Essays on the Transmission and Diffusion of Productive Knowledge in International Economics

Essays on the Transmission and Diffusion of Productive Knowledge in International Economics

A dissertation presented

by

Dany Bahar

to

The Committee on Higher Degrees in Public Policy

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Public Policy

Harvard University

Cambridge, Massachusetts

March 2014

Dissertation Advisors:
Elhanan Helpman (Chair)
Ricardo Hausmann
Pol Antras
Laura Alfaro

Author:
Dany Bahar

Essays on the Transmission and Diffusion of Productive Knowledge in International Economics

Abstract

Numerous empirical studies have shown the difficulties associated with the transmission of knowledge and the limitations of its diffusion process. What are the implications of these difficulties and limitations to international economics? This dissertation deals with this question by looking at how productive knowledge plays a role in the evolution of the comparative advantage of nations and the international expansion of multinational corporations. The first chapter finds that a country is 65% more likely to start exporting a good that is being exported by any of its geographic neighbors, consistently with evidence on the limited geographic patterns of knowledge diffusion. The second chapter finds that migrants, serving as carriers of productive knowledge, play a role in explaining the appearances of new export industries in both their sending and receiving countries. In particular, in terms of their ability to induce exports in the average country, an increase of only 65,000 people in the stock of migrants is associated with about 15% increase in the likelihood of adding a new product to a country's export basket. The figure becomes 15,000 for skilled migrants. The third chapter looks at how the barriers to knowledge transmission within the firm limit the horizontal expansion of multinational corporations. The findings suggest that multinational corporations are, on average, about 12% less likely to horizontally expand a sector that is one standard deviation above the mean in the knowledge intensity scale.

Contents

List of Tables

List of Figures

Acknowledgments

I am very grateful to my extraordinary and very dedicated academic advisors: Laura Alfaro, Pol Antras, Ricardo Hausmann and Elhanan Helpman for their guidance, advise and support. There are no words to describe how grateful I am for their counsel and friendship.

I also thank other Harvard faculty members for their thoughtful comments and willingness to discuss my research ideas since I reached out to them for the first time: Bill Kerr, Robert Lawrence, Marc Melitz, Nathan Nunn, Lant Pritchett, Dani Rodrik and Richard Zeckhauser. I also thank my co-authors César Hidalgo and Hillel Rapoport.

I am indebted to colleagues and friends who have been helping me in thinking through the topics in this dissertation throughout the years: Martin Abel, Sam Asher, Sebastian Bustos, Michele Coscia, Juan Ariel Jimenez, Michael Kransdorff, Frank Nefke, Paul Novosad, Ran Shorrer, Rodrigo Wagner, Muhammed Yildirim, Andres Zahler and Oren Ziv. I am also grateful to participants of numerous seminars at the Harvard Economics Department and Harvard Kennedy School, among many other places.

I am grateful to the staff of Harvard's Center for International Development for providing me an academic home and supporting my research on every front. I acknowledge financial support from Harvard Center for International Development and Harvard Economics Department.

Last but not least, I thank my family for standing behind me. Very special thanks to Jessica Brandt for her unconditional support throughout this process.

Dedicated to the memory of my grandparents, Nico and Erna Colonomos (Z"L). They did not live to see this day; yet, they are my source of inspiration and the reason behind this and all past and future accomplishments. I remain eternally grateful.

§

"And whoever saves a life, it is considered as if he saved an entire world."
Jerusalem Talmud, Sanhedrin 4:1 (22a)

Introduction

About half of cross-country income variation can be explained by differences in productivity levels (Caselli 2005, Hall and Jones 1999). Such an important stylized fact begs the question: what is productivity? Moses Abramovitz (1956) described productivity as “some sort of measure of our ignorance.” Being productive can also be defined as knowing how to do more and better with the same resources. Why, then, is the knowledge that is available in some places is not available in others? For instance, why do some farmers know how to deal with adverse climate conditions, while other farmers in a similar region do not? Why do some countries know how to set up the proper institutional frameworks for economic growth and others just can’t? Why is it that firms know how to manufacture a product much more efficiently than other firms do, even within the same country? These are just few examples of very important but yet unanswered questions economist have studied for decades.

A general answer to all these question is that, certainly, knowledge is not fully mobile. While sometimes knowledge can be easily codifiable (i.e. in an instruction manual, or a textbook), most often it cannot. This hard-to-codify knowledge can sometimes be embedded in goods: one who owns a calculator does not need to know how to add or subtract. Yet, there is still a lot of knowledge that cannot be written down nor embedded in goods (i.e., how to recognize a face, how to ride a bicycle or how to be a good soccer player). This type of knowledge is what Polanyi (1966) referred to as tacit.

A lot of what makes us productive is, in fact, the tacit knowledge we have accumulated. We gain this tacit knowledge through experience and learning and not in school or by reading. That is precisely what makes its transferability and diffusion so difficult. As Kenneth Arrow (1969) pointed out, the channels for knowledge transmission are limited to human interaction rather than written words.

How do the limitations on the transmission and diffusion of knowledge affect economic processes? This dissertation documents the ways in which these barriers play a role in international economics, touching upon topics such as the evolution of comparative advantage of nations and the international expansion of multinational corporations.

The first chapter of the dissertation asks what are the implications of the documented geographic local character of knowledge diffusion on the comparative advantage of nations? My co-authors and I document that the probability a product is added to a country's export basket is, on average, 65% larger if a neighboring country is a successful exporter of that same product. For existing products, growth of exports in a country is 1.5 percent higher per annum if it has a neighbor with comparative advantage in these products. The main contribution of this study is that, using exports as a measure of knowledge acquisition, we document patterns in the evolution of the comparative advantage of nations that are consistent with the widely documented localized character of knowledge diffusion.

The second chapter asks, to what extent are migrants a source of evolution of the comparative advantage of both their sending and receiving countries? The main finding is that migration is a strong and robust driver of productive knowledge diffusion. In terms of their ability to induce exports, we find that an increase of only 65,000 people in the stock of migrants for the average country, is associated with about 15% increase in the likelihood of adding a new product to a country's export basket. We also find that, in terms of expanding the export basket of countries, a migrant is worth about US $30,000 of foreign direct investment. For skilled migrants these same figures become 15,000 people and US $160,000. The main contribution of this chapter is that it presents robust evidence that migrants, as carriers of tacit knowledge, can shape the comparative advantage of nations by inducing exports from their receiving and sending countries to the rest of the world.

Finally, the third chapter of this dissertation asks, to what extent do barriers to knowledge transmission influence a firm's decision to expand? Using a worldwide dataset on foreign subsidiaries, I show that multinational corporations are, on average, about 12% less likely to horizontally expand a sector that is one standard deviation above the mean in the knowledge intensity scale. In addition, I present evidence showing that when firms do expand their knowledge-intensive activities they tend to do so at shorter geographic distances. Finally, I also find that locating a foreign subsidiary in the same time zone as its headquarters tends to reduce barriers to knowledge transmission by easing communication and effectively reducing the distance between them by, on average, 3500 Km. The chapter also includes a conceptual framework that builds on Helpman, Melitz and Yeaple (2004) to formalize the empirical findings. The contribution of this chapter is that it rationalizes the ways in which the cost of knowledge transmission for firms engaged in foreign direct investment affects the mechanisms of the proximity-concentration hypothesis.

Chapter 1

Neighbors and the Evolution of the Comparative Advantage of Nations: Evidence of International Knowledge Diffusion?[1]

1.1 Introduction

Knowledge has become central to modern theories of growth. Knowledge is embodied in goods that are then shipped around at a cost. When these goods are imported, they accelerate productivity growth in the recipient country (e.g. Rivera-Batiz and Romer, 1990; Coe and Helpman, 1993; Coe et. al., 2009). However, significant parts of knowledge are disembodied or tacit (Polanyi, 1962) and its diffusion requires more direct forms of human interaction, which inevitably limits its scope to more localized or idiosyncratic settings (Arrow, 1969).

Previous research has documented the rapid decay of knowledge diffusion with geo-

[1]Co-authored with Ricardo Hausmann and Cesar Hidalgo, published in the Journal of International Economics (Volume 92, Issue 1, January 2014, Pages 111–123)

graphic distance. This literature looked at the impact of distance on the patterns of patent citation (e.g. Jaffe et. al., 1993), of R&D and patent output (e.g. Branstetter, 2001; Bottazzi and Peri, 2003), of R&D and productivity (Keller, 2002), and on the sales of subsidiaries of multinational corporations (Keller and Yeaple, 2013). Keller (2002, 2004) has shown that foreign sources of technology account for up to 90% of domestic productivity growth and that the impact is highly localized.

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

In this paper, we use a novel setting to explore the diffusion of industry-specific productivity increases: the export baskets of countries. The key assumption is that, controlling for product-specific shifts in global demand, firms in a country will be able to incorporate a new good into their export basket only after they have become productive enough to compete in global markets. Additionally, in order to increase their market share, firms will also need to become more productive. If knowledge diffusion decays strongly with distance, countries with the relevant knowledge should induce shifts in productivity in their neighbors—we explore this in both a static and a dynamic setting. We study both the intensive and the extensive 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 therefore also explore the intensive margin, looking at the impact of neighbors in the evolution of a country's market share.

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

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

We can understand our results in the context of an endogenous Ricardian framework, where comparative advantage evolves with the progressive acquisition of knowledge or technologies which diffuse geographically[2]. However, under such a Ricardian framework, a

[2]Alvarez et. al. (2012) provides a useful framework to think about this. In their model, technology diffuses

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 or instruction manuals—should diffuse with more difficulty. How does tacit knowledge diffuse? As mentioned above, Kenneth Arrow argued that knowledge diffusion requires more direct forms of human interaction, which limits its scope to more localized or idiosyncratic settings (Arrow, 1969). Furthermore, the emerging consensus in the literature of knowledge diffusion is that diffusion occurs predominantly within a fairly short range (e.g. Jaffe et al. 1993; Branstetter, 2001; Keller, 2002; Bottazzi & Peri, 2003), an observation that is attributed to the characteristics of tacit knowledge. Hence, if indeed knowledge diffusion translates into productivity shifts that can shape the export basket of countries, then, in a world in which knowledge diffuses preferentially at short ranges, a country's export basket—as well as its evolution—will be shaped by the knowledge available in its neighborhood.

The localized nature of knowledge diffusion should generate the observables that we document in this paper. In particular, if knowledge has been homogenized preferentially at shorter distances, a snapshot view of the export basket of countries (a realization of their comparative advantage) should resemble that of their neighbors. Dynamically, we should also observe a geographically correlated pattern of adoption of new export goods and of changes in market shares. In this interpretation, there is a causal link between the presence of productive knowledge in a country and its diffusion to a neighbor. However, there is always the possibility that these correlated events may be caused by a third factor that is common to neighboring countries and that explains both the static similarity and the

---

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.

time-lapsed pattern of adoption without there being a causal link between the two. We will try to control, as best we can, for these alternative channels but we do not claim to have ruled them out completely. We discuss this more in detail in the body of the paper.

Until now, the burgeoning literature on international knowledge diffusion has relied on three main indicators to measure knowledge acquisition: patent citations (e.g. Jaffe et al. 1993), patent output (e.g. Bottazzi & 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.

1.2 Data and Stylized Facts

1.2.1 Data

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

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

In the static analysis, for which we use a cross-country data of the year 2000,[3] the base sample consists of 123 countries (7503 country pairs)[4]. For the dynamic analysis, the list of countries is reduced to 100, given the exclusion of countries with no geographic neighbors from the sample and those that belonged to the Former Soviet Union (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.

1.2.2 Exploring Static Similarity

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

[3] 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.

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

$$RCA_{c,p} \equiv \frac{exp_{c,p} / \sum_p exp_{c,p}}{\sum_c exp_{c,p} / \sum_c \sum_p exp_{c,p}} \tag{1.1}$$

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:

$$S_{c,c'} \equiv \frac{\sum_p (r_{c,p} - \bar{r_c})(r_{c',p} - \bar{r_{c'}})}{\sqrt{\sum_p (r_{c,p} - \bar{r_c})^2 \sum_p (r_{c',p} - \bar{r_{c'}})^2}} \tag{1.2}$$

where $r_{c,p} = ln(RCA_{c,p} + \varepsilon)$ and $\bar{r_c}$ is the average of $r_{c,p}$ over all products for country c. We chose a log form to prevent the correlation from being driven by the few products that countries export with very high RCA and we add ε, defined as 0.1, to assign a value to the zeroes, while also preventing the correlation being driven by similarities in the RCA of products that countries export very little of or not at all[5].

$S_{c,c'}$ 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 markets. Nevertheless, our analysis is robust to using the F&K similarity index and other variations of our own similarity index (see Section A.3.3 of the web appendix).

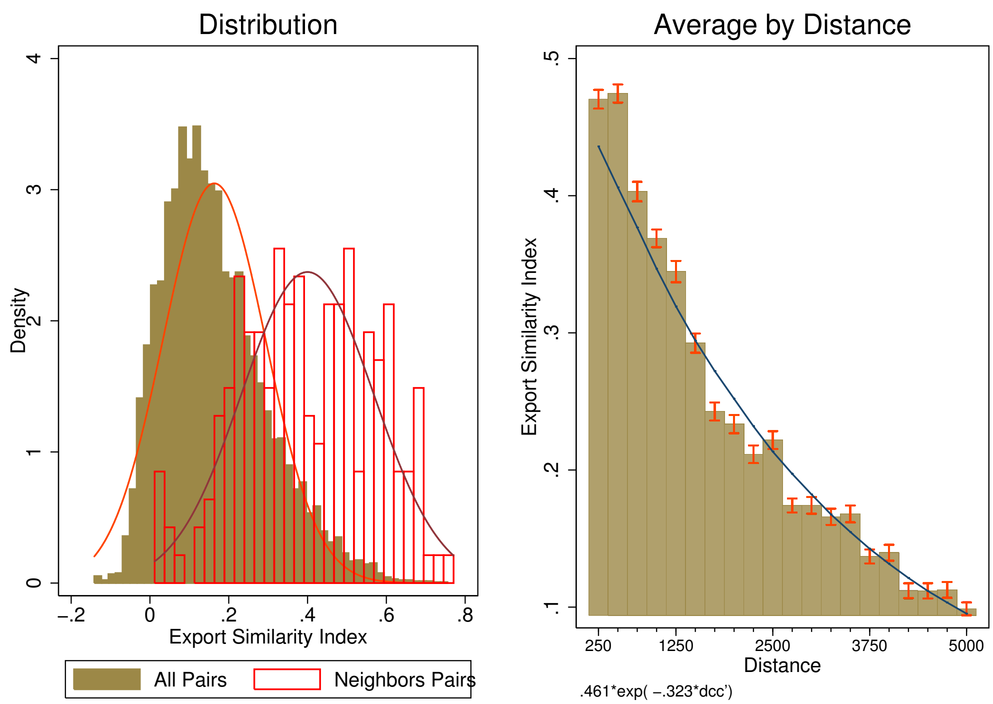
Table 1.1 presents summary statistics for our base sample which contains bilateral country-level data for the year 2000. Note that data on factor endowments is limited to fewer countries.

[5]We test that our results are not driven by the choice of ε. See section A.3.1 of the web appendix for robustness checks of this exercise using different values of ε for constructing $S_{c,c'}$. Also, section A.3.2 of the web appendix present robust results with an alternative $S_{c,c'}$ that does not require a log-transformation.

Table 1.1: *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

Figure 1.1: *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.

The left panel of Figure 1.1 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 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[6]. 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

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

Table 1.2: *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

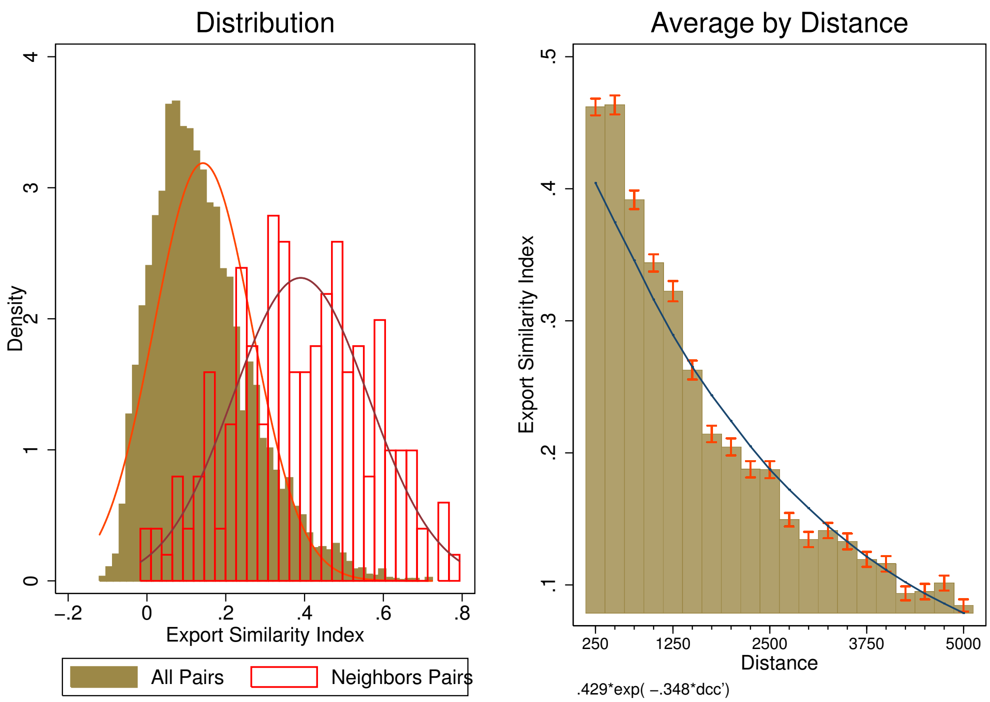
is more likely to be the case for geographic neighbors. To control for this fact, we exclude products from the sample that are confined by geography. We do this by using the technological classification suggested by Lall (2000) that divides products into the categories presented in Table 1.2.

Lall's classification is used to create two categories of products: Primary and Resource Based (PRB) products and Non-Primary or Non-Resource Based (NPRB) products. We consider PRB products as those that are classified as Gold, Primary Products and Resource Based Manufactures (categories 1 thru 4 in Table 1.2), whereas NPRB products are those contained in all other categories.

Figure 1.2 reproduces Figure 1.1 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[7], and the negative relationship between export similarity and geographical distance is equally strong, suggesting that the observed export similarity among neighbors is not driven solely by primary and resource based products. We include more controls in this analysis next.

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

Figure 1.2: *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.

1.2.3 The Correlates of Export Similarity

The fact that, beyond geology and climate, export similarity decays with distance could be due to a number of different reasons. We study the correlates of the Export Similarity Index through an adapted "gravity model" (Zipf, 1946; Tinbergen, 1963). We do so in order to understand whether the role of geographic proximity is actually driven by similarity in other dimensions such as income, size, factor endowments, institutions and culture, among others. Our adapted gravity model follows the functional form:

$$S_{c,c'} = \alpha + \beta \times d_{c,c'} + z_{c,c'}\gamma + l_{c,c'}\theta + b_{c,c'}\delta + \mu_c + \mu_{c'} + \varepsilon_{c,c'} \tag{1.3}$$

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 include 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 1.3. For easier interpretative purposes, we use a normalized version of $S_{c,c'}$ as the dependent variable, with mean zero and unit standard deviation.

The first three columns of table 1.3 correspond to the results with the (normalized) Export Similarity Index computed with all products, while the last three columns uses a version of the Export Similarity Index computed with NPRB products only. The base dataset

Table 1.3: *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 (1.3) 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.
$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

contains 123 countries, which sum up to 7503 unique country pairs in year 2000[8]. 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.3 of the web appendix. Columns 2 and 5 include two variables that represent alternative measures of geographic proximity and are highly correlated with distance: sharing a border and being in the same region. Sharing a border is associated with an export similarity index that is, on average, 0.8 standard deviations above the mean for all products and 0.9 for NPRB goods. We can add to this another 0.4 or 0.35 standard deviations respectively if the two countries are in the same geographical region. This means that we could expect neighboring countries in the same region to have, on average, a similarity index roughly 1.2 standard deviations above the mean relative to non-neighbors from different regions for all goods and 1.25 for NPRB products—this does not take into account the fact that neighbors are a shorter distance apart than the average pair of countries. These variables are always strongly significant in all our robustness checks. This motivates our use of neighboring countries in our dynamic analysis in the next section.

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

8 $\frac{123\times 122}{2}$.

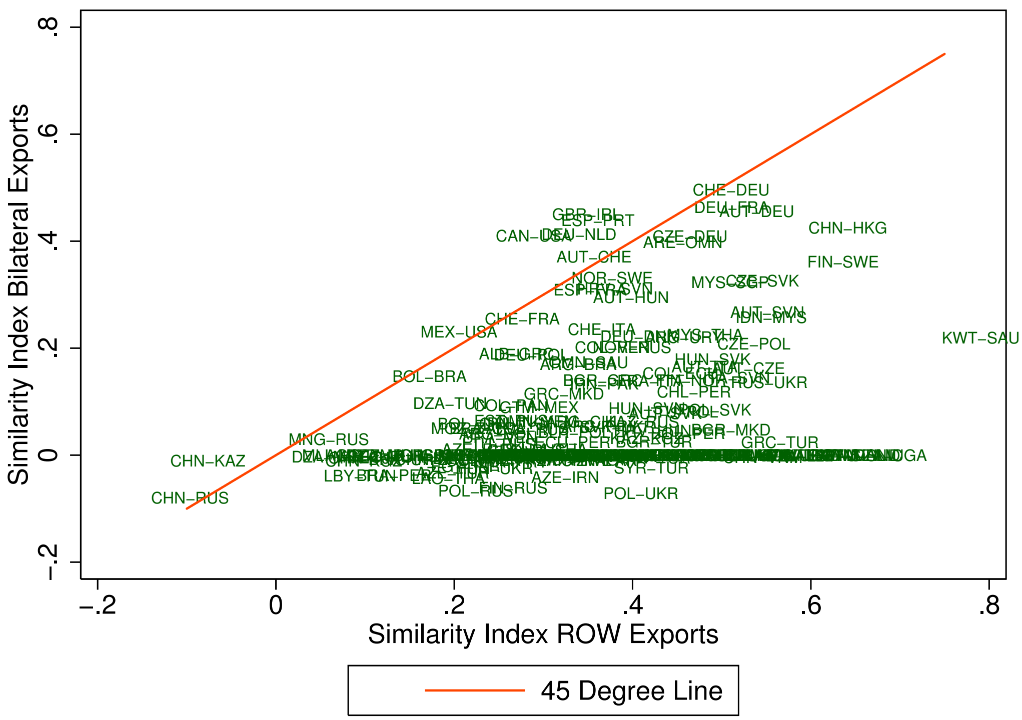
similarity in exports: countries that trade more among themselves are less similar in their export baskets, as would be expected.

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 1.3, 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). Moreover, in a world with integrated supply chains, the similarity in exports could be a result of neighboring countries trading inputs that are classified in the data in the same category as the outputs themselves. Since neighbors trade more intensively, then similarity in bilateral trade may be driving our results. We check for this by comparing the similarity index $S_{c,c'}$ of the bilateral exports of neighbors with the similarity index of their exports to the rest of the world. To do this, we construct a similarity index for each pair of countries based on bilateral exports between each pair of countries, and a similarity index based on each pair's exports compared to the rest of the world (excluding the bilateral exports). Figure 1.3 plots the two measures using data from 2000 and neighboring country pairs only. As Figure 1.3 shows, neighbors are remarkably more similar in terms of what they export to the rest of

Figure 1.3: *Neighbor Similarity (on bilateral exports vis-á-vis ROW exports)*



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

the world than what they trade between themselves. This implies that export similarity is not driven by the composition of bilateral trade between neighbors.

An alternative exercise to explore this point consists of repeating the estimation of model (1.3), using the similarity index of their exports to the rest of the world as the dependent variable. The result is presented in Table 1.4. The relationship with distance—taken to mean sharing a border and being in the same region—holds when considering only exports to the rest of the world as the basis for similarity between all pairs of countries. Moreover, the last column on this table includes the bilateral similarity index as a regressor, although its inclusion does not qualitatively change the results. Section A.2 in the appendix presents further analysis.

