density	10.2808	-7.4406	22.8850	2.9340
	(5.898)*	(9.444)	(5.609)***	(8.664)
Growth Rate Exports (t-1)	-0.0091	-0.0030	-0.0087	-0.0036
	(0.005)*	(0.007)	(0.006)	(0.007)
Zero Exports (t-1)	-2.8801	0.0000	-3.1950	-0.1378
	(0.429)***	(0.634)	(0.452)***	(0.672)
N	262017	136929	262017	136929
r2	0.37	0.45	0.36	0.45

This table presents results using the Compound Average Annual Growth for Export value 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 have 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 ammount 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 A9: Dynamics of Exports Similarity (Excluding 1980)

Panel A: Continous Independent Variable				
	Real		Control	
	All	NPRB	All	NPRB
Ln Max RCA Neighbors	0.0026	0.0020	-0.0025	-0.0032
	(0.000)***	(0.001)***	(0.002)	(0.004)
Baseline Ln RCA	0.0158	0.0093	0.0177	0.0095
	(0.003)***	(0.005)*	(0.003)***	(0.005)**
Baseline Density	0.2253	0.3869	0.2425	0.3993
	(0.034)***	(0.062)***	(0.033)***	(0.059)***
Growth Rate RCA (t-1)	-0.0005	-0.0006	-0.0005	-0.0006
	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Zero RCA (t-1)	0.0012	0.0026	0.0006	0.0022
	(0.001)	(0.002)	(0.001)	(0.002)
N	123300	62866	123300	62866
r2	0.05	0.06	0.06	0.07
Panel B: Binary Independent Variable				
	Real		Control	
	All	NPRB	All	NPRB
Neighbor Exports	0.0080	0.0066	0.0011	0.0007
	(0.001)***	(0.002)***	(0.001)	(0.001)
Baseline Ln RCA	0.0163	0.0095	0.0177	0.0094
	(0.003)***	(0.005)**	(0.003)***	(0.005)**
Baseline Density	0.2288	0.3827	0.2425	0.3988
	(0.034)***	(0.061)***	(0.033)***	(0.059)***
Growth Rate RCA (t-1)	-0.0005	-0.0006	-0.0005	-0.0006
	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Zero RCA (t-1)	0.0011	0.0026	0.0006	0.0022
	(0.001)	(0.002)	(0.001)	(0.002)
N	123300	62866	123300	62866
r2	0.05	0.06	0.06	0.07

This table presents results when excluding period 1980-1990 from the sample. 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 have 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 ammount 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 A10: Dynamics of Exports Similarity (Baseline RCA=0)

Panel A: Continous Independent Variable				
	Real		Control	
	All	NPRB	All	NPRB
Ln Max RCA Neighbors	0.0040	0.0043	-0.0041	-0.0122
	(0.001)***	(0.001)***	(0.003)	(0.008)
Baseline Ln RCA	0.0000	0.0000	0.0000	0.0000
	(.)	(.)	(.)	(.)
Baseline Density	0.2560	0.4565	0.2816	0.4970
	(0.049)***	(0.121)***	(0.049)***	(0.119)***
Growth Rate RCA (t-1)	-0.0004	-0.0002	-0.0005	-0.0002
	(0.000)***	(0.000)	(0.000)***	(0.000)
Zero RCA (t-1)	0.0020	0.0013	0.0016	0.0010
	(0.001)	(0.002)	(0.001)	(0.002)
N	112783	57289	112783	57289
r2	0.10	0.15	0.11	0.15
Panel B: Binary Independent Variable				
	Real		Control	
	All	NPRB	All	NPRB
Neighbor Exports	0.0138	0.0140	-0.0006	-0.0017
	(0.002)***	(0.003)***	(0.002)	(0.003)
Baseline Ln RCA	0.0000	0.0000	0.0000	0.0000
	(.)	(.)	(.)	(.)
Baseline Density	0.2618	0.4608	0.2809	0.4954
	(0.049)***	(0.120)***	(0.049)***	(0.119)***
Growth Rate RCA (t-1)	-0.0004	-0.0002	-0.0005	-0.0002
	(0.000)***	(0.000)	(0.000)***	(0.000)
Zero RCA (t-1)	0.0018	0.0013	0.0016	0.0010
	(0.001)	(0.002)	(0.001)	(0.002)
N	112783	57289	112783	57289
r2	0.10	0.15	0.11	0.15