Table 1.4: *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 (1.3) with the Export Similarity Index computed using all products.

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

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

1.3 Dynamics of Export Similarity

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

To explore this issue, we use a dynamic analysis to study the role neighbors play in the ability of countries to add a particular good to their export basket or to expand their comparative advantage in a product. We start by discussing the extensive margin. More specifically, we study the probability that a country will add a product to its export basket in period T (i.e. "jump" to the product) if it has at least one neighbor that is already exporting that product in period t (with $T > t$). For this task, we use the dataset described in Section 1.2.1, with 100 countries[9]. We divide our sample into four periods: 1970-1980, 1980-1990, 1990-2000 and 2001-2008[10]. For each period, we eliminate all products that were not exported by any country and all countries that did not export any product. The total number of countries in the dataset is 100, and the total number of products is 777.

We define a "jump" as a tenfold or more increase in the RCA of country c in product p, from $RCA_{c,p} \leq 0.1$ to $RCA_{c,p} \geq 1$ within a ten year period[11]. This setting allows us

[9]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.

[10]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.

[11]With the exception of our last period which is seven years long (2001-2008)

to explore the extensive margin of exports. We are interested in studying the probability of a product being exported in the next period, given that it was not being exported (or exported only in very small quantities) at the beginning of the current period. Furthermore, we are interested in products that achieve an RCA above 1, implying significant gains in comparative advantage and increases in its share of world trade[12].

To avoid noise, we restrict jumps to two conditions: first, a jump needs to keep RCA above 1 for four years after the end of the period, year T (the forward condition); second, we restrict jumps to products that had an RCA below 0.1 for two years before the beginning of the period (the backward condition)[13]. These two conditions help rule out the possibility of "temporary jumps" in the data driven by noise, errors, shocks in commodity prices or other exogenous reasons[14].

Table 1.5 presents the ten NPRB and PRB products with the largest frequency of "jumps" in our dataset. For instance, the NPRB product with the largest number of appearances in the data (i.e. with RCA going from less than 0.1 to above 1 in ten years) is SITC 8441 (men's undershirts), in the period 1980-1990. SITC 8441 had 6 occurrences (denoted by O) out of the 74 countries that had an RCA<0.1 in 1980 (denoted by B) . This means that 8% of the eligible countries acquired RCA>1 for SITC 8441 over that period (denoted by P). Seven out of the top ten products for the NPRB categories are garments and textiles in the period 1980-1990.

Table 1.6 presents the ten countries with the largest number of product appearances in our dataset, classified by NPRB and PRB products. When looking at the ranking based on NPRB products, all countries in the list are developing countries, besides Germany, mostly located in Southeast Asia. China, at the top of the list, added 17 NPRB products to its export basket in the period 1980-1990, or 7% of the 234 products that, at the time,

[12]In section A.5 of the web appendix 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 $RCA_{c,p} \geq 5$.

[13]In section A.5 of the web appendix we present robustness checks that limit the sample to observations for which RCA is equal to zero at the beginning of the period.

[14]For the last period (2001-2008) we eliminate the forward condition due to data limitations.

Table 1.5: *Frequency of Jumps by Product*

SITC4	Product Name	Period	O	B	P
NPRB Products					
8441	Men's undershirt	1980-1990	6	74	0.08
8439	Other women outerwear	1980-1990	5	70	0.07
6781	Iron pipes	2001-2008	5	60	0.08
5913	Herbicides	1980-1990	4	83	0.05
5721	Prepared explosives	2001-2008	4	43	0.09
6521	Unbleached cotton woven fabrics	1980-1990	4	40	0.10
8459	Other knitted outerwear	1980-1990	4	64	0.06
8423	Men's trousers	1980-1990	4	65	0.06
8452	Knitted women's suits & dresses	1980-1990	4	72	0.06
8442	Men's underwear	1980-1990	4	83	0.05
PRB Products					
812	Bran, sharps & other cereal residues	1990-2000	5	58	0.09
342	Frozen fish, excluding fillets	1980-1990	4	72	0.06
3415	Coal & water gases	2001-2008	4	73	0.05
3344		1980-1990	4	33	0.12
611	Raw sugar beet & cane	1980-1990	4	66	0.06
9710	Gold, non-monetary	2001-2008	4	40	0.10
723	Cocoa butter & paste	1970-1980	3	72	0.04
344	Frozen fish fillets	1990-2000	3	49	0.06
3510	Electric current	2001-2008	3	61	0.05
6861	Unwrought zinc & alloys	2001-2008	3	70	0.04

This table presents statistics on the SITC4 products with the largest amount of "jumps" in the data.

Table 1.6: *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
Malaysia	1980-1990	11	307	0.04
Syrian Arab Republic	2001-2008	11	315	0.03
Bangladesh	1980-1990	11	382	0.03
Vietnam	1990-2000	11	364	0.03
Tanzania	2001-2008	10	358	0.03
Cambodia	1990-2000	10	422	0.02
Lao PDR	1990-2000	9	412	0.02
Guatemala	1980-1990	9	331	0.03
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
Malawi	2001-2008	8	268	0.03
Lao PDR	2001-2008	8	287	0.03
Botswana	2001-2008	8	304	0.03
Namibia	2001-2008	8	211	0.04
Turkey	1970-1980	7	199	0.04
Uruguay	2001-2008	7	207	0.03

This table presents statistics on the countries with the largest amount of "jumps" in the data.

were being exported with an RCA below 0.1. The bottom list, based on PRB products, shows the repeated appearance of many countries from the top list, but also includes many African countries. The presence of Germany in the list is surprising, given that it is the only developed country that appears. However, its high ranking in the period 1990-2000 mgiht well be due to classification errors associated with the reunification of the country [15].

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

$$J_{c,p,t\rightarrow T} = \alpha + \beta ln(RCA_{c_N,p,t}) + controls_{c,p,t} + \varphi_{p,t} + \mu_{c,c_N,t} + \varepsilon_{c,p,t} \tag{1.4}$$

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

where $J_{c,p,t\rightarrow T}$ is a binary variable that takes the value of 1 when there was a "jump" between year t and T in product p and country c. The variable of interest, $ln(RCA_{c_N,p,t})$, is the natural logarithm of the RCA of the neighbor with the largest RCA in product p for country 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[16]. To correct for undefined growth rates caused by zeros in the denominator, we compute the growth rate using RCA+0.1 for all observations, thus pairing down the rate of growth for very low RCA products. To control for our own correction, we also add a dummy variable indicating whether the RCA was zero at the initial year of the computed growth rate used in the right hand side of the specifications, which are the observations more likely to be distorted. We also control for the "density" of the country in the product at the beginning of the period. The variable "density", which distributes between 0 and 1, was developed by Hausmann and Klinger (2006) and used in Hidalgo et. al. (2007). It measures the intensity with which a country exports products that are strongly co-exported by other countries who also export the product under consideration. In other words, the density of a product proxies for the existence of other exports that share similar technologies or inputs (as measured by their co-occurrence across countries). Density strongly affects the likelihood that a country adds the product to its export basket (Hausmann & Klinger, 2007; C. A. Hidalgo et al. 2007). We use density to control for the likelihood that a country would jump to a product given the initial composition of its export basket[17]. $\varphi_{p,t}$ are product-year fixed effects which control for any time-varying product characteristic such as global demand, price or productivity shocks, particular to

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

[17] 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.

product p. $\mu_{c,c_N,t}$ are country-neighbor-year fixed effects, which use the neighbor with the largest RCA in that product. By adding $\mu_{c,c_N,t}$ we control for time-varying country-neighbor aggregate characteristics such as similarity in institutions, geography, climate, culture, history, productivity, economic development, population, initial factor endowments, inflation, bilateral exchange rates, etc. [18].

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

Table 1.7 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).

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

Table 1.7: *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

That is, the unconditional probability of "jumping" is 1.5% (or 1.6% for NPRB products only). In the right-hand side, there are two variables of interest that we will use interchangeably. First, the continuous variable "Ln Maximum RCA [of] Neighbors", which is the natural logarithm of $RCA_{c_N,p,t}$ (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 1.8. Panel A estimates the model (1.4) with the "Ln Maximum RCA [of] Neighbors" variable as the regressor of interest, while Panel B estimates the same model with the “Neighbor Exports” binary variable. The first two columns in both panels present the results from our original sample. The last two columns in both panels use the control sample with randomly assigned neighbors—as previously explained. Our variables of interests, in both their continuous and binary form, are economically and statistically significant in columns 1 and 2 (which estimate the model with the real sample) and neither economically nor statistically significant in columns 3 and 4 (using the control sample). The economic significance of this result is the following: a doubling in the export intensity of a product by a geographic neighbor (i.e. RCA) at the beginning of the period is associated, on average,with a 0.4 percentage 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%). Panel B of Table 1.8 estimates that if a country has one neighbor who already exports product p with an RCA above 1 at the beginning of the period, then the chance that the country "jumps" to that product increases by 1 percentage point. This represents an increase of roughly 65% for the average product in the probability of "jumping" (from 1.5% to 2.5%) [19].

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 1.9 replaces the dependent variable of specification (1.4) with the compound average annual growth rate of RCA for the same time periods as before. In this exercise we use all the observations in the dataset without the low RCA restriction we used for the extensive margin. The intention is to estimate future growth in exports for a particular product in which a geographic neighbor has revealed comparative advantage, instead of focusing on new appearances. The main difference between the two approaches is that, by looking at the intensive margin, we include products that are already being exported by the country under consideration, and do not limit the sample to non-exported products.

The results in Table 1.9 are organized in the same way as those in Table 1.8. The upper panel estimates the model using the "Ln Maximum RCA [of] Neighbors" variable as the regressor of interest, while the lower panel estimates the same model with the "Neighbor Exports" binary variable. The results in Panel A show a strong positive association between a country's increase in future product RCA growth and the highest RCA of a neighboring country in that product at the beginning of the period. Panel B shows that having a neighbor with RCA>1 is associated with a future annual growth of RCA of 1.5%-1.7% for that product in the next ten year period (or 16 - 18% cumulative). Table A.8 in section A.5 of the web appendix shows that this result is robust to the use of export growth as the dependent variable, rather than RCA growth.

We repeat this analysis for different regions, periods and types of products, to understand whether there are differential effects across any of these dimensions. That is, we are

[19] See section A.5 of the web appendix for a number of robustness checks of these results, varying the definition of the LHS variable, the method, the sample used and the dataset. All tests show full robustness with the results presented here.

Table 1.8: *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. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Table 1.9: *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.

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

interested in understanding which set of observations in the sample are driving the observed results. Table 1.10 summarizes this exercise by presenting results for a different cut of the data in each row.

The left panel of Table 1.10 uses the maximum log RCA of neighboring countries, and the right panel uses the dummy variable which takes the value 1 if a country has a neighbor with RCA>1 in that product (at the beginning of the period). The left and right panels are analogous to the upper and lower panels of Tables 1.8-1.9 respectively. For instance, the first row considers all observations eligible to "jump" (i.e. with a baseline RCA below 0.1). Of these, 2.52% achieved an RCA above 1 in the following ten years if they had a neighbor with an RCA in that same product in the top 25% of the distribution. The same number drops to 1.2% if the best neighboring exporter had an RCA in the bottom 75% of the distribution. The ratio of these two numbers indicates that the first group was 2.1 times more likely to "jump". The table also presents the 95% confidence interval for the estimate of the coefficient on the neighbor RCA variable in model (1.4). The first row is analogous to the results presented in Table 1.8, but every row recalculates the coefficient for each cut of the data.

From Table 1.10 we find that our results are in fact dominated mostly by developing countries, given that, both in the left and right panel, the estimator for β is statistically significant only for non OECD countries. When we divide the world into regions we see the same pattern. The "neighbor effect" is statistically significant for East Asia & the Pacific, Latin America and the Caribbean and Sub-Saharan Africa, in both specifications.

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

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

Table 1.10: *Intensive and Extensive Margin Comparative Statics*

		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**

This table presents statistics on the likelihood of a country adding a newproduct to its export basket based on whether one of its geographic neighbor is also exporting such product at the beginning of the period. To define whether a neighbor is exporting or not the product under consideration the left and right panels use alternative definitions. The left panel distinguishes the case in which a neighbor exports the product under consideration with an RCA in the top 25% of the distribution, as opposed to the case when a neighbor exports such product in the bottom 75% of the distribution. The right panel distinguishes the case in which a neighbor exports the product under consideration with an RCA above 1, as opposed to an RCA below 1.

1.4 Interpretation of the results

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

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

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

While our observations are what would be expected in a world where knowledge diffusion decays strongly with distance, our results could be driven by factors other than knowledge diffusion. The documented similarity of dynamics in the evolution of export baskets between neighbors could be influenced by a common third factor that expresses itself in the region, albeit not simultaneously—there could be both supply or demand stories. On the supply side, for instance, countries may be on a similar development trajectory, moving—for instance—from agriculture to light manufactures and into more complex products, although one country is ahead of the other. Consequently, neighboring countries become good at the same products, but with a time lag. We try to control for this with the highly significant density variable—which captures a country's own predisposition to move into that product—and by the lagged growth rate of the product's exports in the country.

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

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

1.5 Concluding Remarks

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

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

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

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

Foreign direct investment is also a potential channel. 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 what FDI contributes to the evidence on export similarity we document in this paper.

Labor flows or migration could also be a channel for knowledge spillovers. If knowledge resides in brains, it should move with them. If direct human interaction is key to knowledge spillovers, as suggested by much of the literature quoted above, then people could be an important source of knowledge transmission. For instance, Andersen and Dalgaard (2011) show how the ease of travel can explain shifts in aggregate productivity. Other forms of human interaction may also be involved, including the ease of physical and electronic communication, as well as international ethnic/cultural links (e.g. Stein and Daude, 2007; Giroud, 2012; Kerr, 2008). Whether these effects significantly contribute to the observed product-level geographic correlations remains to be shown.

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

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

Chapter 2

Migration, Knowledge Diffusion and the Comparative Advantage of Nations[1]

2.1 Introduction

Franschhoek valley, a small town in the Western Cape province of South Africa, is known today for its beautiful scenery and for its high-quality wineries. The town was founded in the late 17th century by French Huguenot refugees, who settled there after being expelled from France following King Louis XIV elimination of the Edict of Nantes. As of today, the wineries in Franschhoek are among the main producers of South African wine exports. Likewise, Saxenian (2006) relates the story of Dov Frohman, an Israeli scientist who in 1974 returned home after years of having worked in Intel Corporation in the United States. Upon his return to Israel, Frohman founded in Haifa Intel's first design center outside the United States. As of today, Israel is an exporter of semiconductors related technologies. In this paper we explore the role of migrants in developing the comparative advantage of both their sending and receiving countries.

[1]This paper is a part of a larger research project in conjunction with Hillel Rapoport.

Ricardian models of trade assume as given the exogenous productivity parameters that define the export basket of countries which are generated in equilibrium. A large part of the literature has focused on understanding the characteristics of this equilibrium and the mechanisms through which it is conceived. However, a burgeoning literature has been dealing with understanding the evolution of what defines these productivity parameters. This paper contributes to this literature by documenting industry-specific productivity shifts as explained by the variation in international factors movement with particular focus on migration. We study productivity by exploiting changes in the export baskets of countries. The key assumption is that, after controlling for product-specific shifts in demand, firms in a country will be able to export a good only after they have become productive enough to compete in global markets. Of all international factors flows, the results point to migration as the strongest of those drivers. We find that migrants, and even more so, skilled migrants, can explain variation in good-specific productivity as measured by the ability of countries to the export those goods, for products that are intensively exported in the migrants' home/destination countries. In particular we find that, on average, a stock of migrants larger by 65,000 people is associated with a 15% increase in the likelihood of exporting a new product for a given country, whereas the same figure for skilled migrants is reduced to 15,000. Also, in terms of expanding the export basket of countries, a migrant is worth about US $30,000 of foreign direct investment (FDI), while a skilled migrant is worth over $160,000.

This differs from the previous approaches in the literature that look at the link between international factor flows and changes in aggregate productivity, as opposed to industry-specific productivity dynamics. That literature includes, for instance, the work of Coe and Helpman (1993) and Coe et. al. (2009) who study changes in aggregate productivity as a result of importing more from countries with higher R&D investment. Aitken & Harrison (1999) and Javorcik (2004) are among the long list of studies trying to establish whether FDI generates productivity spillovers on domestic firms, with no definite answer emerging from all of them. Andersen and Dalgaard (2011) find a correlation between aggregate productivity and business travel flows.

We consider three alternative explanations on how migration could be associated to good-specific productivity increases. First, if a given country c receives migrants from countries exporters of a given product p, then there could be a local shift in demand for product p, given the plausible shift in aggregate preferences. This could result in a demand-driven productivity shift, which could become exports to either the migrants' sending/receiving country or global exports to the rest of the world, supplying the increase in global demand.[2].

Second, migrant networks could generate lower transaction costs for bilateral trade in specific goods, thus inducing bilateral exports between the sending and receiving country of the migrants (Kugler and Rapoport 2011).

Finally, migrants can serve as a transmission vehicle of product-specific knowledge, which could induce productivity shifts and in turn inducing global competition in certain goods. The acquisition of industry-specific knowledge is an important input for the productivity dynamics of a firm: more knowledge (either through learning or experience) allows economic agents to do more with the same resources. Then, the question remains: can migrants induce exports through knowledge transmission? Bahar et. al. (2014) present evidence suggesting that, after controlling for product-specific global demand, the evolution of the export basket of a country, both in its extensive and intensive margin, could be explained by the documented local geographic character of knowledge diffusion (e.g. Jaffe, Trajtenberg and Henderson 1993; Bottazi and Peri 2003; Keller 2002; Keller 2004). They attribute their result to the fact that knowledge is often non-easily transferrable, mostly because a large component of it is "tacit" (Polanyi 1966). Polanyi explained tacit knowledge by saying that we know more than we can tell, and Kenneth Arrow (1969) suggested that the drivers for knowledge transmission are human minds rather than written words, also when it comes to economic processes. Thus, intuitively, migration would be the natural candidate among all international factor flows to serve as a driver of tacit knowledge and

[2] Linder (1961) suggests, in this case, country c will become a trade partner of the home countries of the migrants.

thus induce exports. This is, precisely, what this paper documents.

The paper uses a worldwide dataset that includes bilateral trade, FDI and migration stock figures for years 1990 and 2000. From it, we construct a sample that includes for each country, product and year the total exports to the rest of the world. The sample also includes the computed total stocks of trade, FDI and migration (disaggregated in immigrants and emigrants) to or from partner countries that export that each product in years 1990 and 2000.

The undertaken empirical exercise looks at how migration figures correlate with a country's extensive and intensive margin of trade. The extensive margin is measured by looking at the future addition of a new product to a country's export basket, while the intensive margin refers to the future annual growth rate of a product that is already exported by a country. We control for global demand of each good by adding product-year fixed effects. We also add country-year fixed effects which would control for all country level variables characteristics that would make a given country more likely to export and receive migrants at the same time. We also calculate all of the specifications using an alteration of the dependent variable, which measures exports to the rest of the world excluding flows to countries where migrants are in or from. These controls, we argue, would rule out the first two explanations discussed above.

We are left with the third explanation, which is our preferred one. Yet, endogeneity concerns are present. They are reduced by adding proper controls and, also, by presenting a set of results that instrument for migration stocks using geographic and cultural bilateral variables between countries the sending and receiving countries of the migrants. The instruments provide an exogenous variation to the number of migrants in/from countries. They are based on the share of the migrants' sending and receiving countries exporting the product under consideration which have a common language, a common colonizer or a (former) colony-colonizer relationship with the country under analysis.

The body of the paper discusses in detail all the data collection, the empirical strategies and present the results. The paper is divided as follows. The next section describes the empirical strategy and the data. Section 2.3 presents the main results, and Section 2.4 discusses them. Section 3.6 concludes.

2.2 Empirical Strategy

2.2.1 Research Question and Empirical Challenges

The empirical strategy studies the relationship between international factor flows and the dynamics in the export basket of the receiving and sending countries, with emphasis on migration. In particular, the question is: can migrants induce product-specific productivity shifts in their sending (destination) countries, on products already intensively exported in their destination (sending) countries?

For the sake of better understanding, we use the following hypothetical example. Suppose there are two countries in the world: Italy (a pizza exporter), and the US (a hamburger exporter). The analogous question then becomes whether the presence of more Italians in the US is associated with the ability of the US to export pizza, and, whether this same presence is also associated with the ability of Italians to export hamburgers.

There are two main empirical challenges in studying the relationship between productivity and international factor flows (i.e. goods, capital and people). First, all flows are highly correlated among themselves. Moreover, several empirical studies have shown that migration networks are an important determinant of bilateral trade flows and bilateral FDI.[3]

Kugler and Rapoport (2011) even find evidence of complementarity between the three types of flows. More specifically, the authors claim, migration generate links that lower transaction costs inducing bilateral FDI and trade.

Hence, the positive correlation between international flows of capital, goods and labor is a matter of consideration to any study of this kind. In fact, in the sample for year 2000, the correlation matrices between total migration, FDI and trade across countries are all positive,

[3]e.g. Gould, 1994; Rauch and Trindade, 2002; Combes, Lafourcade and Mayer, 2005; Iranzo and Peri 2009; Felbermayr and Jung, 2009; Tong, 2005; Kugler and Rapoport, 2007, 2011; Javorcik et. al. 2011

Table 2.1: *Correlation Matrix International Flows (log)*

Variables	Migrants (log)	FDI (log)	Trade (log)
Migrants (log)	1.000		
FDI (log)	0.455	1.000	
Trade (log)	0.457	0.738	1.000

Table 2.2: *Correlation Matrix International Flows (per capita)*

Variables	Migrants (p.c.)	FDI (p.c.)	Trade (p.c.)
Migrants (p.c.)	1.000		
FDI (p.c.)	0.159	1.000	
Trade (p.c.)	0.423	0.538	1.000

and above 0.4, with the exception of migration and FDI per capita (see Tables 2.1 and 2.2). That is, countries that receive/send more FDI tend to also receive/send more migrants and export/import in larger quantities. Hence, to deal with this challenge, the empirical specification controls for all three factors simultaneously.

The second challenge refers to the risk of having biased estimates generated by endogeneity, even after controlling for all factor flows. For instance, migrants could relocate themselves based on future potential of specific sectors of growing, or a third variable (i.e. "openness" shock) could induce migration and induce exports at the same time (though not necessarily of specific products). While the controls in the empirical specification intents to deal with some of these problems, the identification problem remains. In order to further reduce these concerns, we implement a number of instrumental variables, which exploit the variation in migration stocks explained by bilateral cultural/historic characteristics between the sending and receiving countries of these migrants.

2.2.2 Empirical Specification

The aim of the paper is to study the dynamics of the extensive and intensive margin of trade (with exports to the rest of the world) given different levels of migration stocks, controlling for FDI and trade stocks. The specification will disentangle between immigration and emigration, and between all vs. skilled migrants.

Throughout the paper we will use the concept of Revealed Comparative Advantage (RCA) by Balassa (1965), which will be used to construct export-related variables both in the left-hand-side and right-hand-side of the specification. RCA is defined as follows:

$$RCA_{c,p} \equiv \frac{exp_{c,p} / \sum_{p} exp_{c,p}}{\sum_{c} exp_{c,p} / \sum_{c}\sum_{p} exp_{c,p}}$$

where $exp_{c,p}$ is the exported value of product p by country c. This is a yearly measure.

For example, in the year 2000, soybeans represented 4% of Brazil's exports, but accounted only for 0.2% of total world trade. Hence, Brazil's RCA in soybeans for that year was $RCA_{Brazil,Soybeans} = 4/0.2 = 20$, indicating that soybeans are 20 times more prevalent in Brazil's export basket than in that of the world.

The empirical specification is defined as follows:

$$\begin{aligned} Y_{c,p,t\to T} &= \beta_{im} \sum_{c'} immigrants_{c,c',t} \times R_{c',p,t} + \beta_{em} \sum_{c'} emigrants_{c,c',t} \times R_{c',p,t} \\ &+ \beta_{FDI} \sum_{c'} FDI_{c,c',t} \times R_{c',p,t} + \beta_{trade} \sum_{c'} trade_{c,c',t} \times R_{c',p,t} \\ &+ \gamma Controls_{c,p,t} + \alpha_{c,t} + \eta_{p,t} + \varepsilon_{c,p,t} \end{aligned} \tag{2.1}$$

The definition of the dependent, or left hand side (LHS) variable, $Y_{c,p,t\to T}$, alternates according to whether the specification is studying the intensive or the extensive margin of trade for a specific product p and country c. When studying the *extensive* margin, $Y_{c,p,t\to T}$ is 1 if country c achieved an RCA of 1 or more in product p in the period of time between t and T (conditional on having an $RCA_{c,p,t} = 0$. That is:

$$Y_{c,p,t\to T} = 1 \text{ if } RCA_{c,p,t} = 0 \text{ and } RCA_{c,p,T} \geq 1$$

In addition, we also condition $Y_{c,p,t\to T} = 1$ to whether whether $RCA_{c,p,t-1} = 0$, where $t-1$ refers to the beginning of the previous period (i.e. for the period 2000-2010, it is 1990)[4]. This eliminates cases in which the country was already exporting such product but for some

[4]For the period 1990-2000, $t-1$ is 1985 given data limitations

reason they stopped doing so at the beginning of the period under consideration.

When studying the *intensive margin*, $Y_{c,p,t \to T}$ is the annual (log) growth rate in the exports value of product p, between years t and T, conditional on having $exports_{c,p,t} > 0$. That is:

$$Y_{c,p,t \to T} = \frac{ln(exports_{c,p,T} + 1) - ln(exports_{c,p,t} + 1)}{T - t} \text{ if } exports_{c,p,t} > 0$$

The independent variables include the following:

- The sum of the stock of immigrants and of emigrants (in logs) from and to other countries (denoted by c′) at time t, weighted by a dummy $R_{c',p,t}$ which is 1 if $RCA_{c',p,t} \geq 1$[5].

- The sum of stock of FDI and stock of trade (in logs) according to the previous logic.

- A vector of controls of baseline variables (when applicable): the baseline level of exports (in logs) for that same product; the average annual (logarithmic) growth rate of the export value in the previous ten year period (in order to control for previous trends in the export dynamics for that product); in order to correct for undefined growth rates caused by zeros in the denominator, we compute the growth rate using $exports_{c,p,t} + 1$ for all observations; in addition, to control for our own correction, we also add as a control a dummy variable indicating whether $exports_{c,p,t} = 0$, which correspond to the observations most likely to be distorted.

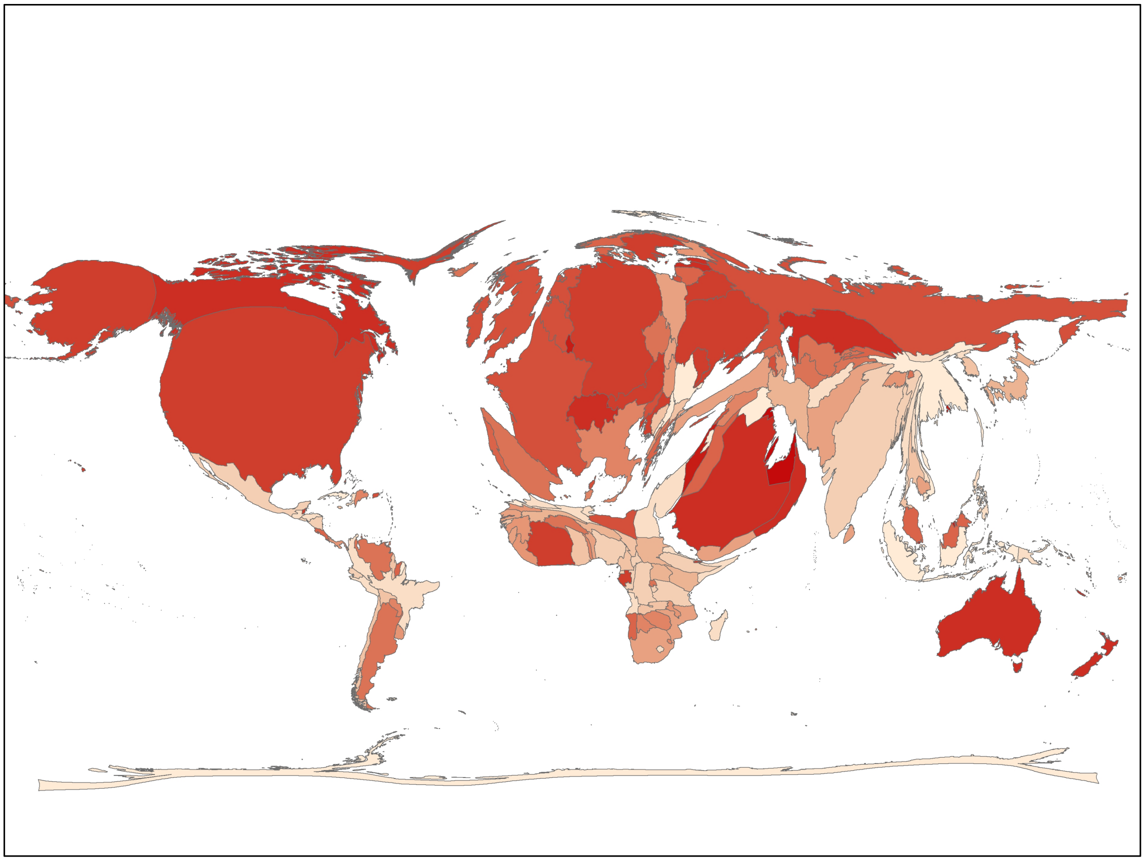
- Country-by-year and product-by-year fixed effects.

2.2.3 Data and Sample

Bilateral migration data comes from Docquier, Ozden and Perry (2010). The dataset consists of total bilateral working age (25 to 65 years old) foreign born individuals in 1990 and 2000. The data provide figures for skilled and non-skilled migrants at the bilateral level as well.

[5]In fact, we define several thresholds for $R_{c',p,t} = 1$, which are$RCA_{c',p,t} \geq x$ where $x \epsilon [1, 5]$

Figure 2.1: *Cartogram Share of Migrants, Year 2000*



Skilled migrants are considered to have completed some tertiary education at the time of the census. Figures 2.1 and 2.2 represent the migration data in year 2000.

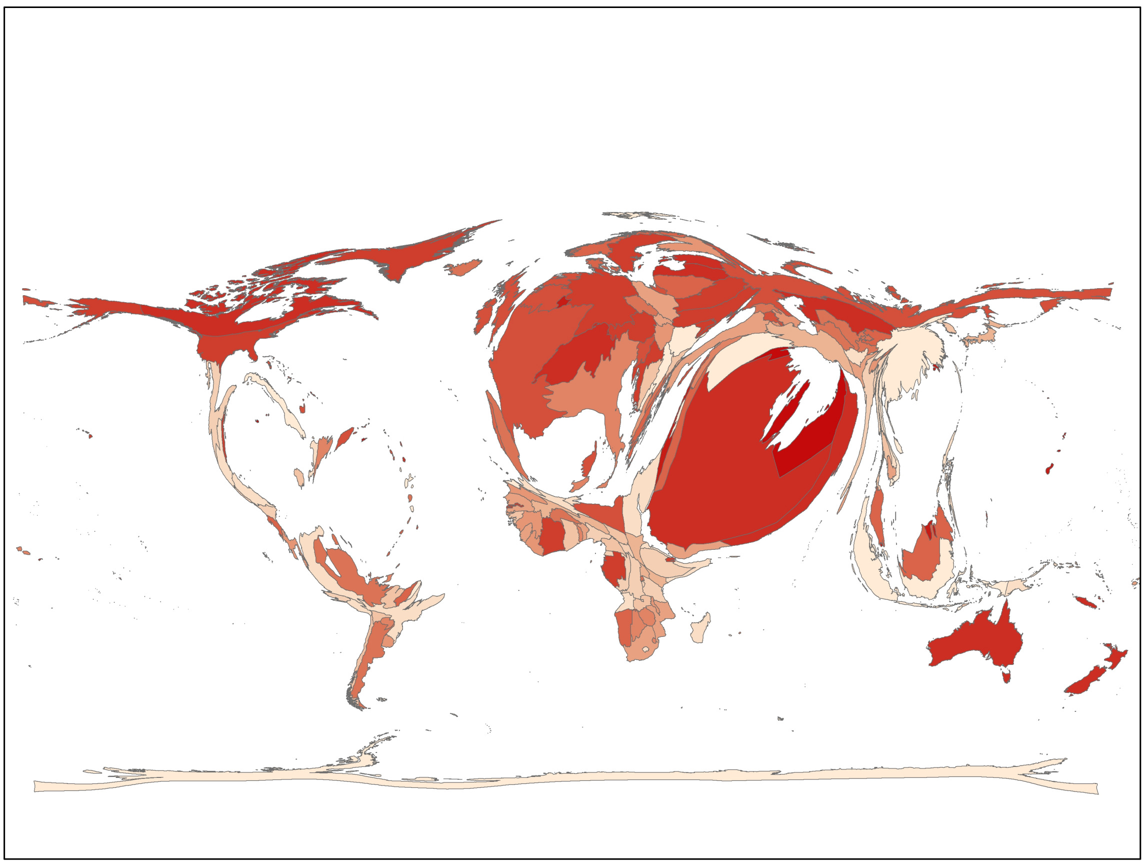
Bilateral FDI positions are from the OECD International Direct Investment Statistics (2012). It tracks FDI from and to to OECD members since 1985 until 2009. Using this data we compute 10-year stocks of capital flows for each country in 1990 and 2000[6]. Negative FDI stocks are treated as zero[7].

Bilateral trade data comes from Hausmann et. al. 2011, based on the UN Comtrade data from 1984 to 2010. We use the trade dataset to construct two variables. First, total exports per

[6]For 1990 we use the stock from 1985 to 1990 due to limitations of the data.

[7]This follows the same methodology suggested by Kugler and Rapoport (2011)Kugler and Rapoport (2011). Only 1.7% of the original dataset is affected by this.

Figure 2.2: *Cartogram of Migrants Per Capita, Year 2000*



product per country to the rest of the world, to be used to compute the dependent variable in the empirical specifications, to study the intensive and extensive margin of trade. The list of products is fairly disaggregated. An example of a product is "Knitted/Crocheted Fabrics Elastic Or Rubberized" (SITC code 6553), or "Electrical Measuring, Checking, Analyzing Instruments" (SITC code 8748). The words product, good and industry interchangeably referring to the same concept throughout the paper. Similarly to FDI, we also compute 10-year stocks for bilateral trade (imports plus exports) to be used in the RHS. Both the 10-year Trade and FDI stocks are deflated using the US GDP deflator (base year 2000) from the World Development Indicators by the World Bank.

Former Soviet Union countries are excluded from the sample given their poor trade data in the period 1990-2000. The final sample consists of 136 countries and 781 products. We define two 10-year periods for the analysis due to the limitations imposed by the bilateral migration data, which are 1990-2000 and 2000-2010.

Furthermore, we incorporate variables from the GeoDist dataset from CEPII on bilateral relationships such distance, common colonizer, colony-colonizer relationship, and common language, to construct the instrumental variables (Mayer and Zignago 2011).

The summary statistics for the variables to be used in the analysis are in Table 2.3. Panel A presents the summary statistics for the extensive margin sample, while Panel B does it for the intensive margin sample. The left and right sides of both panels present the same variables used in the RHS varying the threshold $R_{c',p,t} = 1$, based on $RCA_{c',p,t} \geq 1$ and $RCA_{c',p,t} \geq 5$, respectively. The variables that do not depend on this threshold are reported in the left panel only.

From Panel A we see that the unconditional probability of achieving an RCA above 1 (starting with an none exports) for the average country-product is 3.5%. Similarly, from Panel B , the average country-product exports value annual growth rate is close to zero in the data. The tables also include the sum of immigrants and emigrants for the average country and year from and in countries exporting a product with RCA above 1 (left) and above 5 (right). It presents the same statistics for aggregated FDI and Trade figures in

Table 2.3: *Summary Statistics*

	RCA=1			RCA=5		
Variable	**N**	**Mean**	**sd**	**N**	**Mean**	**sd**
Panel A: Extensive Margin Sample ($exports_{c,p,t} = 0$)						
New Product (RCA>1)	83,397	0.035	0.185	-	-	-
Immigrants	83,397	14233.76	69694.14	83,397	3869.30	32469.86
Emigrants	83,397	49819.20	139403.55	83,397	6657.21	57416.82
Immigrants (HS)	83,397	1862.31	9711.80	83,397	450.21	4396.31
Emigrants (HS)	83,397	12818.14	36970.25	83,397	987.11	5668.17
FDI (total, mn USD)	83,397	59.26	1767.27	83,397	8.12	482.81
Trade (total, mn USD)	83,397	1256.47	4014.39	83,397	129.22	733.90
Panel B: Intensive Margin Sample ($exports_{c,p,t} > 0$)						
Growth Exports	129,035	-0.003	0.349	-	-	-
Baseline Log Exports	129,035	13.259	3.723	-	-	-
Immigrants	129,035	169327.61	582922.35	129,035	32345.20	165795.17
Emigrants	129,035	189433.86	445525.94	129,035	16373.44	90889.80
Immigrants (HS)	129,035	48736.68	220836.55	129,035	7964.71	53017.41
Emigrants (HS)	129,035	55520.94	120329.89	129,035	3422.09	18897.87
FDI (total, mn USD)	129,035	14387.69	60995.89	129,035	644.94	6656.45
Trade (total, mn USD)	129,035	29353.97	67416.95	129,035	2152.64	9568.08

million USD, after the delation process explained above. Note that FDI and Trade variables total inwards and outwards stock figures.

2.3 Results

Table 2.4 presents the estimation of specification (2.1) for all products in the dataset. The upper panel estimates the extensive margin (measured by the likelihood of adding a new product to a country's export basket) while the lower panel estimates the intensive margin (measured by the annual growth in exports of a product already in the country's export basket). It is important to notice that the dependent variables in both panels are computed using exports from country c to product p to the rest of the world. The columns titled "R1" indicate that the threshold used for constructing the RHS variables was $RCA_{c',p,t} \geq 1$, whereas the columns titled "R5" indicate the use of $RCA_{c',p,t} \geq 5$, instead. The table also presents results using all migrants (columns 1 and 3) and skilled migrants (columns 2 and 4) in the RHS. The upper panel of Table 2.4 uses country-product pairs which had zero exports in the baseline years (1990 and 2000), which corresponds to almost 84,000 observations (thus, baseline variables are not included because lack of variation).

The results in Panel A indicate that a country with 10% increase in its stock of immigrants from nations exporters of product p, is associated with an increase of 0.4% in the likelihood the receiving country will export product p with an RCA above 1 in the next ten years.[8] Similarly, a 10% increase in the stock of emigrants residing in countries exporters of product p, will tend to increase the same likelihood for country c by 1%. On average in the sample these numbers amount to about 1,400 immigrants and 5,000 emigrants. The correspondent correlation for an increase of 10% in the total stock of (incoming and outgoing) FDI is about 0.4%[9]. That is, a 3.5 times increase in the stock of FDI, which is about USD 210 million based on the sample average, would account for the same marginal effect of a stock increase of about 6500 migrants (both immigrants plus emigrants). In other words, each migrant is worth about $30,000 of FDI in this context.

Column 2 in Table 2.4 limits the migration figures to skilled immigrants and emigrants only. This significantly reduces the variation in the RHS, as can be seen in the summary statistics. The results suggest that, in the case of immigration, the coefficient estimator is almost doubled and statistically significant. This implies that the probability of a country adding a new product p to its export basket is larger by 0.8% for each 10% increase its stock of immigrants from countries exporters of p (about 200 people, on average). In terms of skilled emigration, a 10% increase in the stock of emigrants in nations exporters of product p (about 1300 people), is associated with an increase of 0.8% in the likelihood a given country will start exporting p in the next 10 years. In total, the figures suggest that, on average, an increase of 1500 in the stock of migrants in and from countries with comparative advantage in p, is associated with a probability of exporting p larger by 1.6%. To achieve the same

[8] A 10% increase is associated with an increase in such likelihood of 0.015 percentage points, which based on the unconditional probability of 3.5% (see Table 2.3), corresponds to a 0.4% increase.

[9] Or even zero, given that the estimator for the FDI coefficient is statistically insignificant.

Table 2.4: *Fixed Effects*

	R1		R5	
Panel A: Extensive Margin				
	All	Skilled	All	Skilled
Ln Immigrants	0.0015	0.0028	0.0005	0.0011
	(0.000)***	(0.001)***	(0.000)	(0.001)
Ln Emigrants	0.0032	0.0026	0.0019	0.0018
	(0.001)***	(0.001)**	(0.001)***	(0.001)***
Ln FDI, total	0.0015	0.0014	0.0001	0.0001
	(0.001)	(0.001)	(0.000)	(0.001)
Ln Trade, total	-0.0080	-0.0084	0.0014	0.0014
	(0.004)*	(0.004)*	(0.001)	(0.001)
N	83397	83397	83397	83397
r2	0.15	0.15	0.15	0.15
Panel B: Intensive Margin				
	R1		R5	
	All	Skilled	All	Skilled
Ln Immigrants	0.0030	0.0043	0.0029	0.0021
	(0.001)***	(0.001)***	(0.001)***	(0.001)**
Ln Emigrants	0.0069	-0.0006	0.0042	0.0043
	(0.002)***	(0.002)	(0.001)***	(0.001)***
Ln FDI, total	-0.0015	-0.0013	-0.0002	-0.0002
	(0.000)***	(0.000)***	(0.000)	(0.000)
Ln Trade, total	0.0094	0.0165	0.0012	0.0021
	(0.005)*	(0.005)***	(0.001)*	(0.001)***
Baseline Log Exports	-0.0298	-0.0295	-0.0290	-0.0285
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Previous Exports Growth	-0.1373	-0.1379	-0.1397	-0.1413
	(0.013)***	(0.013)***	(0.013)***	(0.013)***
Previous Exports Zero	-0.0044	-0.0042	-0.0007	0.0001
	(0.013)	(0.013)	(0.013)	(0.013)
N	73193	73193	73193	73193
r2	0.35	0.35	0.35	0.35

All specifications include country-by-year and product-by-year fixed effects. SE clustered at the country level presented in parenthesis

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

result with FDI would require a fourfold increase in FDI, which amounts to about USD 240 million. Using the same logic as above, this means that each skilled migrant is worth about USD $160,000 in terms of expanding a country's export basket in the ways studied by this paper.[10]

Panel B of Table 2.4 uses the annual export value (logarithmic) growth rate as the dependent variable, in order to study the intensive margin of trade. The number of observations is different than the sample used for Panel A, because we are using all products with export value above zero in the baseline year. The results present evidence that both the presence of immigrants from and of emigrants in countries exporters of product p, is associated with a larger future rate of growth in export value of product p in the country under consideration. In particular, for a given product p, a 10% increase in the stock of immigrants from countries exporting such product (about 17,000 people) is associated with an increase in the future annual growth rate in export value for the receiving country of about 0.03 points. Similarly, future annual growth in exports value of product p tends to be 0.07 percentage points higher with a 10% larger stock of emigrants in countries exporters of the same product (about 19,000 people). That is, about 40,000 more migrants, on average, is associated with a annual growth rate that is larger by 0.1 points. The coefficients in Column 2, which use skilled migration in the right hand side, are larger in magnitude in the case of immigrants, though the coefficient for emigration becomes statistically insignificant for the case of emigrants.

Columns 3 and 4 repeat the exercise of Columns 1 and 2 in both panels, but redefining $R_{c',p,t} = 1$ if $RCA_{c',p,t} \geq 5$. This means that the right hand side variables are weighted by whether the partner countries have a revealed comparative advantage that is 5 times the world average in product p at time t. Intuitively, the average migrant from and/or in these countries will have a higher likelihood of being exposed to the productive knowledge that is required to efficiently produce (and thus, being able to export) product p. At the same

[10]The comparison of migrants with trade figures is not possible given that the estimated coefficients for trade are negative. This is a natural result given that countries tend to trade less with other countries that export the same goods.

time, the variation for all variables is considerably less.

For this case, the results show no significant correlation with immigrants. However, the results suggest that a 10% increase in the stock of emigrants, which amounts to about 660 people, is associated with a larger probability of achieving comparative advantage by 0.5%. The same figure for skilled emigrants, which corresponds to about 100 people on average, is also 0.5%. That is, less than 1000 emigrants, on average, in countries that export product p with an RCA above 5, could increase the likelihood of their sending countries achieving comparative advantage in the same product by 1%. In this case, the comparison with FDI suggest that each emigrant is worth USD 800,000, or USD 900,000 of trade!

An interesting implication of the results is that FDI and Trade figures, in most cases, seem not to correlate with the ability of countries to expand their the export baskets under the studied context. That is, trading with countries which are exporters of a particular product is not associated with the likelihood of gaining comparative advantage in that same product. However, when it comes to the intensive margin of trade seem to positively correlate with the future annual growth of export value. Precedents of this result tracks to Coe and Helpman (1995), where they find evidence on how trade leads to increases in aggregate productivity.

Based on valid concerns on how much of these results are being driven by good pinned down by geology or climate conditions, we estimate specification (2.1) excluding those products from the sample. The results are robust to their exclusion. See Section B.2.2 in the Appendix for more details.

All the specifications presented above include product-by-year fixed effects and country-by-year fixed effects. The former set of fixed effects would control for global demand for all products. Given that we are looking at exports to the rest of the world, the shifts we identify must be related to the supply side. The country-by-year fixed effects would control for time invariant countries' characteristics, as well as country-level aggregate demand and supply shocks, that would rule out a country-level third factor that positively correlates with both migrant figures and overall productivity, such as an openness shock.

2.3.1 Increased productivity or lower bilateral trade costs?

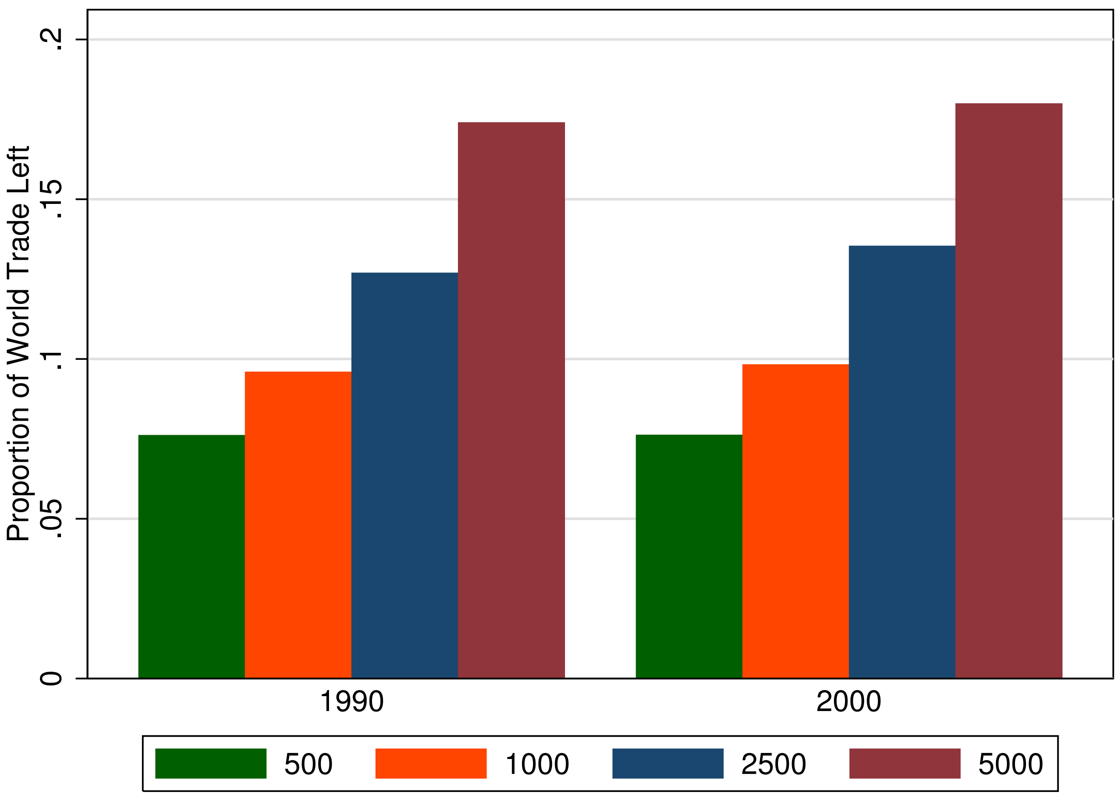
A valid concern would be that the partial correlations we are observing are being driven by bilateral trade: the country is exporting more of the product to those countries where the migrants are from/in. This relates to the evidence presented by Kugler and Rapoport (2011), who find that migrants facilitate the creation of business networks which induces bilateral trade and capital flows. Under this possibility, it would be harder to attribute the results to a gain in productivity, but to a decrease in trade costs. In order to deal with this we estimate again the same specification, but we exclude from the dependent variable all exports to countries where migrants are from/in. That is, we reconstruct the dataset such that the export value to the rest of the world for each product and country combination, excludes exports to nations that send or receive that same country's migrants.