This table presents results limiting the observations to those having an initial RCA zero at the beginning of each period. 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 have 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 ammount 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 A11: Dynamics of Exports Similarity ($RCA_{c,p,T} \geq 2$)

Panel A: Continous Independent Variable				
	Real		Control	
	All	NPRB	All	NPRB
Ln Max RCA Neighbors	0.0026	0.0029	-0.0021	-0.0058
	(0.000)***	(0.001)***	(0.002)	(0.005)
Baseline Ln RCA	0.0014	-0.0033	0.0029	-0.0025
	(0.002)	(0.003)	(0.002)	(0.003)
Baseline Density	0.0892	0.1900	0.1099	0.2069
	(0.019)***	(0.042)***	(0.020)***	(0.049)***
Growth Rate RCA (t-1)	-0.0003	-0.0003	-0.0004	-0.0003
	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Zero RCA (t-1)	0.0031	0.0056	0.0026	0.0051
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
N	173433	90811	173433	90811
r2	0.05	0.07	0.05	0.07
Panel B: Binary Independent Variable				
	Real		Control	
	All	NPRB	All	NPRB
Neighbor Exports	0.0073	0.0069	0.0000	-0.0012
	(0.001)***	(0.002)***	(0.001)	(0.002)
Baseline Ln RCA	0.0020	-0.0028	0.0029	-0.0024
	(0.002)	(0.003)	(0.002)	(0.003)
Baseline Density	0.0955	0.1972	0.1095	0.2058
	(0.019)***	(0.042)***	(0.020)***	(0.049)***
Growth Rate RCA (t-1)	-0.0003	-0.0003	-0.0004	-0.0003
	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Zero RCA (t-1)	0.0030	0.0056	0.0026	0.0051
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
N	173433	90811	173433	90811
r2	0.05	0.07	0.05	0.07

This table presents results redefining the left-hand side variable to be 1 if $RCA_{c,p,t} \leq 0.1$ and $RCA_{c,p,T} \geq 2$ (instead of $RCA_{c,p,T} \geq 1$). 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 have 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 ammount 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 A12: Dynamics of Exports Similarity ($RCA_{c,p,T} \geq 5$)

Panel A: Continous Independent Variable				
	Real		Control	
	All	NPRB	All	NPRB
Ln Max RCA Neighbors	0.0013	0.0010	-0.0010	-0.0023
	(0.000)***	(0.000)**	(0.001)	(0.003)
Baseline Ln RCA	0.0014	0.0000	0.0021	0.0006
	(0.001)	(0.002)	(0.001)*	(0.002)
Baseline Density	0.0430	0.0974	0.0588	0.1105
	(0.009)***	(0.020)***	(0.010)***	(0.022)***
Growth Rate RCA (t-1)	-0.0001	-0.0001	-0.0002	-0.0001
	(0.000)***	(0.000)*	(0.000)***	(0.000)**
Zero RCA (t-1)	0.0007	0.0009	0.0004	0.0008
	(0.000)*	(0.001)	(0.000)	(0.001)
N	173433	90811	173433	90811
r2	0.04	0.05	0.04	0.05
Panel B: Binary Independent Variable				
	Real		Control	
	All	NPRB	All	NPRB
Neighbor Exports	0.0035	0.0021	-0.0006	-0.0012
	(0.001)***	(0.001)**	(0.001)	(0.001)
Baseline Ln RCA	0.0017	0.0002	0.0021	0.0006
	(0.001)	(0.002)	(0.001)*	(0.002)
Baseline Density	0.0468	0.1004	0.0586	0.1102
	(0.009)***	(0.020)***	(0.010)***	(0.022)***
Growth Rate RCA (t-1)	-0.0001	-0.0001	-0.0002	-0.0001
	(0.000)***	(0.000)*	(0.000)***	(0.000)**
Zero RCA (t-1)	0.0006	0.0009	0.0004	0.0008
	(0.000)	(0.001)	(0.000)	(0.001)
N	173433	90811	173433	90811
r2	0.04	0.05	0.04	0.05