Of course, a critical caveat is that the exclusion requires defining a threshold on the number of sending/receiving migrants. If one migrant is enough to activate this rule we will probably clean all world trade, given that there is always one alien citizen of every country in most developed nations, which generate the largest share of world trade. In this sense, we define a number of arbitrary thresholds which are 500, 1000, 2500 and 5000 migrants. For example, let's suppose we are looking at Canadian exports of television sets to the rest of the world in year 1990. We will exclude from that figure exports of TV sets from Canada to countries that (1) have a number X of Canadians migrants and (2) a number Y of their citizens are migrants in Canada, as long as X+Y is larger than 500, 1000, 2500 and 5000. The assumption is that in order to create an effective business network one would need more than 500, 1000, 2500 or 5000 migrants among the two countries.

In fact, Figure 2.3shows the magnitude of the reduction of total trade figures after revising the exports figures as explained above. For instance, with the 500 threshold world trade figures are reduced by about 92.5%; while using the 5000 threshold reduces total trade figures by about 83%.

Nevertheless, despite the strong decline in the variation of the dependent variable, the results show consistent patterns with the previous results. For instance, Table 2.5 shows

Figure 2.3: *Proportion of World Trade Left*



results using the 500 threshold (the most conservative one), while the tables using the other thresholds are presented in the Appendix Section B.2.2. The results are qualitatively the same as in Table 2.4. However, all in all, the observed correlation between migrant stocks and increased exports is not explained by migration-induced lower trade costs, but rather good-specific productivity increases. In fact, for Panel A, the estimates are similar in magnitude than in Table 2.4. In Panel B, the results are less robust, but the correlation with emigrants in countries with an RCA above 5 is consistent with the previous results.

In spite of having shown that the results are not driven by bilateral migrant networks there is still room for endogeneity concerns, which keeps us from concluding anything causal on the relationships we have found so far. The next subsection deals with this issue and attempts to solve some of these concerns.

2.3.2 Instrumental variables approach

The documented correlations can be partly driven by endogeneity: migrants relocate themselves following potential growth in products they are familiar with. In order to reduce endogeneity concerns, we generate a number of instrumental variables that will serve as exogenous variation to migration figures.

Given that there are two regressors we are interested in (immigrants and emigrants), an instrumental variable approach would require at least two instruments. In the spirit of Frankel & Romer (1999) we construct a number of instruments based on bilateral characteristics of the country under consideration and the countries of its immigrants/emigrants, using data from CEPII's GeoDist database.

For each combination c, p (with $RCA > 1$) and t there is a defined set of countries c' where immigrants are *from* and emigrants are *in* (i.e. Θ^I and Θ^E respectively). Based on this, there are three pairs of instruments that are constructed. They are the *share* of countries Θ^I and Θ^E that:

- speak a common language as c
- have the same current or historic colonizer as c

Table 2.5: *Fixed Effects, excluding bilateral exports (500 migrants threshold)*

Panel A: Extensive Margin				
	R1		R5	
	All	Skilled	All	Skilled
Ln Immigrants	0.0019	0.0030	0.0008	0.0008
	(0.000)***	(0.001)***	(0.000)**	(0.001)
Ln Emigrants	0.0032	0.0023	0.0012	0.0008
	(0.001)***	(0.001)**	(0.000)***	(0.000)*
Ln FDI, total	0.0008	0.0008	-0.0001	-0.0001
	(0.001)	(0.001)	(0.000)	(0.000)
Ln Trade, total	-0.0063	-0.0063	0.0028	0.0029
	(0.004)	(0.004)	(0.001)**	(0.001)**
N	119783	119783	119783	119783
r2	0.11	0.11	0.11	0.11
Panel B: Intensive Margin				
	R1		R5	
	All	Skilled	All	Skilled
Ln Immigrants	-0.0007	0.0019	0.0016	0.0012
	(0.001)	(0.002)	(0.001)*	(0.001)
Ln Emigrants	0.0084	0.0012	0.0028	0.0024
	(0.002)***	(0.003)	(0.001)***	(0.001)*
Ln FDI, total	-0.0009	-0.0008	0.0010	0.0010
	(0.001)	(0.001)	(0.000)**	(0.000)***
Ln Trade, total	0.0097	0.0147	0.0021	0.0027
	(0.006)	(0.006)**	(0.001)**	(0.001)***
Baseline Log Exports	-0.0038	-0.0037	-0.0029	-0.0026
	(0.002)*	(0.002)*	(0.002)	(0.002)
Previous Exports Growth	-0.1077	-0.1073	-0.1103	-0.1114
	(0.019)***	(0.019)***	(0.019)***	(0.019)***
Previous Exports Zero	-0.0645	-0.0650	-0.0599	-0.0595
	(0.018)***	(0.018)***	(0.019)***	(0.018)***
N	55580	55580	55580	55580
r2	0.32	0.32	0.32	0.32

All specifications include country-by-year and product-by-year fixed effects. The dependent variable in all specifications is constructed using exports of country c to the whole world excluding to countries c' where total migration between c and c' exceeds 500 people.

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

- have/had a colony-colonizer relationship with c

For the instruments to be valid, the exclusion restriction must be that, *product specific* exports to the *the whole world* are not correlated with common bilateral cultural/historic ties with its migrants' countries, once we control for country-year fixed effect. Furthermore, we assume that countries do not engage into *product specific* export-inducing agreements based on their cultural or historical ties, which *are not* captured via flows such as FDI or trade.

The assumptions above are not testable, and their validity -as any other exclusion restriction- are not fully guaranteed. However, following previous literature, we believe the assumptions are reasonable, and present the results using these instruments.

The relevance of the instrument is fully testable. For intuition purposes, figure 2.4 presents the analogous of a first stage in a one endogenous variable 2SLS regression, using the United States and South Africa as examples. In both the left and right panels, each observation is a product labeled with its SITC 4-digit code. The left panel uses the United States in year 2000 as an example. The vertical axis measures the number of immigrants (in logs) it has received from countries exporters of each product, while the horizontal axis measures what share of those countries speak english. The right panel uses South Africa in year 2000, and measures the same relationships through emigrants, rather than immigrants. In both panels we see a strong positive correlation, which represents what would be a first stage.

However, since the specification includes n>1 endogenous regressors, testing for the relevance of the instruments is not straightforward. Stock and Yogo (2002) define critical values on the a number of cases involving n>1 endogenous regressors, to be used with the Kleibergen-Paap F statistic (when not assuming homoskedasticity). The critical value is 15.72 for the case of 2 endogenous and 6 instruments. Being above this value implies the bias caused by a weak instrument is lower than the OLS bias with 95% certainty. The Kleibergen-Paap F statistic will be reported in all regressions.

Results using the instrumental variables through the generalized methods of moments are presented in Table 2.6.

Figure 2.4: *First stage, common language*

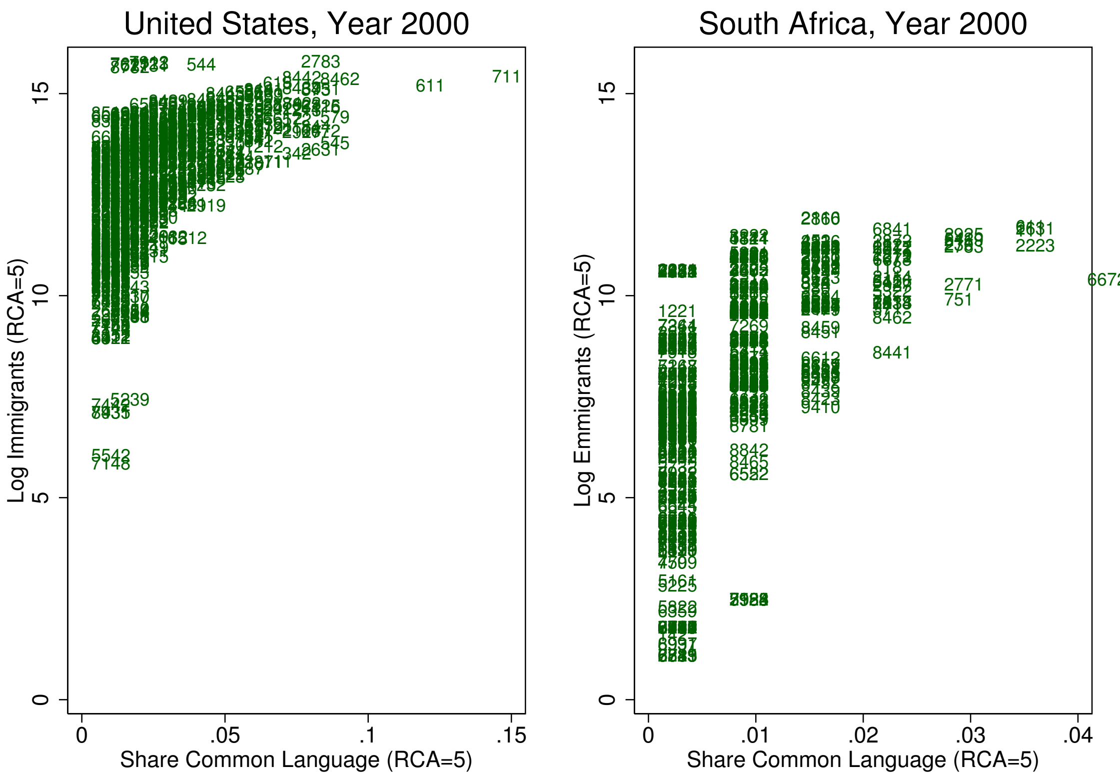


Table 2.6: *Instrumental Variables Estimation (GMM)*

Panel A: Extensive Margin				
	R1		R5	
	All	Skilled	All	Skilled
Ln Immigrants	-0.0004	-0.0014	0.0002	-0.0005
	(0.002)	(0.004)	(0.001)	(0.002)
Ln Emigrants	0.0110	0.0154	0.0050	0.0082
	(0.003)***	(0.006)***	(0.001)***	(0.002)***
Ln FDI, total	-0.0001	-0.0003	-0.0004	-0.0004
	(0.001)	(0.001)	(0.000)	(0.000)
Ln Trade, total	-0.0072	-0.0084	0.0003	0.0004
	(0.005)	(0.005)*	(0.001)	(0.001)
N	83396	83396	83396	83396
r2	0.14	0.14	0.15	0.14
KP F Stat	20.38	15.34	84.01	24.19
Panel B: Intensive Margin				
	R1		R5	
	All	Skilled	All	Skilled
Ln Immigrants	0.0165	0.0224	-0.0022	-0.0033
	(0.008)**	(0.007)***	(0.004)	(0.004)
Ln Emigrants	-0.0074	-0.0193	0.0146	0.0162
	(0.013)	(0.014)	(0.004)***	(0.004)***
Ln FDI, total	-0.0013	-0.0009	-0.0004	-0.0005
	(0.001)**	(0.001)	(0.000)	(0.000)
Ln Trade, total	0.0056	0.0098	-0.0002	0.0005
	(0.009)	(0.009)	(0.001)	(0.001)
Baseline Log Exports	-0.0306	-0.0307	-0.0295	-0.0287
	(0.002)***	(0.002)***	(0.002)***	(0.001)***
Previous Exports Growth	-0.1270	-0.1284	-0.1328	-0.1330
	(0.013)***	(0.013)***	(0.013)***	(0.013)***
Previous Exports Zero	-0.0105	-0.0098	-0.0033	-0.0050
	(0.013)	(0.013)	(0.012)	(0.012)
N	73193	73193	73193	73193
r2	0.34	0.34	0.34	0.34
KP F Stat	3.95	7.60	15.33	14.50

All specifications include country-by-year and product-by-year fixed effects. SE clustered at the country level presented in parenthesis.

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

Note that in all specifications in Panel B, the Kleibergen-Paap F statistics reveal the weakness of the instruments, so there is little we can say about the intensive margin of trade according to these results. In Panel A, however, the instrument seems to have a strong enough first stage (besides in Column 2), which allow us to draw conclusions.

First, immigration seems not to be correlated with the ability of countries to add new goods to their export baskets. However, when it comes to emigration, the magnitude of the coefficients are much larger, as compared to the ones in Table 2.4. For instance, based on Columns 1, a 10% increase in the stock of emigrants (about 5000 people) increases the likelihood of exporting a product exported in the receiving countries of these migrants by 3% (or 0.11 percentage points). Column 2 shows that the same increase in the stock of skilled emigrants (about 1300 people) could increase by 4% (or 0.15 percentage points) the likelihood of including a new good to the export basket. Again, we see how skilled emigrants have a much stronger correlation.

Columns 3 and 4 show a similar picture. While the magnitude of the estimators are smaller, the variance in the number of emigrants using the $RCA \geq 5$ threshold is also lower. Thus, the effect is much larger when measuring according to the actual number of emigrants that could boost the export basket of their sending countries. That is, only about 700 emigrants and 100 skilled emigrants residing in countries that export a particular good with an RCA that is five times the world average, would increase the likelihood of exporting that same good by 1.5% or 2.2%, respectively.

If the exclusion restrictions presented before are valid, and the results cannot be attributed to a third uncontrolled for variable, then these results are particularly strong and a solid contribution to the literature. The presence of migrants from or in nations that export a particular good induce a productivity shift in the sending and receiving country of the migrants, which results in the diversification of their export baskets.

2.4 Discussion and Interpretation

The results in the previous section show through a variety of ways that migration, in both directions, is a determinant of the evolution of the comparative advantage of nations. What is the logic behind that?

If knowledge is tacit, and thus it requires human interaction for its transmission and diffusion, then we could expect that migrants are a driver of such process, which results in increased productivity of the particular sectors that are especially productive in the sending and receiving countries of the migrants.[11] The results are consistent with such hypothesis.

The idea that immigrants could bring productive knowledge is obvious. They are physically present in their receiving country, and thus they interact with the local population in ways that could lead to the diffusion of knowledge. But, why would we see a similar pattern with emigrants? There are two possible mechanisms in place: return migration or communication with their family and friends back home.

Return migration seems like a very plausible one. After all, estimates show that about 30% of emigrants return to their home countries after some period of time (e.g. Borjas and Bratsberg, 1996). These migrants spend enough time in the foreign country to be part of the labor force, which eventually could lead to generate industry-specific productivity shifts back home. More recently, Choudhury (2014) shows how Indian return migrants induce productivity improvements in their firm back home, after spending time in the multinational corporation headquarters abroad.

The second mechanism relates to links and open communication channels between the emigrants and their co-nationals back home. Thus, industry-specific knowledge diffusion are more prevalent whenever these links between individuals across nations are open: more communication, more short-term travel, etc. The identification of the exact mechanism, however, is part of our future research agenda.

[11]Section B.1 of the Appendix outlines a simple model that formalizes this idea.

2.5 Concluding Remarks

This paper presents evidence suggesting that migrants are a source of evolution for the comparative advantage of nations; a relationship that has not been documented in the literature thus far. The results contribute to the growing literature that aims to explain the evolution of industry-specific productivity of countries, and to the literature of international trade that aims to understand, in a Ricardian framework, dynamics of the comparative advantage of nations. It also contributes to the literature of international knowledge diffusion by studying the possible drivers of knowledge across borders, using the setting suggested by Bahar et. al. 2014, which uses product-level exports figures as a measure of knowledge acquisition, after controlling for global demand.

The main result in all these settings is that people matter: by serving as international drivers of productive-knowledge, they can shape the comparative advantage of nations. In all of the specifications we include controls for a set of variables that leave us with empirical evidence suggestive that this is the mechanism in place. The instrumental variables approach also reduces possible endogeneity concerns.

This finding is particularly important to understand some known characteristics of knowledge diffusion. First, the short-ranged character of knowledge diffusion can be explained by the fact that part of knowledge is embedded in people, that tend to move in a more localized manner than goods or capital. Second, the fact that diffusion of knowledge and technology is more widespread today than decades ago (i.e. the diffusion process has accelerated over time) can be explained by the fact that people flows, such as migration or short term travel, have also increased rapidly.

The results suggest a process that is observable. In fact, the sample contains examples that are consistent with anecdotical evidence. For instance, the Tanzanian soap industry which benefited from Kenyan migration[12]; as well as paper products exports from Chile which occurred simultaneously to the presence of Chileans in Sweden during Chile's military

[12] http://www.unitedworld-usa.com/usatoday/tanzania/interviews/hemal_shah.htm

dictatorship.

The importance of these results, however, go beyond the pure relationship between migration and exports. It serves as a proof of concept that mobility is a crucial element in the evolution of the comparative advantage of countries and productivity, which is known to be highly explanatory of income and growth.

Chapter 3

Heavier than Air? Knowledge Transmission within the Multinational Firm

3.1 Introduction

About fifty percent of cross-country income variation is explained by differences in productivity.[1] This begs the question: if productivity-inducing knowledge[2] is available in some places, why isn't it available in others? Arrow (1969) suggests that the transmission of knowledge is difficult and costly. These difficulties arise because effective knowledge transmission involves human interaction, which cannot be fully replaced with written words[3] (e.g., even in today's world, business trips have not been fully replaced by emails). A firm, as any other economic agent, also faces difficulties when transferring knowledge among different divisions and affiliates. When a firm operates across borders, different time

[1]e.g., Caselli 2005, Hall and Jones 1999

[2]The Merriam-Webster dictionary defines knowledge as the set of information, understanding, and/or skills that one gets from experience or education.

[3]Knowledge that resides in human minds is usually referred to as tacit (Polanyi 1966). Tacit knowledge is information that cannot be easily explained, embedded or written down.

zones, languages and cultures can raise knowledge transmission costs further. This study contributes to the literature by addressing the effects of knowledge transmission costs on the expansion of multinational corporations (MNCs).

This paper presents two main empirical findings and formalizes them in a theoretical framework. First, MNCs are less likely to horizontally expand their knowledge intensive activities to foreign locations, compared to non knowledge intensive industries. Second, when they do expand, they tend to do so at a shorter geographic distance. Interestingly, however, geographic distance becomes less relevant for horizontal expansion when a firm and its subsidiary are located within the same time zone, and thus able to communicate in real time.

These findings cannot be explained by most theoretical models on MNC fragmentation, which implicitly or explicitly assume zero cost, or costs orthogonal to distance, of transferring knowledge between headquarters and subsidiaries (i.e., Helpman 1984; Markusen 1984; Brainard 1993; Markusen et. al. 1996; Markusen 1997; Carr et. al. 2001; Helpman, Melitz and Yeaple 2004; Keller and Yeaple 2013). A number of empirical studies have tested the validity of these models' predictions, but there has been little or no emphasis on testing the assumption that knowledge transmission is costless.[4]

Thus, to explain the results, I augment the model by Helpman, Melitz and Yeaple (2004) by incorporating the marginal cost of knowledge transfer faced by firms engaged in foreign direct investment (FDI). The augmented model's main assumption is the existence of a marginal cost of knowledge transmission that increases with the level of knowledge intensity and the distance between headquarters and subsidiary. This departs from the traditional view, which assumes that the fixed costs of initial setup are the only costs incurred by the headquarters when creating a foreign affiliate. In reality, the costs of maintaining and interacting with a foreign subsidiary are present throughout the subsidiary's lifetime, and do not end the day the plant is built.

For the empirical analysis, I use a sample derived from the Worldbase dataset by Dun &

[4] e.g., Brainard 1997; Carr et. al. 2001; Markusen and Maskus 2002.

Bradstreet,[5] comprising more than 60,000 foreign subsidiaries of MNCs with information on their physical location and primary economic activity, as defined by the 1987 Standard Industry Classification (SIC). From this dataset, I identify those foreign subsidiaries that represent a horizontal expansion of the associated MNC. Using geocoded location data, I measure the precise distance between each foreign affiliate and its MNC global headquarters. I then compute industry-specific knowledge intensity measures. These indicators reflect the accumulated experience and training of workers in any given industry, using occupational characteristics defined in the O*NET project dataset. I link these indicators to the industry reported by each foreign subsidiary in the dataset. Finally, I exploit variation in the knowledge intensity of subsidiaries to study the model's predictions about the nature of knowledge transmission for horizontal foreign subsidiaries as opposed to the MNC's domestic affiliates as well as non-horizontal foreign subsidiaries.

The data reveals that firms are less likely to have horizontal foreign subsidiaries producing knowledge intensive goods. This result controls for transportation costs, characteristics of the host country relative to the headquarters country, and MNC fixed effects. More specifically, manufacturing industries that are one standard deviation above the knowledge intensity mean are, on average, about 3.6 percentage points less likely to be replicated abroad as a horizontal foreign subsidiary, or 12% based on the actual proportion of foreign affiliates in the sample. For example, a semiconductor manufacturing plant is about 30 percentage points less likely to be replicated abroad than a meat packing plant.

Using the whole portfolio of foreign subsidiaries of a firm, which include both horizontal and non-horizontal subsidiaries, I find that horizontal subsidiaries are characterized by being located at shorter geographic distances to the headquarters, a result that supports the model's assumption that the cost of transferring knowledge to horizontal foreign subsidiaries increases with distance. In light of this result, I explore the relationship between distance and knowledge intensity for horizontal subsidiaries. The assumptions of the model imply

[5]The dataset was privately acquired from D&B and is not publicly accessible. It has been previously used in the literature by Lipsey (1978), and more recently by Harrison et. al. (2004), Black and Strahan (2002), Acemoglu, Johnson & Mitton (2009), Alfaro and Charlton (2009), Alfaro and Chen (2012).

that firms, in order to maximize profits, face a tradeoff that drives them to locate foreign horizontal subsidiaries nearby—especially when they produce a knowledge intensive good. This is supported by the data, which shows a negative partial correlation between knowledge intensity and the distance between a headquarters and its horizontal foreign subsidiaries. For instance, an American MNC with a meat packing horizontal subsidiary located in Turkey would locate its horizontal semiconductors plant in Ireland.

Much of the literature would posit that these results are driven by transportation costs of intermediate goods, which are assumed to be more prevalent for knowledge intensive industries (Irrazabal et. al. 2013; Keller and Yeaple 2013).[6] I find evidence in the data to rule out this mechanism. More specifically, I find that when a headquarters and its subsidiaries are located in the same time zone, distance losses relevance in a firm's decision to expand horizontally. This implies that real-time communication decreases the cost of transferring knowledge by effectively "reducing" the distance between headquarters and subsidiary by two thirds, or by 3500 Km. for the average foreign subsidiary. This suggests that the cost of shipping intermediate goods (which would be just as relevant within the same time zone, because north-south shipping is equally as expensive as east-west shipping), is not the only factor driving a firm's location decisions. Rather, the evidence suggests that the cost of transferring knowledge plays an important role by incentivizing firms to locate their knowledge intensive subsidiaries at shorter distances. Speaking a common language also seems to effectively reduce distance between a headquarters and a subsidiary (though not as much), while the existence of a non-stop commercial flight between a headquarters and its subsidiary does not. This implies that real-time remote interaction and cultural similarities are more important in lowering the cost of knowledge transfer than the ease of face-to-face interaction.

The results show that the cost of knowledge transmission is a determinant of MNC activity. These findings have larger implications for a number of yet-unanswered questions

[6]Keller and Yeaple (2013) assume that knowledge is substitutable with intermediate goods, inducing intra-firm trade in knowledge intensive sectors and thus worsening the performance of distant foreign affiliates.

in economics. For instance, high barriers to knowledge transmission may explain persistent differences in productivity levels between countries and the divergence of their incomes over time (e.g., Pritchett 1997, Hall and Jones 1999), because productivity-inducing knowledge does not diffuse easily.

The rest of the paper is divided as follows. Section 3.2 summarizes the related literature. Section 3.3 outlines a theoretical framework that explores how knowledge transmission can affect MNC decisions and provides guidance for the empirical analysis. Section 3.4 describes the dataset and the construction of relevant variables. Section 3.5 discusses the empirical strategy and presents results and their interpretation, while Section 3.6 concludes and addresses areas for future research regarding the role of knowledge in economic activity.

3.2 Related Literature

The determinants of MNC expansion and fragmentation have been explored in the literature for years.[7] Helpman (1984) suggests vertical fragmentation is motivated by differences in factor abundance between the host and recipient country. Markusen (1984) models the case when horizontal expansion can arise between two identical countries, based on the assumption that a headquarters' activities can be geographically distant from production processes. Brainard (1993) modeled the "proximity-concentration hypothesis," in which both transportation costs and increasing returns play a role in international horizontal expansion of MNCs.

Carr, Markusen and Maskus (2001) –building on Markusen et. al. (1996) and Markusen (1997)– endogenize the vertical and horizontal decisions of firm in what is known as the "knowledge capital model." The model is based on three critical assumptions. First, knowledge-based assets may be fragmented from production; second, knowledge-based assets are skilled labor intensive; and third, the services of knowledge-based assets are (at least partially) joint inputs (i.e., homologous to a public good within the firm) into

[7]See Antras and Yeaple (2013) for recent review on this topic.

multiple production facilities. In this model, vertical fragmentation arises from the first two assumptions, while horizontal expansion is a result of the third one.

Papers such as Brainard (1997), Carr et. al. (2001), Markusen and Maskus (2002) empirically test for the predictions of the above mentioned models, with little emphasis on testing the validity of the zero-cost assumption concerning knowledge transmission.

In the literature on heterogenous firms, Helpman, Melitz and Yeaple (2004) present a model in which horizontal FDI substitutes for exports. In it, a firm's potential profit determines such tradeoff, based on fixed costs and transportation costs. More recently, Keller and Yeaple (2013) augment this model by adding the knowledge component. In their model, once the firm expands horizontally beyond its borders, it faces the tradeoff between creating an upstream plant in that remote location (which locally provides the knowledge) or, alternatively, shipping the knowledge-embedded intermediate good from the headquarters' site (being the main assumption that knowledge can be fully embedded into an intermediate good). The model predicts that a firm will decide to do the latter for knowledge-intensive products. Under this framework, the lower profitability that characterizes distant subsidiaries active in more knowledge intensive industries is explained by intra-firm trade. That is, firms face higher trade costs for knowledge intensive industries given their optimal choice of importing the "ready-to-go" knowledge embedded in intermediate goods from its headquarters. They present empirical evidence supporting this hypothesis.

I introduce tacit knowledge as an additional component to this discussion. Michael Polanyi (1966) referred to tacit knowledge as information that is difficult to transfer: it cannot be easily explained, embedded or written down.[8] Firms possess tacit knowledge both in their specific processes and in the minds of their employees (e.g., Kogut and Zander 1992). It is in the interest of the firm to transfer this knowledge, as efficiently as possible, to all of its subsidiaries. However, the tacit character of this knowledge implies it cannot be embedded in intermediate goods, and that there are difficulties associated with its transmission. If

[8]Others in the management and strategy literature have referred to this type of knowledge as "sticky information" (e.g., Von Hippel 1994; Szulanski 1996, 2002).

these difficulties are large enough, we would expect them to have an impact on the pattern of MNCs' decisions regarding foreign expansion.

Is it reasonable to think that such difficulties exist? In fact, the consensus in the existing literature on the economics of knowledge is that the transmission of knowledge is not immediate, and that knowledge diffusion strongly decays with distance. For instance, the paper by Jaffe, Trajtenberg and Henderson (1993) was among the first to make this claim, showing that patent citations are more frequent within the same geographic area. Bottazi and Peri (2003) followed up using European data. Along the same lines, Keller (2002) showed that knowledge spillovers decrease with distance by looking at productivity changes as explained by foreign R&D investment. He documents that the half-life of such spillovers is 1200Km. More recently, Bahar et. al. (2014) show that a country is 65% more likely to add a new product to its export basket whenever a geographic neighbor is a successful exporter of the same good, a finding that is attributed to the local character of knowledge diffusion.[9]

In this context, this paper aims to contribute to the literature by presenting unexplored evidence on the role tacit knowledge transmission plays in the activity of MNCs, focusing on horizontal expansion.

3.3 Conceptual Framework

In this section I augment the model by Helpman, Melitz and Yeaple (2004) – referred to as HMY hereafter – by including a new parameter capturing the intra-firm cost of transmitting knowledge between headquarters and foreign subsidiaries. This extension allows us to understand how the cost of knowledge transmission faced by firms affects their decision to serve foreign markets. First the common set-up is described and then the proper adaptation is incorporated.

As in HMY, there are N countries producing H+1 sectors with labor as the only input of production. H sectors (indexed 1, 2,...,H) produce a differentiated good, while the other

[9] See Keller (2004) for a review of this literature.

sector (indexed 0) produces an homogenous good (which serves as the numeraire). In any given country, individuals spend a share $\beta_h > 0$ of their income on sector h, such that $\sum_{0 \le h \le H} \beta_h = 1$. Country i is endowed with L^i units of labor and the wage rate in this country is denoted by w^i.

Consider now a particular differentiated sector, h. For simplicity of notation, the index h is dropped in the next few paragraphs, but it is implicit that all sector specific variables may vary across sectors.[10] In order to enter the industry in country i a firm bears a fixed and sunk cost f_E denominated in units of labor. After bearing this cost, the potential entrant learns its labor-per-unit cost, a, drawn from a common and known distribution $G(a)$. Upon observing this cost, the firm may choose not to enter, and thus bear no additional costs and receive no revenues. If it chooses to produce, however, an additional cost of f_D units of labor is incurred. There are no other fixed costs if the firm chooses to produce and sell in the local market only.

The firm can choose to serve a foreign market either by exporting or creating a foreign subsidiary. If the firm chooses to export, it bears an additional cost of f_X (per country it exports to). If it chooses to create a foreign affiliate, it incurs an additional cost of f_I for every foreign market it chooses to serve this way. Similar to HMY, f_X can be interpreted as the cost of forming a local distribution and service network in the foreign market, and f_I includes all of these costs, as well as the cost of forming a subsidiary in the foreign country and the overhead production costs embodied in f_D. Hence, $f_I > f_X > f_D$.

The homogenous good is freely traded at no cost.[11] Differentiated goods that are exported from country i to country j are subject to a "melting-iceberg" transport cost $\tau(t, d_{ij})$ which is an increasing function of the per unit shipping cost of the good (denoted by t, and proxies for weight or other good specific characteristics) and the distance between countries i and j (denoted by d_{ij}). It is assumed that that $\tau(t, d_{ij}) > 1$. That is, a firm in country i has

[10]Some sector-specific variables are explicitly kept in the notation, such as t and k, since these variables will be relevant in the empirical analysis.

[11]Thus, as long as the numeraire good is produced in all countries the wage rate is equalized.

to ship τ units of a good for 1 unit to arrive in country j.

Analogously, serving a foreign market through an affiliate is subject to a marginal cost $\kappa(k, d_{ij})$ related to the transfer of knowledge. $\kappa(k, d_{ij})$ is assumed to be an increasing function of both the knowledge intensity of the good (represented by k) and distance (d_{ij}). The cost of transferring knowledge includes resources and time used for communicating with foreign affiliates to transmit proper knowledge required for efficient production. It is assumed that $\tau(t, d_{ij}) > \kappa(k, d_{ij}) > 1$ for all goods. The last inequality implies that for a multinational corporation, the cost of selling 1 unit of a good through a foreign affiliate is $\kappa(k, d_{ij})$.

The cost of knowledge transmission in knowledge intensive sectors being higher is justified given that these sectors require higher interaction and communication among their workers. Thus, firms pay for business travel and communication services that occur more often within these sectors. In addition, and perhaps more importantly, knowledge intensive activities usually encompass tasks with higher probability of failure and thus requiring trained and experienced workers. This too raises operational costs.

Assuming that knowledge transmission costs are increasing in distance is consistent with empirical evidence (e.g., Jaffe, Trajtenberg and Henderson 1993, Bottazi and Peri 2001, Keller 2002, Keller 2004, Bahar et. al. 2014). This evidence is reviewed in the previous section.

All the producers which serve a market engage in monopolistic competition. Consumer preferences across varieties of a differentiated product h have the standard CES form, with an elasticity of substitution $\varepsilon = \frac{1}{1-\alpha} > 1$. It is well known that these preferences generate a demand function $A^i p^{-\varepsilon}$ for each product in the industry in country i, where $A^i = \frac{\beta}{\int_0^{n^i} p^i(s)^{1-\varepsilon} ds} E^i$, n^i is the measure of firms active in the industry in country i, and $p^i(s)$ is the consumer price for a product indexed s.

In this setting, an active producer with labor requirement of a optimally sets a price of $\frac{w^i a}{\alpha}$. Consequently, the price of a locally produced good is $\frac{w^i a}{\alpha}$, the price of a good which is exported to country j is $\frac{\tau(t,d_{ij}) w^j a}{\alpha}$, and the price of a good that is sold by a foreign affiliate in country j is $\frac{\kappa(k,d_{ij}) w^j a}{\alpha}$, where a is the labor required for the producer to manufacture one unit of the product.

In what follows, it is shown that the balance of forces ruling the tradeoff of serving a foreign market through exports or FDI is influenced by the knowledge intensity of the product.

The assumption that the numeraire good is produced in each country simplifies the analysis, as it implies that the wage rate is equalized across all countries and is equal to 1. Hence, the operating profit for a firm in country i with a labor coefficient of a from serving the domestic market maybe expressed as $\pi^i_D = a^{1-\varepsilon}B^i - f_D$, where $B^i = (1-\alpha)\,\frac{A^i}{\alpha^{1-\varepsilon}}$. The additional profits from exporting to country j are $\pi^i_X = \left(\tau(t, d_{ij}) \cdot a\right)^{1-\varepsilon} B^j - f_X$ and those from selling in country j through a foreign affiliate are $\pi^i_I = \left(\kappa(k, d_{ij}) \cdot a\right)^{1-\varepsilon} B^j - f_I$. B^i represents demand parameters for country i and are considered exogenous to each individual firm.

Hence, in this setting, the productivity parameter a will be critical for a firm's decision of whether to serve the local market only or to serve foreign markets, either through exports or FDI. The sorting pattern is similar to the one in HMY and is based in the following equations:

$$(a_D)^{1-\varepsilon} \cdot B^i = f_D, \qquad \forall i \tag{3.1}$$

$$\left(\tau(t, d_{ij}) \cdot a_X\right)^{1-\varepsilon} \cdot B^j = f_X, \quad \forall i, \forall j \neq i \tag{3.2}$$

$$\left[\kappa(k, d_{ij})^{1-\varepsilon} - \tau(t, d_{ij})^{1-\varepsilon}\right] \cdot a_I^{1-\varepsilon} \cdot B^j = f_I - f_X, \quad \forall i, \forall j \neq i \tag{3.3}$$

Similar to HMY, the first two equations define the productivity thresholds after which firms will sell domestically or export, respectively. The minimum productivity threshold after which firms will engage in FDI is derived from Equation (3.3).[12] This threshold is defined as:

[12] Condition (3.3) will have a positive solution if we assume $\kappa(k, d_{ij})^{\varepsilon-1} f_I > \tau(t, d_{ij})^{\varepsilon-1} f_X > f_D$, which is homologous to condition (1) in HMY (with equal wages across countries), but including κ.

$$a_I^{1-\varepsilon} = \frac{f_I - f_X}{\left[\left(\kappa(k, d_{ij})\right)^{1-\varepsilon} - \left(\tau(t, d_{ij})\right)^{1-\varepsilon}\right] B^j}, \forall i, \forall j \neq i \tag{3.4}$$

Predictions derived from this model will serve as the basis for the empirical analysis. The implications of the original HMY model are straightforward. An increase in $\tau(t, d_{ij})$, either through an increase in either t or d_{ij}, will result in lower π_E making it more likely to substitute exports with FDI. This is part of the mechanism of the concentration-proximity tradeoff. However, with the inclusion of $\kappa(k, d)$ into the model, some new predictions arise, assuming full symmetry in fixed costs and demand variables for all sectors and countries. The propositions are presented in terms of $\phi(a_I) = a_I^{1-\varepsilon}$.

Proposition 1 *As k increases, the profitable FDI threshold ($a_I^{1-\varepsilon}$) increases, implying fewer firms will substitute exports towards FDI.*

$$\frac{\partial \phi(a_I)}{\partial k} = \frac{\partial \phi(a_I)}{\partial \kappa} \cdot \frac{\partial \kappa}{\partial k} > 0 \tag{3.5}$$

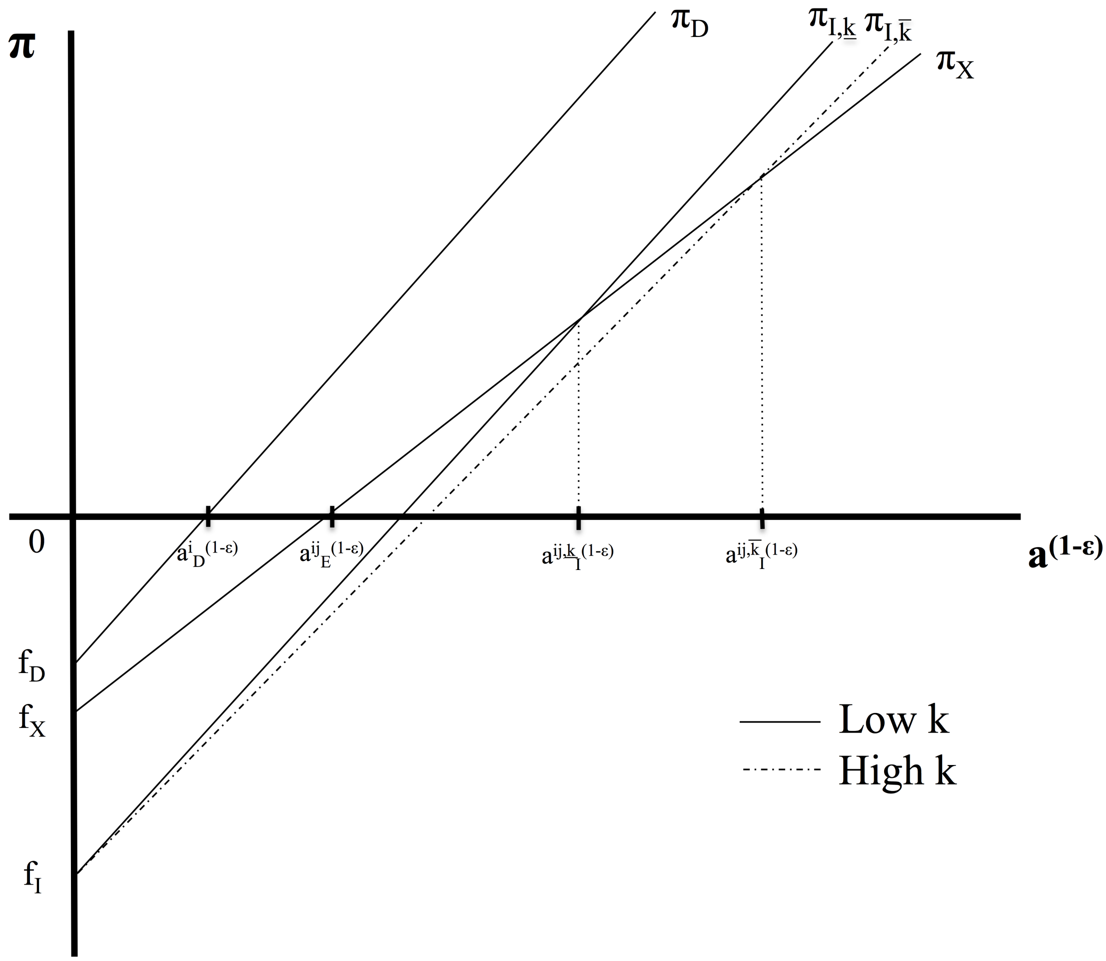
Proposition (1) is a direct consequence of adding κ into the model. Thus, *ceteris paribus*, FDI will be less likely for sectors with higher k. The graphical representation of the model in Figure 3.1 shows the case for two sectors that differ in their knowledge intensity, $\underline{k}$ and $\overline{k}$ (where $\overline{k} > \underline{k}$). Notice that the profit functions for both sectors originate in the same fixed cost value f_I, but the function is flatter for the sector $\overline{k}$ (dashed line). Hence, the productivity threshold required for a firm to substitute exports with FDI becomes higher for sectors with higher levels of k. That is, $(a_I^{ij,\overline{k}})^{1-\varepsilon} > (a_I^{ij,\underline{k}})^{1-\varepsilon}$.

Proposition 2 *As d increases, the change in $a_I^{1-\varepsilon}$ is ambiguous.*

$$\frac{\partial \phi(a_I)}{\partial d} = \frac{\partial \phi(a_I)}{\partial \tau} \cdot \frac{\partial \tau}{\partial d} + \frac{\partial \phi(a_I)}{\partial \kappa} \cdot \frac{\partial \kappa}{\partial d} = \begin{cases} \geq 0, & \left[\frac{\tau(t,d)}{\kappa(k,d)}\right]^{\varepsilon-1} \geq \frac{\epsilon_{\tau,d}}{\epsilon_{\kappa,d}} \\ < 0, & otherwise \end{cases} \tag{3.6}$$

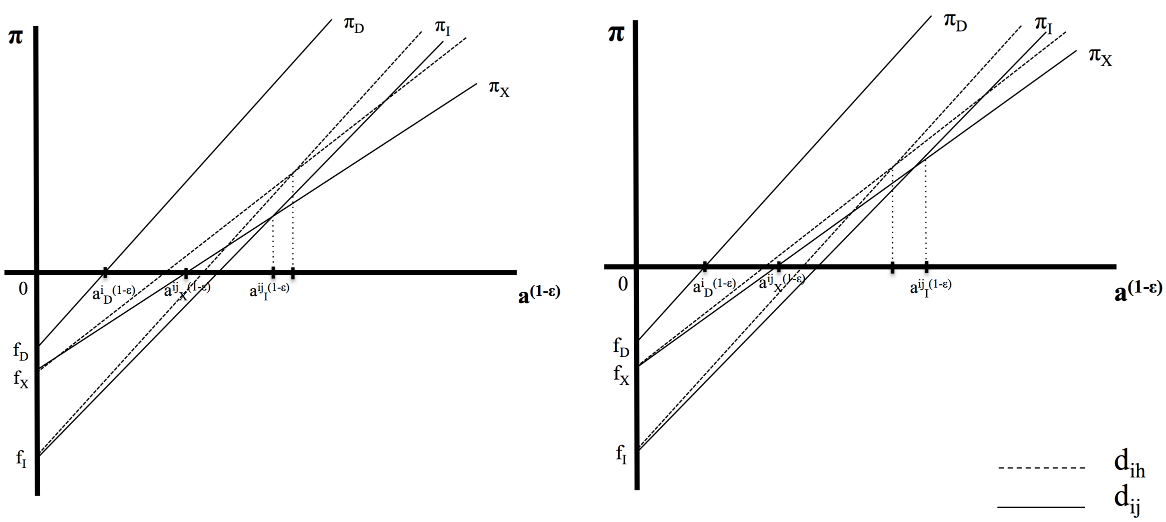
To understand Proposition (2) suppose there are two foreign countries h and j such that $d_{ij} > d_{ih}$. In the original HMY model, longer distances will reduce π_E hence making it *always* more profitable for a given level of a to engage in FDI instead of exports. However,

Figure 3.1: *Increase in k (knowledge intensity)*



Graphical representation of the model, for a case considering two sectors with different levels of k, where $\overline{k} > \underline{k}$. The result suggests that the threshold a_I is an increasing function of k. Thus, FDI will be less likely for sectors with higher k.

Figure 3.2: *Increase in d (distance)*



Graphical representation of the model, for a case considering a firm serving two foreign markets h and j, where $d_{ij} > d_{ih}$. The left panel shows the case where the threshold a_I is a decreasing function of d, while the right panel shows the case where the threshold a_I is an increasing function of d. The case in the left panel assumes that π_E is more elastic to changes in distance than π_I, while the case in the right panel assumes otherwise.

with the inclusion of κ in the model, longer distances will reduce both π_E and π_I. Thus, the equilibrium point can shift either way, depending on the elasticity of profits with respect to distance. The left panel of Figure 3.2 shows the case when $(a_I^{ih})^{1-\varepsilon} > (a_I^{ij})^{1-\varepsilon}$. Intuitively this happens whenever π_E is more elastic to changes in distance than π_I (or given the condition stated in Equation (3.6), where ϵ represents elasticity; see Appendix Section C.1 for more details on this condition). This case is qualitatively the same result as in the HMY model. The right panel of Figure 3.2, however, shows another possibility. In it $(a_I^{ih})^{1-\varepsilon} < (a_I^{ij})^{1-\varepsilon}$. In this case, FDI will be less profitable for longer distances hence resulting in *fewer* firms substituting exports with FDI.

The predictions coming out of the model following the inclusion of an intra-firm cost of transferring knowledge (κ) have testable implications in the data. First, *ceteris paribus*, industries with higher levels of knowledge intensity will be *less likely* to expand horizontally to foreign destinations. Second, horizontal expansion will be *less likely* in foreign locations that are located at longer distances under certain conditions. Regarding Proposition (2), given there are no empirical priors on whether the stated condition holds, letting the data speak will provide guidance on the assumptions of the developed model. That is, if horizontal FDI correlates negatively with longer distances, then there is empirical support to assume that ${}^{\partial\kappa}/_{\partial d} > 0$.

The next section presents the sample and the variables used to perform the empirical analysis.

3.4 Data and Definitions of Variables

3.4.1 Worldbase dataset by Dun & Bradstreet

This paper uses the Worldbase dataset by Dun & Bradstreet (from May 2012) as its main data source. The dataset has information on more than one hundred million establishments worldwide. Each establishment is uniquely identified and linked to its global headquarters (referred to as the "global ultimate"). For this study I focus on foreign plants engaged in manufacturing industries (SIC codes 2000 to 3999) owned by MNCs. As suggested by Caves (1971), an MNC is "an enterprise that controls and manages production establishments – plants – located in at least two countries."[13]

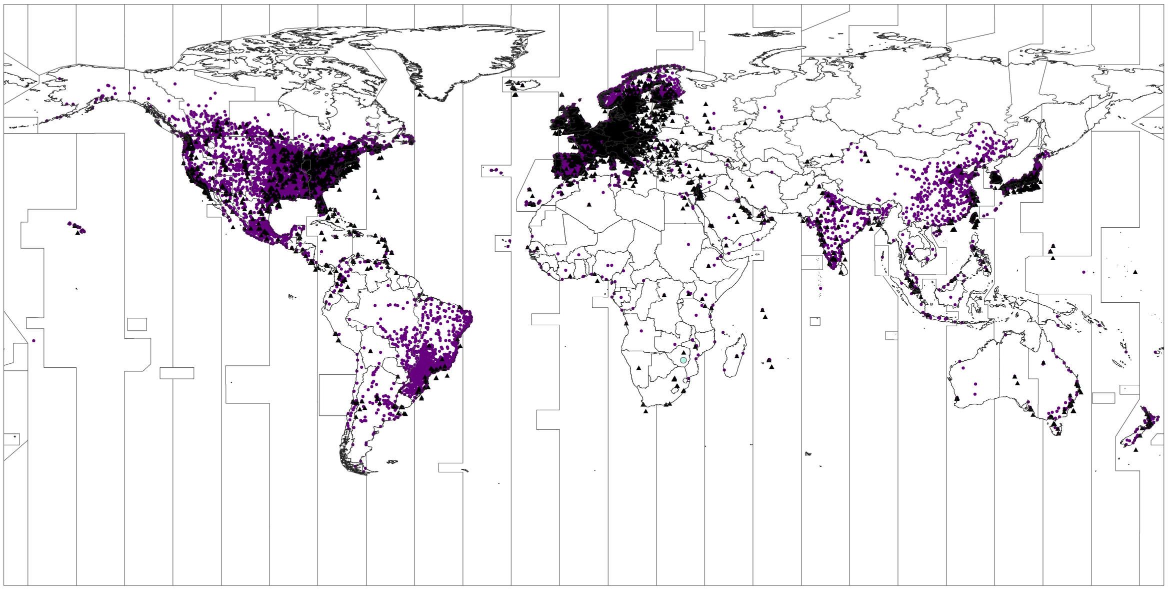
Two different samples are obtained from the dataset. The first one, uses both domestic and horizontal foreign subsidiaries of MNCs. The second one, exclusively uses the complete portfolio of foreign subsidiaries of MNCs, which include horizontal and non-horizontal subsidiaries.