This table presents results redefining the left-hand side variable to be 1 if $RCA_{c,p,t} \leq 0.1$ and $RCA_{c,p,T} \geq 5$ (instead of $RCA_{c,p,T} \geq 1$). 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 have 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 ammount 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 A13: Dynamics of Exports Similarity (Logit)

Panel A: Continous Independent Variable				
	Real		Control	
	All	NPRB	All	NPRB
Ln Max RCA Neighbors	0.1279	0.1425	-0.0740	-0.6157
	(0.034)***	(0.070)**	(2.885)	(4.262)
Baseline Ln RCA	1.0458	0.3783	1.0935	0.3460
	(0.234)***	(0.420)	(0.236)***	(0.385)
Baseline Density	11.9902	12.0677	14.1376	13.1071
	(4.537)***	(7.918)	(4.690)***	(8.794)
Growth Rate RCA (t-1)	-0.0379	-0.0448	-0.0380	-0.0445
	(0.007)***	(0.010)***	(0.008)***	(0.011)***
Zero RCA (t-1)	0.3621	0.5512	0.3118	0.4870
	(0.236)	(0.420)	(0.226)	(0.362)
N	22792	10551	22792	10551
r2_p	0.19	0.26	0.19	0.26
Panel B: Binary Independent Variable				
	Real		Control	
	All	NPRB	All	NPRB
Neighbor Exports	0.3081	0.2148	-0.0523	-0.1043
	(0.117)***	(0.200)	(0.269)	(0.435)
Baseline Ln RCA	1.0590	0.4017	1.0932	0.3311
	(0.232)***	(0.420)	(0.235)***	(0.380)
Baseline Density	12.7418	13.0036	14.1460	13.1778
	(4.547)***	(8.006)	(4.759)***	(8.566)
Growth Rate RCA (t-1)	-0.0382	-0.0440	-0.0380	-0.0447
	(0.007)***	(0.010)***	(0.007)***	(0.010)***
Zero RCA (t-1)	0.3534	0.5445	0.3117	0.4834
	(0.236)	(0.418)	(0.221)	(0.369)
N	22792	10551	22792	10551
r2_p	0.19	0.26	0.19	0.26

This table presents results using a logit estimation, limiting the sample to the last period in our dataset (2001-2008). 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 have 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 ammount 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 A14: Dynamics of Exports Similarity (HS4)

	Real	Control	Real	Control
Ln Max RCA Neighbors	0.0034	0.0054		
	(0.001)***	(0.005)		
Neighbor Exports			0.0093	-0.0002
			(0.002)***	(0.002)
Baseline Ln RCA	0.0254	0.0272	0.0262	0.0272
	(0.005)***	(0.005)***	(0.005)***	(0.005)***
Baseline Density	0.2752	0.3201	0.2750	0.3207
	(0.050)***	(0.051)***	(0.050)***	(0.051)***
Growth Rate RCA (t-1)	-0.0009	-0.0009	-0.0009	-0.0009
	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Zero RCA (t-1)	-0.0001	-0.0002	-0.0003	-0.0003
	(0.002)	(0.002)	(0.002)	(0.002)
N	45589	45589	45589	45589
r2	0.06	0.06	0.06	0.06

This table presents results using the Harmonized System classification disaggregated at the 4-digit level. The first two columns use the maximum RCA among all geographic neighbors of a country for a particular product, in natural logarithm, as the independent variable. Columns 3 and 4 uses a dummy variable which takes the value 1 if at least one of the neighbors of a country have 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 ammount 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.

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