The first sample includes about 64,462 subsidiaries, both domestic and foreign (the latter defined as being in a different country than their global ultimate). The second sample consists of 60,621 foreign subsidiaries. Overall, headquarters are scattered across 89 countries while subsidiaries are in over 100 countries.

For the analysis I will use the reported main SIC code as the only indicator of a plant's

[13]I exclude MNCs for which 99% of their subsidiaries or employees are in the home country, besides them having plants in two or more countries. This drops a small number of Chinese MNCs with one or two subsidiaries in Hong Kong and the rest in China.

Figure 3.3: *Unique locations of headquarters and subsidiaries*



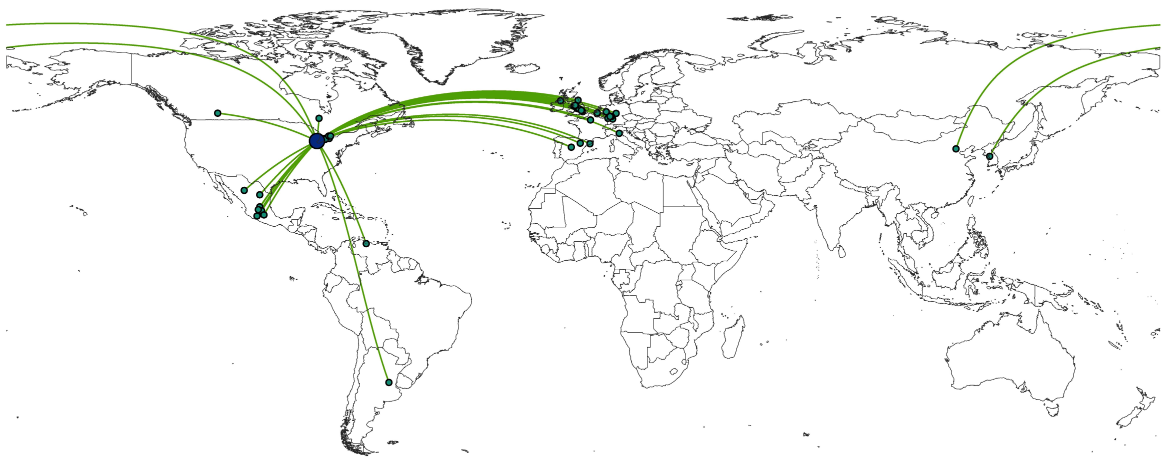
The figure shows a World map with the geocoded location of all the headquarters (triangles) and foreign subsidiaries (dots) in the sample.

economic activity. There are about 450 unique SIC 4-digit codes (in manufacturing) reported by subsidiaries as their main economic activity in the dataset (see Appendix Section C.2 for more details on this).

In order to obtain the precise location of each plant I geocode the dataset using Google Maps Geocoding API to find the exact latitude and longitude of its headquarters and each one of its foreign subsidiaries. With this I computed the exact distance between each headquarters and its foreign subsidiaries. Figure 3.3 maps the unique locations of all foreign subsidiaries (dots) and headquarters (triangles) in the sample.

For instance, Figure 3.4 shows the headquarters and subsidiaries of an American car manufacturing multinational firm. The firm, headquartered in the US, has a number of foreign subsidiaries on different continents. The lines originating from the headquarters represent the geographic distance to each subsidiary.

Figure 3.4: *Headquarters and foreign subsidiaries of an American MNC*



The figure is an example of the resolution of the data. It shows a World map with the geocoded location of the headquarters of an American car manufacturing firm and all of its subsidiaries.

3.4.2 Definitions of Variables

Horizontal Foreign Subsidiary

I define a foreign subsidiary as a horizontal expansion based on its SIC code vis-á-vis all the SIC codes reported by the firm, in all of its domestic subsidiaries in the home country. This resolves the data issues that arise when the economic activity of the headquarters does not necessarily represent the main business of the firm. For instance, in the dataset, the headquarters of a well known worldwide multinational in the cosmetic world is defined under SIC code 6719 ("holding company"). However, many of its domestic subsidiaries are classified under SIC code 2844 (perfumes, cosmetics, and other toiletries), which would be a more natural classification for the firm as a whole. Hence, by limiting the definitions to the global ultimate's SIC category only, horizontal relationships would be underestimated.

Following the methodology used by Alfaro and Charlton (2009) I exclude from the definition of horizontal expansions those foreign subsidiaries that fall in both horizontal and vertical classification (see Appendix Section C.3 for more details).

When limiting the sample to domestic and foreign horizontal affiliates only, the latter are about 29% of all plants. When looking at the broader foreign affiliates portfolio of a MNC, (which includes all types of foreign subsidiaries, and excludes domestic subsidiaries), around 34% of all foreign subsidiaries are classified as horizontal expansions. Of the remaining 64% of non-horizontal affiliates, only a handful can be classified as vertical foreign subsidiaries,[14] while the majority are subsidiaries classified in industries that are unrelated to the firms' core business, as measured by the sectors the firm is producing at home.[15]

Knowledge Intensity Measures

In order to estimate the knowledge intensity of industries I construct indicators that measure the accumulated experience and training required for optimal performance of the different occupations associated with each industry. These measures attempt to proxy for the knowledge parameter k referred to in the theoretical framework above.

Knowledge is defined as the set of information, skills and understanding that one acquires through experience and education. The tacit component of knowledge is the one that resides mostly in people's brains, and cannot be codified. Thus, in order to quantify the intensity of the tacit knowledge that characterizes an specific industry I compute the average experience and training of that industry's representative workforce. This differs from other measures that would capture only the codified component of knowledge such as patent counts or years of schooling of workers. To the best of my knowledge, these are the first measures that attempt to capture the tacit knowledge intensity of an industry.

To construct the knowledge intensity measures I use data from the Occupational Employment Statistics (OES) from the Bureau of Labor Statistics,[16] and occupational profiles

[14] 3,062 observations can be classified as vertical foreign subsidiaries, while 8,108 are classified as "complex", implying they fall in both horizontal and vertical categories. These "complex" subsidiaries are considered neither horizontal nor vertical. Appendix Section C.3 expands on this discussion.

[15] This is an interesting finding in and of itself, and is also noted by Alfaro and Charlton (2009). While attempting to explain this finding is out of the scope of this paper, it is a part of the future research agenda.

[16] Data from 2011, downloadable from ftp://ftp.bls.gov/pub/special.requests/oes/oesm11in4.zip

compiled by the Occupational Information Network (O*NET) project.[17] OES breaks down the composition of occupations for each industry code,[18] based on a list of about 800 occupations. These occupations can be linked to occupational profiles generated by O*NET, which includes results from a large number of survey questions on the characteristics of each occupation.

The relevant questions in the survey that capture the learning component of the workers, as mentioned above, are the ones related to experience and training. The exact form of the questions from O*NET are:

- How much related experience (in months) would be required to be hired to perform this job?
- How much "on-site" or "in-plant" training (in months) would be required to be hired to perform this job?
- How much "on-the-job" training (in months) would be required to be hired to perform this job?

Using these questions I generate the main knowledge intensity measure that I will be using in the empirical analysis section.[19] The measure, which I refer to it as *"Experience plus training"* throughout the paper, is constructed by measuring the (wage-weighted) average months of experience plus on-site and on-the-job training required to work in each industry.

Using this measure, industries related to legal, financial and engineering services rank highly in the list among the knowledge intensive industries. In the manufacturing sector,

---

[17] O*NET is the successor of the US Department of Labor's Dictionary of Occupational Titles (DOT). I use the O*NET database version 17, downloadable from http://www.onetcenter.org/download/database?d=db_17_0.zip. Costinot et. al. (2011) also use O*NET to create an industry level measure of task routineness for 77 sectors. Keller and Yeaple (2013) also present results making use of knowledge intensity variables constructed with O*NET in the web appendix.

[18] I used Pierce & Schott (2012) concordance tables to convert industry codes from NAICS to 1987 SIC. The concordance table is downloadable from http://faculty.som.yale.edu/peterschott/files/research/data/appendix_files_20111004.zip.

[19] Appendix Section C.6.2 presents robustness tests of the empirical analysis using a measure averaging the experience indicators only (excluding the training indicators).

Table 3.1: *KI Measures Correlations*

Variables	Experience + Training	R&D share (N&T)	R&D share (K&Y)
Experience + Training	1.000		
R&D share (N&T)	0.354	1.000	
R&D share (K&Y)	0.420	0.682	1.000

The table shows the Pearson correlation coefficients between the O*NET based measures of knowledge intensity and R&D share in sales, used previously in the literature as proxies of knowledge intensity by Nunn and Trefler (2008) and Keller and Yeaple (2013).

industries ranking highly are computer related (SIC 3573, 3571 and 3572), communications equipment (SIC 3669, 3663 and 3661) and electronics and semiconductors (SIC 3672, 3674 and 3676). Appendix Section C.5 expands on this discussion.

One limitation of this measure is that it is based on US data. Full precision would require to compute these weighted averages using data on occupations per industry for each country separately. However, this data is unavailable, and I will assume the ranking in the knowledge intensity of industries based on US data proxies that of the rest of the world.

I find that this measure correlates positively with other knowledge intensity measures used in the literature, such as the average R&D share of sales per industry (e.g., Nunn & Trefler 2008; Keller & Yeaple 2013), as evidenced in Table 3.1.[20]

The R&D based measures, however, have three main shortcomings that could generate significant biases. First, these measures assign a zero value to about half of the industries, because most firms within those industries have no R&D investment whatsoever. For these industries in the lower end of the distribution, the intensity of their knowledge is indistinguishable.[21] Second, since these measures are computed by averaging across each industry the R&D share of sales reported by a (random or not) sample of firms, they are likely to favor industries in which larger firms are more prevalent. This might happen

[20]It also correlate positively with other measures that could proxy for knowledge intensity or complexity. The correlation coefficient with the share of non-production workers in total employment, from the NBER-CES Manufacturing Industry Database (Becker et. al. 2013), is 0.68. Similarly, the correlation coefficient with the Product Complexity Index, developed by Hausmann et. al. (2011), is 0.49.

[21]See Appendix Section C.4.

in industries for which the barriers to entry are higher, and not necessarily knowledge intensive industries. Third, R&D investment might not be equally accounted for across all industries.

The O*NET based measures solve these issues. Their distribution is smoother (see Appendix Section C.5), they do not rely on a sampling of firms, and they use the same standardized measure for all industries. Hence, I use these indicators as the main proxies for knowledge intensity throughout the paper.

Unit shipping costs

Unit shipping costs for SIC manufacturing industries are computed using data from Bernard, Jensen & Schott (2006).[22] This industry-level measure aims to proxy for t, referred to in the theoretical framework as the unit shipping cost variable, which accounts for how costly it is to transport one unit of that good irrespective of industry. For instance, goods with the highest unit shipping costs in the dataset include ready-mixed concrete and ice, which require special forms of transportation.

The variable measures the amount of US dollars required to transport 1$ worth of a good per every 100Km. It is computed by averaging the same measure per industry across all countries exporting to the US in year 2005. To deal with long tails, this variable will be used in a logarithmic scale in all the different empirical specifications.

Ease of Communication Proxies

In order to proxy for the ease of communication between a subsidiary and its headquarters, I use three variables: non-stop flights, working hours overlap and common language.

The first variable is used because the existence of non-stop flights would proxy for the ease of managers and workers to do more frequent business trips, given the convenience of a direct flight. Business trips, by allowing face-to-face interaction, would facilitate the

[22] Downloadable from http://faculty.som.yale.edu/peterschott/files/research/data/xm_sic87_72_105_20120424.zip

transmission of tacit knowledge. However, it is important to note that business trips, even if convenient, happen much less often than phone calls due to the elevated costs associated with them. In order to compute the existence of a non-stop air route between a headquarters and its subsidiary, I matched all the existing airports within a 100Km radius (conditional on being in the same country), using the geocoded latitude and longitude. The data for airports (with their respective coordinates) and active air routes come from OpenFlight.com.[23] Through this matching I create a dummy variable which takes the value of 1 if there is a non-stop flight between the headquarters and its subsidiary.[24]

The second variable, overlap in working hours, aims to capture the "real-time" communication ability between managers and workers in the two plants. Being in the same time zone allows workers to use phone or videoconference communication more frequently (substituting partially for means such as fax or email). This allows for better transmission of tacit knowledge, which is valuable for troubleshooting or crisis solving. In order to compute the overlap in working hours I use the geocoded longitude of each subsidiary to find its time zone, and compare it to that of its headquarters. Assuming that working hours run from 8:00am to 6:00pm (10 hours in total), the variable measures, for a single day, the number of hours that overlap in the working schedule of both the headquarters and its subsidiary.

Finally, a common language captures cultural proximity, and also better ability to communicate between workers in both locations of the same firm. The common language comes from CEPII's GeoDist database (Mayer and Zignago, 2011). Two countries have a common language if at least 8% of the population in both countries speak such language.

The sample is merged with data that proxies for the ease of communication between the headquarters and its foreign subsidiaries: the existence of a non-stop air route between their nearby airports, the number of overlapping working hours in a given day and whether there is a common language spoken in their respective countries.

---

[23] Data downloadable from http://openflights.org/data.html. Downloaded in June 2013.

[24] I also compute the minimum number of non-stop flights required to travel between two given airports by using the shortest path algorithm. The results using this measure, however, are qualitatively the same as the ones that use the non-stop flight dummy. Thus, this measure is omitted in the analysis.

3.5 Empirical Analysis

This section first discusses the broad empirical strategy and then presents descriptive statistics from the sample. The following subsection presents results of the empirical analysis that test the propositions presented in the theoretical framework section. The remaining subsections present additional evidence consistent with the assumption that the barriers of transferring knowledge are higher for longer geographic distances.

3.5.1 Empirical Strategy

The conceptual framework outlined above is useful to understand the determinants of horizontal expansion for MNCs. The empirical section focuses on understanding the role of knowledge in particular.

In spite of the lack of firm-level export data in the sample, I test for the implications of the model through *reduced forms* that look at the determinants of horizontal FDI, as compared to both domestic subsidiaries and non-horizontal foreign subsidiaries.

The first empirical exercise will deal with testing Proposition (1) of the conceptual framework: are knowledge intensive activities less likely to be replicated abroad? To do so, I will look at a sample of domestic and foreign (horizontal) subsidiaries, and estimate the likelihood of an industry being replicated abroad given its knowledge intensity.

In order to test Proposition (2), that is, whether longer distances makes the knowledge transmission process more costly for firms, I rely on the complete portfolio of foreign affiliates of the MNCs in the sample. Thus, the question the empirical specification asks is: conditional on having a foreign plant in a given industry and location, is it likely to be an horizontal one, given its distance to the headquarters and the knowledge intensity of its economics activity? More broadly, the goal of the exercise is to test whether the patterns for horizontal subsidiaries in the data are consistent with the mechanisms described in the model. The underlying assumptions for this analysis to serve as proof of the raised question are described in the next section.

Following this, I relax the assumptions of the previous analysis, and use only the horizontal foreign subsidiaries to test a deviation of the conceptual framework. More specifically, I test for a negative correlation between distance and knowledge intensity, which follows the model's prediction.

It is important to clarify that this exercise does not substitute for, nor it intends to, using firm-level export data as part of the identification strategy. Yet, the exercise adds value by presenting stylized facts that are robust, and at the same time, consistent with the mechanisms of the conceptual framework with respect to the role of knowledge transmission in explaining the existence and location of horizontal subsidiaries of a MNC.

3.5.2 Descriptive statistics

This section provides descriptive details about the sample, in terms of the distribution of foreign affiliates across regions of the world and developing vs. developed countries.

Table 3.2 presents descriptive statics which compare domestic to horizontal foreign subsidiaries in the sample. This sample includes domestic subsidiaries and foreign subsidiaries that replicate production abroad (i.e. an horizontal expansion). In total, there are 64,462 subsidiaries that are owed by 1540 MNCs. Domestic subsidiaries tend to be more numerous than foreign ones (on average, 29% of these subsidiaries are foreign). The table also includes the knowledge intensity variable measured in standard deviations from the mean (denoted by KI), averaged over domestic and over foreign subsidiaries. The last column presents the difference, with stars denoting the correspondent p-value level.

As it can be seen, on average, industries of the foreign subsidiaries are roughly as half as knowledge-intensive as the industries manufactured by their domestic counterparts. The same pattern holds for all presented cuts of the data, besides for few firms based on non-OECD countries (with a p-value of 0.10), and for few firms based on Western Europe, for which the difference is not statistically significant. This statistic is consistent with Proposition (1) of the conceptual framework. I use this sample in the next subsection to analyze this proposition further.

Table 3.3 summarizes the number of records in the sample that includes only foreign

Table 3.2: *Descriptive Statistics (Domestic Vs. Foreign Subsidiaries)*

	MNC	# Subs	Foreign (%)	$KI_{Foreign}$	$KI_{Domestic}$	Δ
All Observations	1540	64462	.29	.19	.33	$-.14^{***}$
Non OECD	28	958	.12	.42	.34	$.086^{*}$
OECD	1512	63504	.29	.19	.33	$-.14^{***}$
East Asia & Pacific	306	17008	.11	.37	.46	$-.087^{***}$
Latin America & Caribbean	18	1920	.58	-.38	-.3	$-.083^{***}$
North America	508	24891	.2	.24	.33	$-.089^{***}$
South Asia	15	370	.16	.17	.29	$-.12^{*}$
Western Europe	693	20273	.51	.2	.18	.014

The table presents descriptive statistics from the sample. It presents for different cuts of the sample, based on the home country of the MNC, the total number of MNC firms, the number of subsidiaries, the proportion of those subsidiaries that are foreign (horizontal) subsidiaries, the average knowledge intensity of the foreign subsidiaries, the average knowledge intensity for the domestic subsidiaries, and the difference between these averages, denoted by Δ. Stars represent statistical significance of the difference: $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

subsidiaries, both horizontal and non-horizontal. Overall, there are 8,266 MNC firms which have 60,621 foreign subsidiaries. Of those subsidiaries, 34% are defined as horizontal expansions, while the rest could be vertical subsidiaries or, simply, a foreign subsidiary in a non-related industry[25]. The average distance in the sample between headquarters and subsidiaries is 5,152Km. Regarding communication proxies, a subsidiary and its headquarters overlap, on average, 7.3 working hours in a given day, and for about 25% of subsidiaries there exists a commercial non-stop flight from their headquarters. The following rows present the same statistics across different cuts of the sample, based on the headquarters' country. For instance, most of the foreign subsidiaries are located in OECD countries (49,936 vs. 10,685). Similarly, the table shows that most of the foreign subsidiaries in the sample are located in Western Europe and North America.

I also present results of the distribution of sectors among foreign affiliates, to understand whether in the sample there are some sectors that are more likely to appear (i.e. be reported) than others. In terms of industries, the distribution of different sectors in the sample is not homogenous, as can be seen in Figure 3.5. Some sectors are more prevalent than others in the data. The industries that appear the most in the data are Ready-Mixed Concrete

[25]For instance, this could be the result of a MNC diversifying its portfolio by acquiring foreign firms.

Table 3.3: *Descriptive Statistics (Foreign Subsidiaries)*

	MNC	# Subs	H %	Dist	WH	DF
All Observations	8266	60621	34	5152	7.3	.25
Non OECD	2590	10685	32	7697	7.5	.2
OECD	6520	49936	35	4608	7.3	.27
East Asia & Pacific	2074	5560	26	7329	6.1	.24
Eastern Europe	68	125	30	1738	9.2	.2
Latin America & Caribbean	981	7394	36	8453	7.5	.093
Middle East & N. Africa	67	93	40	7549	7.6	.19
North America	2246	16944	37	6698	6.1	.02
South Asia	410	2405	45	7877	6.4	.2
Sub-Saharan Africa	33	51	65	7832	7.7	.098
Western Europe	4969	28049	33	2686	8.3	.44

The table presents descriptive statistics from the sample. It presents for different cuts of the sample the total number of MNC, foreign subsidiaries (Sub), the percentage of subsidiaries classified as horizontal expansion (H%), the average distance in kilometers between subsidiaries and headquarters (Dist), the average number of overlapping working hours between the subsidiaries and the headquarters (WH) and the proportion of subsidiary-headquarter pairs that have a direct flight in between them (DF).

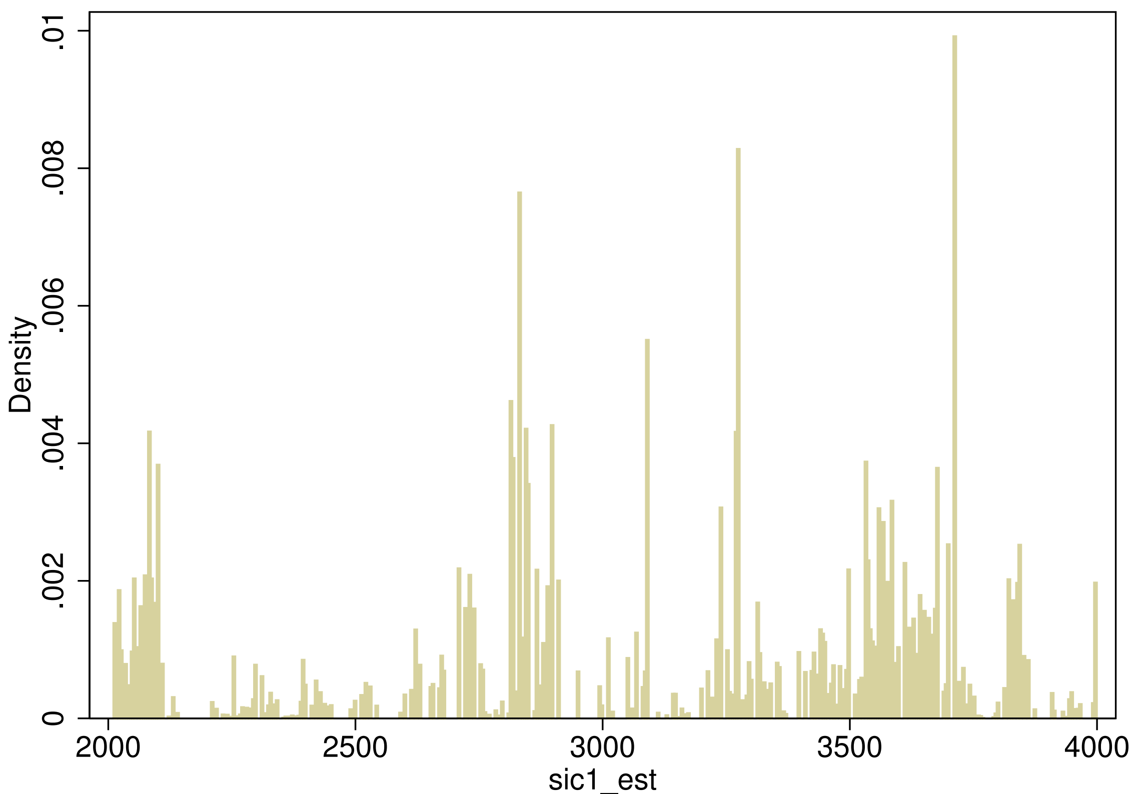
(SIC 3273), Pharmaceutical Preparations (SIC 2834) and Motor Vehicles Parts (SIC 3714). To alleviate concerns on how this distribution could affect the results, all the standard deviations calculations allow for clustering at the industry level.

In addition, it is worth emphasizing that each foreign subsidiary in the sample manufactures a specific product. Hence, if a MNC has several foreign subsidiaries, then each one of those could be manufacturing a different product (in its 4 digit classification). The sample that a single MNC that has more than one foreign subsidiary could be manufacturing more than one product. Figure 3.6 shows that larger MNCs (as measured by number of affiliates) tend to make a larger number of different products.

Notes on the Reliability of the Data

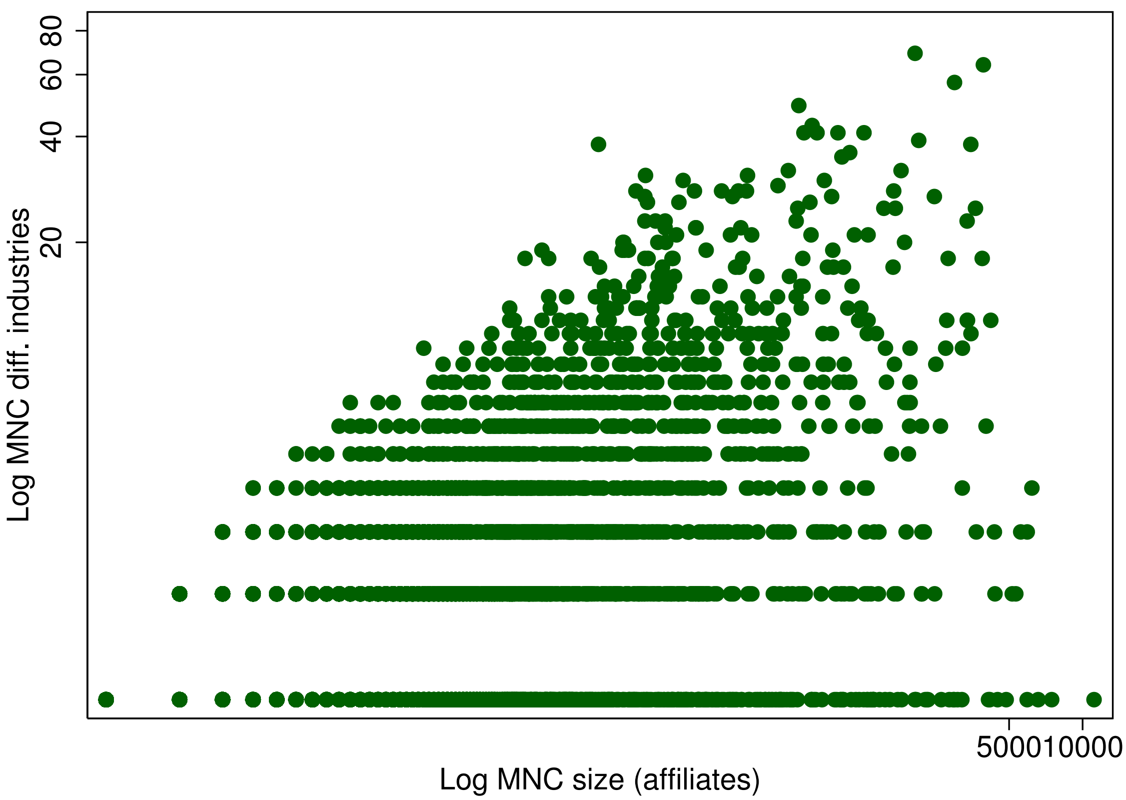
The Worldbase dataset collected by Dun & Bradstreet is sourced from a number of reliable organizations all over the world, including public registries. According to Dun & Bradstreet's website, "the data undergoes a thorough quality assurance process to ensure that our

Figure 3.5: *Histogram of SIC codes in the sample*



The figure is an histogram of the SIC industries reported in the dataset. Each bin represents the frequency of a particular SIC code within the manufacturing sector. Notice that the SIC classification is not fully continuos, what explains the zero values in the figure.

Figure 3.6: *Number of different industries Vs. MNC size*



The figure plots the relationship between MNC size and total number of (different) industries the MNC is active in through its foreign affiliates. The figure reveals that larger MNCs (measured in terms of number of subsidiaries) tend to make a larger number of different products.

customers receive the most up-to-date and comprehensive data available".[26] However, it is important to acknowledge that, given the lack of access to public registries for every country, it is not possible to asses with full accuracy the representativeness of the data. Alfaro and Charlton (2009) compare the dataset with the US multinational firms sample by the US Bureau of Economic Analysis, and find consistencies between the two datasets. Moreover, the regional breakdown of foreign subsidiaries presented in Table 3.3 below seems to be consistent with aggregate figures of FDI inflows across world regions.

Some basic relationships drawn from the sample also behave as expected. For instance, the number of countries in which an MNC has foreign affiliates is related to the overall size of the MNC. Figure 3.7 presents the relationship between the size of firms (in number of establishments in the left panel, and in total number of employees in the right panel[27]) against the number of foreign countries in which their subsidiaries are located (on the vertical axis). Each observation in the scatterplot is an MNC labeled with its headquarters' country ISO3 code. The figure shows smaller MNCs are present in fewer countries, while larger MNCs tend to be more spread out in terms of the number of countries they have a presence in.

Focusing the analysis on the within-firm dimension significantly diminishes the sampling concerns further. This is because, while methods for gathering information may not be symmetric across countries, they would not systematically differ by firm or by industry. The per-country likelihood of missing data would be the same for all firms and industries, controlling for the location of the MNC. Thus, concerns regarding biases caused by possible sampling asymmetries are not particularly large for the purpose of this empirical exercise.

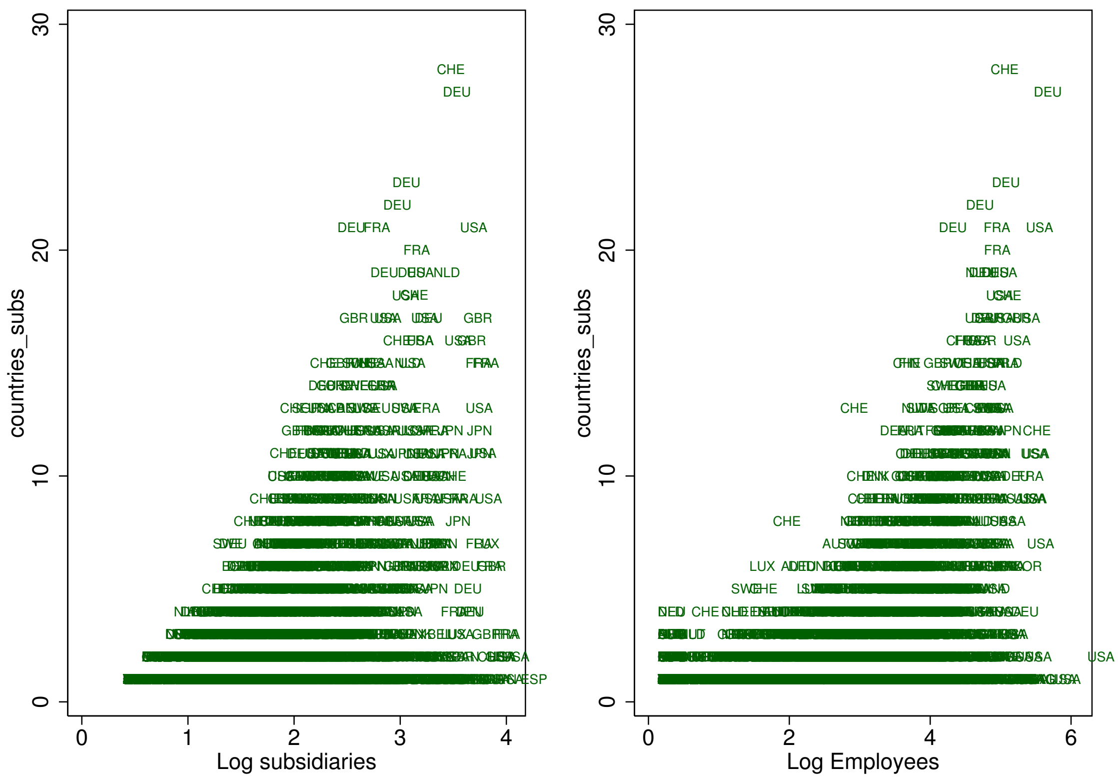
3.5.3 Knowledge intensity as a determinant of horizontal expansion

The first empirical exercise deals with understanding the determinants of horizontal expansion, with guidance of the theoretical model outlined above. It uses the sample that includes

[26]http://dnb.com.au/Credit_Reporting/The_quality_of_DandBs_data/index.aspx

[27]Including their domestic plants for both.

Figure 3.7: *MNC size vs. number of countries*



The figure shows the relationship between the size of MNC (horizontal axis) and the number of foreign countries they are active in (vertical axis). In the scatterplots, each observation is an MNC, labeled with the ISO3 code of the country where its headquarters is located. The left panel measures the firms' size by the total number of subsidiaries it has (both domestic and foreign), while the right panel uses the total employees (both in domestic and foreign plants).

both domestic and foreign horizontal subsidiaries (described in Table 3.2). Thus, the unit of analysis is a subsidiary. The empirical specification is a *reduced form* of the exports-FDI tradeoff. That is, the analysis aims to understand the differential patterns between domestic and foreign horizontal subsidiaries. In terms of the theory presented above, the underlying assumption is that, whenever a firm sell a particular product to a foreign market, domestic subsidiaries are associated with exports whereas foreign subsidiaries are associated with FDI, and thus substitute for exports. The question asked in this exercise is, what characteristics of an industry make it more likely to be replicated abroad (i.e. that exist as a foreign subsidiary)?

According to Equation (3.3), $a_I^{1-\varepsilon}$, which represents the productivity threshold after which a firm engages in FDI, is a function of $\tau(t,d)$ and $\kappa(t,d)$ as well as the fixed costs and demand variables. It is important to clarify that industries may vary in their $a_I^{1-\varepsilon}$ threshold, and its value will determine the likelihood of horizontal expansion for that industry (given the distribution of productivity for firms within each sector). That is, controlling for demand variables and fixed costs, industries with a lower $a_I^{1-\varepsilon}$ will be more likely to be horizontally expanded, and vice-versa.

In this context, for a given firm and location, if a subsidiary is replicating production abroad (i.e. foreign and horizontal), it implies that the productivity level of such MNC goes beyond the minimum industry-specific threshold $a_I^{1-\varepsilon}$ for which FDI becomes more profitable than exports, in that industry. Controlling for MNC productivity, thus, exploiting variation in observed variables will shed light on the the determinants of the $a_I^{1-\varepsilon}$ value for different industries, or alternatively, the likelihood of horizontal expansion[28]:

$$Foreign_s = \beta_k \cdot k_s + \beta_t \cdot log(t_s) + controls_{h,s} + \varphi_h + e_{h,s} \tag{3.7}$$

Where the independent variable is a dummy which takes the value 1 if the subsidiary is a foreign horizontal affiliate of the firm (and 0 if it is a domestic one). k_s is a measure of knowledge intensity of the economic activity (i.e., the manufactured good or product) of

[28]That is, $Prob(a^{1-\varepsilon} > a_I^{1-\varepsilon})$.

the foreign subsidiary. t_s is the unit shipping cost for the good manufactured in the foreign subsidiary. $controls_{h,s}$ is a vector of variables that control for the size of the market and factor endowments of the host country relative to that of the country of the headquarters,[29] which controls for aggregate demand and cost of producing in the host country. φ_h represents MNC fixed effects, which controls for the productivity level a of the firm. It is worth mentioning that subsidiaries within a single MNC might differ in their economic activity, thus allowing for within-firm variation in the right hand side variables of the empirical specification (see Figure 3.6). Finally, $e_{h,s}$ is the error term.

According to the theoretical framework presented above, we expect the following β_k to be negative (see Equation (3.5) and Figure 3.1).

The results are presented in Table 3.4. All the columns include the control variables. The table uses the *experience plus training* measure discussed above as a proxy for k, which is measured in standard deviations from the mean.

Column 1 presents the complete specification. The results suggest that, everything else equal, industries that are one standard deviation above the mean in terms of their knowledge intensity, are 3.6 percentage points less likely to be replicated abroad. This represents a reduction of about 12% given the unconditional probability of being a foreign affiliate in the sample (which is 29% as shown in Table 3.2). For instance, according to this estimation, semiconductors (SIC 3674), which is characterized by having workers with an average of over 80 months of required experience plus training[30], is about 30 percentage points less likely to be replicated abroad than a meat packing plant (SIC 2011), which its workers have, on average, 37 months of experience plus training.

The estimator for β_k is robust across all specifications. This result controls for the unit shipping cost, and for the size of the market and factor endowments of the host country relative to that of the country of the headquarters[31]. According to the theoretical framework

---

[29]i.e., $y_{h,s} = log(y_h) - log(y_s)$.

[30]see Appendix Section C.5

[31]This is 1 for all domestic subsidiaries, naturally.

Table 3.4: *Determinants of Foreign Replication of Production*

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0359	-0.0349	-0.0230	-0.0348
	(0.017)**	(0.016)**	(0.013)*	(0.014)**
log(t)	-0.0235			-0.0198
	(0.023)			(0.022)
GDP per capita ratio	-0.3952		-0.4008	0.3863
	(0.131)***		(0.128)***	(0.131)***
Population ratio	0.0848		0.0866	-0.0668
	(0.019)***		(0.019)***	(0.028)**
Capital per worker ratio	0.3299		0.3326	-0.2328
	(0.080)***		(0.078)***	(0.069)***
Human Capital ratio	0.9534		0.9499	0.0294
	(0.180)***		(0.176)***	(0.072)
Land per worker ratio	-0.1029		-0.0994	0.1103
	(0.018)***		(0.018)***	(0.045)**
Constant	0.2220	0.2961	0.2614	0.9500
	(0.046)***	(0.009)***	(0.007)***	(0.104)***
N	61410	64462	64389	61410
R-squared	0.52	0.40	0.51	0.56
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	Y

The table presents results for the estimation of Specification (3.7) using a sample of domestic and foreign subsidiaries that replicate home production. The left hand side variable is a binary variable that takes the value 1 if the subsidiary is foreign. The variables in the right hand side include the unit shipping cost associated with the industry, knowledge intensity measures (in standard deviations from the mean) and other controls. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parentheses.

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

above, this is a straightforward result from the assumption that $\partial\kappa/\partial k > 0$

The inclusion of host country fixed effects in Column 4 rules out other potential stories that could be driving the results. For instance, poor intellectual property regulation in different countries.[32] The estimate of β_k is robust to the inclusion of this set of fixed effects in terms of its magnitude, negative sign and its statistical significance.[33]

The analysis presented in Table 3.4 seems to support Proposition (1) of the model. The next subsection focuses on Proposition (2).

3.5.4 Is the cost of knowledge transfer increasing in distance?

The previous sample is not useful to test the implications of distance, because there is no information on where are the domestic subsidiaries exporting to, if at all. Thus, to test the implications of Proposition (2), I will use a dataset that includes only foreign affiliates. The idea is to understand whether there are differential patterns in the data for horizontal affiliates (i.e. replication of production), using as a comparison group the non-horizontal subsidiaries. The underlying assumption of using non-horizontal foreign subsidiaries as a counterfactual is that the marginal cost of transferring knowledge is zero (or very little) for non-horizontal subsidiaries. While there is likely a fixed cost of transferring knowledge to non-horizontal subsidiaries when they are created or acquired, the assumption of zero *marginal* costs for this type of subsidiaries relates to the intuition that there is less the headquarters can do to offer ongoing troubleshooting or to train workers in these plants when it comes to production lines that are essentially different from the ones that exist at home. Therefore, controlling for variables that would explain the decision of a firm to locate a foreign subsidiary in a given location (regardless of whether it is horizontal or not), the residual differences could be attributed to the cost of transferring knowledge, and more so if they are consistent with the conceptual framework. Yet, given the assumption is not

[32] Appendix Section C.6.4 presents results excluding China from the sample, to alleviate possible biases this country might generate in the results due to IP concerns. The results are robust to the exclusion of China.

[33] The results are robust to using parent industry (2-digit) interacted with host country fixed effects, to allow for differential policies at the country level for different types of industries.

testable, Section 3.5.4 below relaxes it, and find consistent results.

The empirical specification for this exercise is described in equation 3.8:

$$HOR_s = \beta_k \cdot k_s + \beta_d \cdot log(d_{h,s}) + \beta_t \cdot log(t_s) + controls_{h,s} + \varphi_h + e_{h,s} \tag{3.8}$$

Where the independent variable is a dummy which takes the value 1 if the subsidiary (indexed by s) in that observation is a horizontal foreign affiliate, and 0 if, instead, is a non-horizontal foreign subsidiary. Again, k_s is a measure of knowledge intensity in standard deviations from the mean, associated with the foreign subsidiary. $d_{h,s}$ is the distance between the headquarters and the foreign subsidiary. t_s is the unit shipping cost for the good manufactured in the foreign subsidiary. $controls_{h,s}$ is the same vector as in specification 3.7. Similarly to before, φ_h represents MNC fixed effects and $e_{h,s}$ is the error term. If the mechanisms of the model are in place, we could expect a negative β_d, which could *only* be explained if κ increases with distance

The results are presented in Table 3.5. All the columns include the control variables. The table uses the *experience plus training* measure discussed above as a proxy for k.

Column 1 presents the complete specification, while the other columns vary in the number of variables used in the regression. The estimator for β_k is negative and statistically significant; the estimator for β_d is also negative and statistically significant; and the estimator for β_t has the expected positive sign, but lacks statistical significance.

Before turning into the coefficient of interest for this exercise (β_d), it should be noted that the negative sign for the estimator of β_k is consistent with the previous results in Table 3.4 and Equation (3.5) of the theoretical framework. More specifically, an industry with a knowledge intensity measure one standard deviation above the mean is about 8.7 percentage points less likely to be horizontally expanded, compared to non-horizontal affiliates. Hence, once again, the data suggests that the barriers associated with the transmission of knowledge from the headquarters to the subsidiaries are an important determinant of horizontal expansion.

Across all specifications that include $log(d)$, the estimator for β_d remains negative and

Table 3.5: *Determinants of Horizontal FDI*

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0877	-0.0872		-0.0898
	(0.043)**	(0.043)**		(0.043)**
log(d)	-0.0242		-0.0239	-0.0230
	(0.009)**		(0.009)**	(0.010)**
log(t)	0.0233	0.0239	0.0611	0.0229
	(0.066)	(0.066)	(0.056)	(0.065)
GDP per capita ratio	0.1297	0.1311	0.1282	0.9039
	(0.056)**	(0.057)**	(0.056)**	(0.243)***
Population ratio	0.0128	0.0210	0.0142	0.3127
	(0.007)*	(0.007)***	(0.007)**	(0.077)***
Capital per worker ratio	-0.0833	-0.1005	-0.0792	-0.6599
	(0.045)*	(0.047)**	(0.046)*	(0.182)***
Human Capital ratio	-0.0052	0.0370	-0.0098	-0.4287
	(0.063)	(0.063)	(0.064)	(0.282)
Land per worker ratio	-0.0131	-0.0106	-0.0131	0.0124
	(0.007)*	(0.008)	(0.007)*	(0.054)
Constant	0.6264	0.4426	0.6745	0.7577
	(0.161)***	(0.135)***	(0.148)***	(0.226)***
N	55136	55137	55136	55136
R-squared	0.47	0.47	0.47	0.47
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	Y

The table presents results for the estimation of Specification (3.8) using a sample of foreign subsidiaries of MNCs. The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as a horizontal expansion. The variables in the right hand side include the distance from the MNC headquarters to the foreign subsidiary (in logs), the unit shipping cost (in logs), knowledge intensity measures (in standard deviations from the mean) and other controls. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parentheses.

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

statistically significant. As explained above, in a model that ignores the cost of transferring knowledge, an increase in distance will unequivocally *increase* the incentives for horizontal FDI. However, only the inclusion of κ in the model as an increasing function of distance would explain the obtained results, which suggests that an increase in d would *reduce* the likelihood of horizontal FDI.[34]

What does $\beta_d < 0$ imply? The theoretical model, as summarized in the right panel of Figure 3.2, contemplates a case in which a firm serving a further away market would be better-off by exporting than by setting up a foreign affiliate, because transmitting knowledge to this remote location will significantly lower profits from FDI relative to profits from exports. However, given that the empirical specification is a reduced-form of the theoretical implications, it is not possible to distinguish between the case in which the firm effectively *substitutes* FDI with exports, or alternatively, the case in which the firm *reduces* its horizontal FDI in absolute terms, driven by a reduction of total sales (both through FDI and exports) in a further away location. In both cases, though, the negative sign of β_d implies that the cost of transferring knowledge is increasing with distance. In fact, given that the control group includes vertical subsidiaries, there are less reasons to expect this result. In theory, vertical subsidiaries are located closer to the headquarters as compared to horizontal subsidiaries (given the transportation costs associated with importing the intermediate goods from the vertical subsidiary to the headquarters). Therefore, a negative estimator for β_d is even more striking.

Finally, the estimator for β_t is positive in sign, though statistically insignificant across all specifications. The positive sign is consistent with the proximity-concentration hypothesis: firms will tend to serve foreign markets through foreign affiliates for goods with larger trade costs (e.g., Brainard 1993, 1997; Helpman, Melitz and Yeaple 2004).

Similarly to Table 3.4, Column 4 includes host country fixed effects, which would control for poor intellectual property regulation in different countries. It is important to acknowledge that the specification lacks variables that control for industry-specific fixed

[34]See Appendix Section C.1 for more details on the theoretical conditions for this to happen.

costs of exporting and creating new plants. It can be argued that, as long as fixed costs are not dependent on k or d, then the results are indicative of the explained mechanisms. If fixed costs are the same across industries and countries, then their exclusion should not bias the results. If the fixed costs are country-dependent, then the controls included in Column 4 would account for them.

A trade-off between distance and knowledge intensity

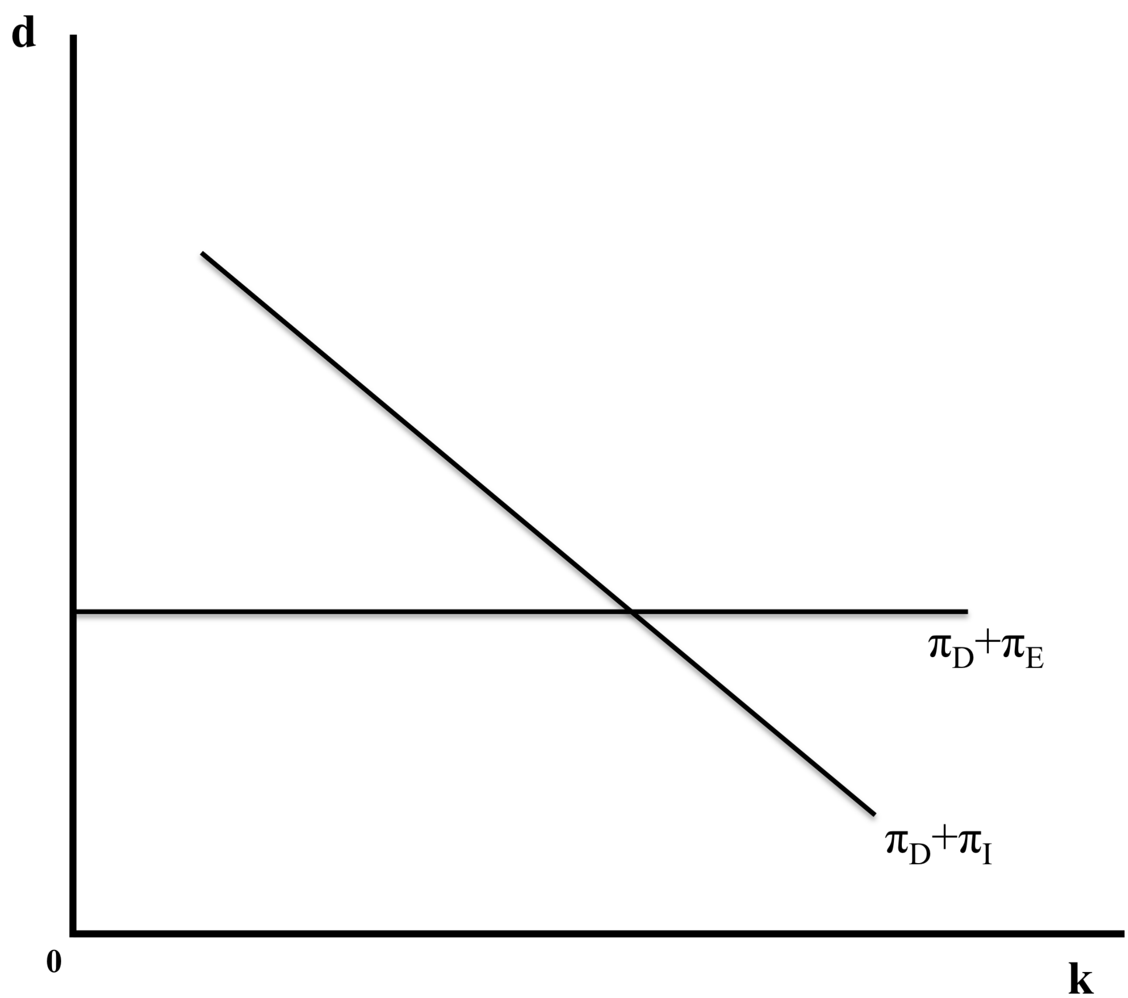
This subsection relaxes the underlying assumption which was required to compare horizontal to non-horizontal foreign subsidiaries stated above, which is critical to correctly interpret the estimation of β_d.

The theoretical framework above provides guidance to address this question in a different way. A firm's total profit when it engages in FDI is $\pi = \pi_D + \pi_I$. Given that π_I is subject to cost κ, and κ increases both in k and d, then $\partial\pi/\partial k < 0$ and $\partial\pi/\partial d < 0$. Figure 3.8 abstracts from the model the expected relationship between d and k that drives a firm's decision to engage in FDI. The figure includes the case that assumes linear relationships. In it, each line represents a profit function. The curve $\pi_D + \pi_E$ represents total profits for an exporting firm, while the curve $\pi_D + \pi_I$ represent total profits for a firm engaging in FDI instead. The profit for an exporting firm does not vary with the level of knowledge intensity (k) of the good, whereas the profit for the same firm if it engages in FDI instead does vary with k. Both profits functions decrease in distance, represented by d.

In all cases, however, it can be seen that for higher levels of k and d (i.e., knowledge intensity and distance, respectively) firms would be better off by substituting exports with FDI. The opposite happens for cases in which both k and d are low. Moreover, even when firms engage in FDI, their profits decrease with both distance and knowledge intensity. Thus, MNCs in knowledge intensive products would be better off by locating their foreign subsidiaries at closer geographic distance.

I explore whether relationship between k and d described above is seen using only the horizontal foreign subsidiaries in the data. That is, conditional on being an horizontal

Figure 3.8: *Profit curves, in the k and d dimension*



The figure is a graphical representation of a firm's profit as a function of k and d. The curve $\pi_D + \pi_E$ represents total profits for an exporting firm, while the curve $\pi_D + \pi_I$ represent total profits for a firm engaging in FDI instead.

foreign subsidiary, do we see a clear negative relationship between the knowledge intensity of its sector, and the distance to its headquarters? The proper way to do this is to analyze these variables after controlling for the regressors in Specification (3.8). Hence, this exercise has two steps.

First, I decompose distance and knowledge intensity and keep the part that is not explained by these other regressors (i.e., the residuals). That is, I define:

$$\begin{aligned} U\left[log(d)\right] &= log(d) - \gamma_t^1 log(t) - controls'_{h,s}\gamma_c^1 - \varphi_h \\ U\left[k\right] &= k - \gamma_t^2 log(t) - controls'_{h,s}\gamma_c^2 - \varphi_h \end{aligned}$$

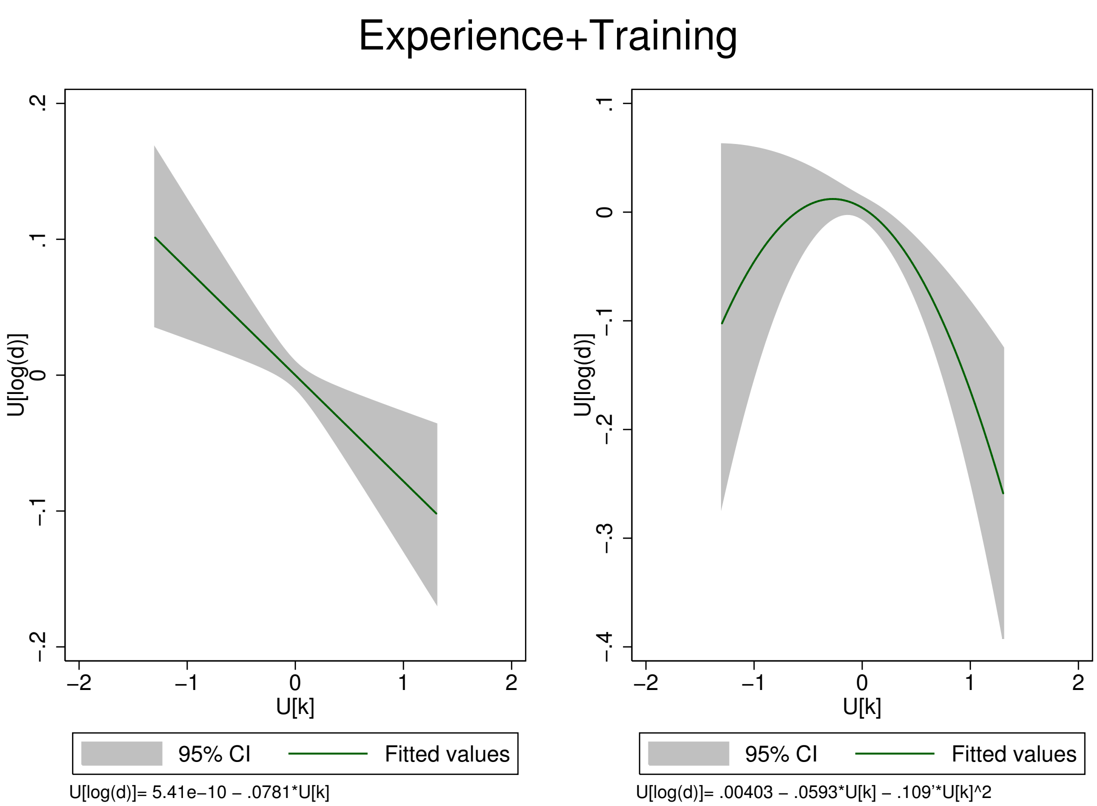
Where the γ coefficients are estimated by regressing $log(d)$ and k on the regressors, limiting the sample to horizontal foreign subsidiaries only. Notice that the inclusion of MNC firm fixed effect, imply that the residuals will contain within firm variation only.

The second step is to estimate $U\left[log(d)\right]$ and $U[k]$ using the sample, and to find a functional form that properly fits the relationship under consideration. As explained above, we expect this relationship to be negative. Figure 3.9 presents the results of this exercise, using the *experience plus training* indicator as a proxy for k. The left column performs a linear fit between k and d while the right column performs a quadratic firm between the two.

The linear fit shows a monotonic decreasing relationship between k and d, as depicted in Figure 3.8. In its linear form, the calculation suggests that the distance to the headquarters is shorter by 7.8% for every standard deviation above the mean in knowledge intensity. This implies that, for an American MNC, a meat packing subsidiary would be located in Turkey (approximately, 10000Km from USA), while a semiconductor plant would be located in Ireland (approximately 6500km from USA), *ceteris paribus*. The evidence hence suggests, that, indeed, firms face a trade-off between distance and knowledge intensity, providing further evidence on the fact that the cost of knowledge transmission is a function of both these terms.

Interestingly, the quadratic fit suggests an inverted U-shaped relationship. That is, the

Figure 3.9: *Estimated relationship of $U[k]$ and $U[log(d)]$*



The figure presents the empirical fit for the relationship between d and k (the latter proxied by the *experience plus training* measure). The left column performs a linear fit between k and d while the right column performs a quadratic firm between the two. The grey area represents the 95% confidence interval for the estimated relationship.

estimated quadratic relationship does not seem to be monotonically decreasing for the lower values of k (although a flat or even negative slope in that area cannot be rejected in the data either). However, and perhaps more importantly, for higher levels of knowledge intensity there is a clear negative relationship with distance. This result is qualitatively important, given that it would be consistent with the idea that distance appears to matter much more for higher levels of knowledge intensity. Intuitively, this means that after certain level of knowledge intensity, the more sophisticated products are the closer the foreign subsidiaries will be located to the headquarters. The negative second derivative implied in the fit suggests that the documented negative relationship intensifies with the level of knowledge intensity.

Overall, the data supports the existence of a trade-off between distance and knowledge intensity for horizontal foreign subsidiaries. This implies that, if MNC do expand horizontally, the foreign affiliates will tend to be geographically closer to the headquarters the more knowledge intensive the product under consideration is.[35]

These results, in their reduced form, are consistent with the ones found by Keller and Yeaple (2013). They find that distant foreign affiliates in knowledge intensive sectors perform worse. The authors, however, attribute these results to additional trade costs due to the substitution of transferring knowledge with intermediate goods. The framework and results presented above present an alternative explanation to this finding, in which the performance of subsidiaries is highly affected by the inefficiencies of transferring knowledge at longer distances, and not higher intra-firm cost induced by intermediate goods. The next section explores this issue further.

3.5.5 Ease of communication and knowledge transmission

The empirical results, while consistent with the theoretical framework, might be driven by factors other than knowledge not accounted for, in the presence of omitted variable bias.

[35] Appendix Section C.6.5 replicates these results excluding foreign subsidiaries located in Western Europe owned by a Europe-based MNC, given the relative shorter distances within the continent. Results are robust to the exclusion of these firms.

For instance, a conventional explanation in the literature would be that knowledge intensive sectors are associated with higher intra-firm trade of intermediate goods, making it less profitable to locate those plants in far away locations (Irrazabal et. al. 2013; Keller and Yeaple 2013).

Keller and Yeaple, in particular, assume that knowledge can be fully embedded in intermediate goods, that are in turn shipped to remote locations. However, this assumption is not feasible for tacit knowledge. Thus, it could well be that it is the cost of transmitting tacit knowledge which drives the documented relationship.

This subsection performs a test that disentangles between both explanations. If the cost of transferring knowledge is indeed an increasing function of distance – as argued – and thus, a determinant in the location decisions of firms, then easier communication between headquarters and subsidiaries would work as a cost-reducing mechanism for the purpose of transmitting knowledge. This would be hard to explain with the intra-firm trade mechanism, given that the ease of communication is orthogonal to the transportation costs of intermediate goods.

I test for this hypothesis by estimating an extended version model (3.8) which includes variables that proxy for the ease of communication within the firm. These variables, all measured for each subsidiary and its headquarters, are (1) the existence of a commercial non-stop air route (between airports within 100Km); (2) the number of overlapping working hours in a business day; and (3) a binary variable indicating whether the countries of both the headquarters and the subsidiary speak a common language[36] (see Section 3.4.2 for more details on the construction of these variables).

The purpose of utilizing these variables is to proxy for forms of communication that allow for the transmission of tacit knowledge, though they are quite different between themselves. As explained above, business travel provides the opportunity to work face-to-face, though it occurs with less frequency, given the high costs of traveling.[37] Being in the

[36]Defined as a language that is spoken by 8% or more of the population in both countries.

[37]Giroud (2012) finds that the existence of commercial air routes between subsidiaries and headquarters

same time zone allows for convenient real-time, day-to-day, communication, significantly reducing waiting time between the two ends for problem solving or consulting about specific tasks.[38] Lastly, if two countries speak a common language, it is more likely that the workers in both the local and remote locations of the same firm can communicate more easily, either in person or remotely, and better communicate (and more often) with each other.

The results are presented in Table 3.6. All columns use the *experience plus training* indicator to proxy for k.

Column 3 of Table 3.6 shows that the estimator for β_d is reduced in magnitude by two thirds of its original value, and becomes statistically insignificant when using the number of overlapping working hours as a control (as compared to Column 1, which replicates the first specification of Table 3.5). This result suggests that being in the same time zone reduces the barriers to transferring knowledge induced by the distance component (given that the estimator for β_k maintains its magnitude and negative sign in those specifications, implying only the distance channel in $\kappa(k, d)$ is affected). That is, real-time communication effectively "reduces" the distance between the headquarters and its subsidiaries, by about two thirds. For the average foreign subsidiary, being in the same time zone is equivalent to being geographically closer to the headquarters by about 3500 Km. The ability to communicate on real-time for troubleshooting or other purposes, avoiding long waiting times, seems to ease knowledge transmission more than the ease of face-to-face interaction. The costs associated with the knowledge intensity component still seem play a role, regardless of communication.

An alternative explanation of the previous results that relies on shipping costs of intermediate goods can be ruled out: transportation costs should be just as expensive north to south as they are east to west.

In terms of the other variables, it can be seen in Column 2 that the existence of a non-stop

---

positively affects the profitability of the former.

[38]Stein and Daude (2007) find that time zone is an important determinant of aggregate FDI flows, which they attribute to better monitoring.

Table 3.6: *Determinants of Horizontal FDI, Ease of Communication*

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0877	-0.0878	-0.0868	-0.0875
	(0.043)**	(0.043)**	(0.043)**	(0.043)**
log(d)	-0.0242	-0.0254	-0.0076	-0.0187
	(0.009)**	(0.009)***	(0.014)	(0.008)**
log(t)	0.0233	0.0232	0.0239	0.0233
	(0.066)	(0.066)	(0.066)	(0.066)
Non-stop Flight		-0.0084		
		(0.010)		
Working hours overlap			0.0091	
			(0.005)*	
Common Language				0.0497
				(0.024)**
GDP per capita ratio	0.1297	0.1302	0.1271	0.1279
	(0.056)**	(0.056)**	(0.056)**	(0.055)**
Population ratio	0.0128	0.0131	0.0123	0.0139
	(0.007)*	(0.007)*	(0.007)*	(0.007)**
Capital per worker ratio	-0.0833	-0.0841	-0.0892	-0.0800
	(0.045)*	(0.046)*	(0.044)**	(0.044)*
Human Capital ratio	-0.0052	-0.0023	-0.0033	-0.0107
	(0.063)	(0.062)	(0.062)	(0.064)
Land per worker ratio	-0.0131	-0.0124	-0.0079	-0.0075
	(0.007)*	(0.007)*	(0.007)	(0.008)
Constant	0.6264	0.6383	0.4316	0.5718
	(0.161)***	(0.160)***	(0.217)**	(0.158)***
N	55136	55136	55136	55132
R-squared	0.47	0.47	0.47	0.47
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	N

The table presents results for the estimation of Specification (3.8) using a sample of foreign subsidiaries of MNCs. The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as a horizontal expansion. The variables on the right hand side include the distance from the MNC headquarters to the foreign subsidiary (in logs), the unit shipping cost (in logs), knowledge intensity measures (in standard deviations from the mean), and other controls. The right hand side also includes variables measuring the ease of communication between a headquarters and its subsidiaries. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parentheses.

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

commercial air route seems not to change the original results, thus hinting that face-to-face interaction plays a lesser role in the stated mechanisms. However, Column 4 shows that having a common language also effectively reduces the distance between a headquarters and its subsidiary by less than half.

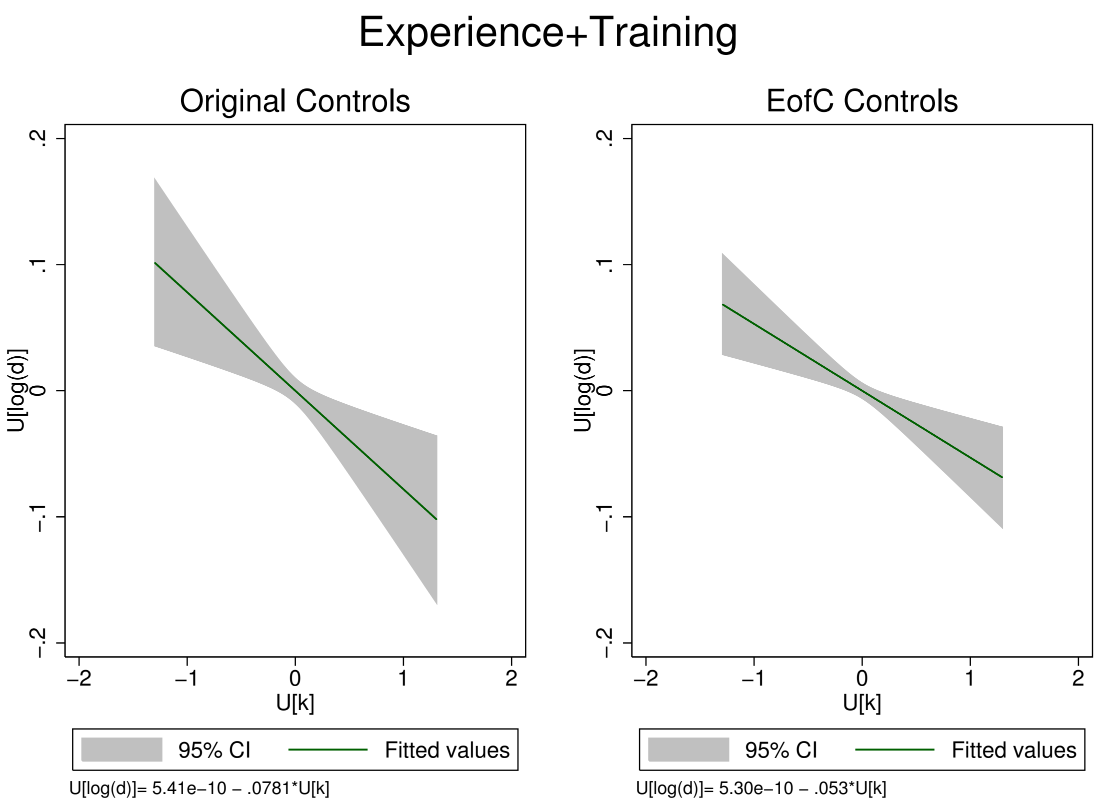
Figure 3.10 replicate Figure 3.9, this time adding as additional controls all the ease of communication variables that are included in Table 3.6. The left panel shows the linear fit with the original controls, while the right panel controls also for the ease of communication. It can be seen that the slope that defines the relationship between k and d after controlling for the ease of communications is about 33% flatter. While the negative relationship still seems to hold, the reduction in the slope is consistent with the results presented in this section.

These findings are insightful on their own. The results suggest that being in the same time zone and speaking a common language seems to facilitate the transmission of knowledge. The ability of managers in the headquarters to communicate with colleagues in foreign locations, for troubleshooting or consulting on an open-ended range of issues, is more efficient when communication happens in real time, without long waiting times. This might be even more relevant for transmitting tacit knowledge, given that complicated problems would require real-time interaction, and not just explanations being sent through fax or email. Furthermore, this logic could also serve as an example for arguing that the barriers of transmitting knowledge is increasing with distance: managers and workers in the headquarters might require working extra hours to communicate with their peers in foreign subsidiaries, incurring additional compensation and operational costs.

3.6 Concluding Remarks

This paper has provided evidence on the important role of knowledge, and the difficulties associated with its transmission in the day-to-day activities of MNCs. Sizable costs of transferring knowledge, even within firms, would have an impact on their strategies to either export or undertake foreign investment, directly affecting the global economy in

Figure 3.10: *Estimated relationship of* $U[k]$ *and* $U[log(d)]$



The figure presents the empirical fit for the relationship between d and k (the latter proxied by the *experience plus training* measure). The left column performs a linear fit between k and d using the original controls, while the right panel repeats the exercise adding the ease of communication variables as controls. The grey area represents the 95% confidence interval for the estimated relationship.

terms of trade and capital flows. Furthermore, the empirical analysis presents evidence of a tradeoff firms face, which drives them to locate foreign subsidiaries producing knowledge intensive goods in geographic locations that are closer to the headquarters. Thus, knowledge is not lighter than air. Rather, its diffusion is difficult and costly, and hence it has implications on economic activity.

These findings are not inconsistent with the mechanism of the proximity-concentration hypothesis (e.g., Brainard 1997), yet they introduce a new and unexplored dimension. The cost of transferring knowledge plays a role that counteracts the incentives to engage in FDI driven by transportation costs. Hence, FDI does not necessarily become more profitable than exports for all industries with high transportation costs.

More generally, the fact that geographic distance hinders the process of knowledge transmission is a result that defies the traditional way economists have thought about FDI and MNC activity. In most of the international economics research, it is taken as a given that knowledge is fully transferrable without incurring any costs whatsoever – not even for different types of technologies or goods. However, if one takes into account the large variety of different industries that exist in the world, and how they can dramatically differ in almost any dimension, it follows that we can expect each firm to set a strategy that is dependent on the types of products they produce and sell. In a globalized economy, being able to sell products at a global scale requires a minimum level of productivity, which firms achieve by acquiring productive knowledge. The way firms acquire and maintain this knowledge is through their workers in the headquarters and all of its relevant subsidiaries (domestic and foreign). The finding that knowledge transmission incurs costs that are dependent on distance would thus have a significant impact on the expansion decisions of MNCs.

Nonetheless, this paper has left open some other specific questions that will shed light on our general understanding of knowledge. For instance, is the cost of knowledge transmission a relevant determinant for service provider firms, as it is for manufacturing firms? Given the difference in the nature of services vs. manufacturing industries in terms of their tradability, we can expect different patterns in the data. Also, how does the knowledge intensity of the good relate to the existence of regional hubs, as opposed to different plants serving every foreign market? What tools and means are at a firm's disposal to enhance the process through which it transfers knowledge to its subsidiaries and workers? These and other questions are an essential part of the future research agenda.

Naturally, this research agenda also contains questions that have relevant policy implications. While governments intend to develop their private sectors by attracting foreign investment, designing an effective policy should answer questions such as: is there enough infrastructure in place to allow effective communication for foreign firms? Should the focus be on specific types of firms and specific industries for which knowledge transmission will be easier? Do all types of products have the potential to generate productivity spillovers to domestic firms, or only those for which the cost of knowledge transmission is low?

All in all, despite the fact that productivity outweighs factor accumulation in growth accounting exercises (Hall and Jones 1999, Caselli 2005), the process through which knowledge is accumulated by economic agents is still an under-researched area. However, a better understanding of this process is critical to answering open questions in economics. The difficulties associated with transferring and acquiring knowledge, which translates into productivity shifts, are not unique to MNCs. They can also relate to domestic firms (e.g., Bloom et. al. 2013; Kalnins and Lafontaine 2013), investors (e.g., Coval and Moskowitz 2001), innovation (e.g., Kerr 2008) and even countries' export baskets diversification (Bahar et. al. 2014). At a larger scale, the documented evidence reinforces the importance of knowledge transmission in overall economic activity. Thus, understanding the ways knowledge affects economic activity lies at the core of important and unanswered questions on convergence, development and growth. Knowledge and its diffusion, after all, are significant phenomena that can alter global economic patterns in as-of-yet unexplored ways.

References

ABRAMOVITZ, M. (1956). Resource and Output Trends in the United States since 1870. *The American Economic Review*, **46** (2), 5–23.

ACEMOGLU, D., JOHNSON, S. and MITTON, T. (2009). Determinants of Vertical Integration : Financial Development and Contracting Costs. *The Journal of Finance*, **LXIV** (3), 1251–1290.

AITKEN, B. and HARRISON, A. (1999). Do domestic firms benefit from direct foreign investment? Evidence from Venezuela. *American Economic Review*, **89** (3), 605–618.

ALFARO, L. and CHARLTON, A. (2009). Intra-Industry Foreign Direct Investment. *American Economic Review*, **99** (5), 2096–2119.

— and CHEN, M. X. (2012). Surviving the Global Financial Crisis: Foreign Ownership and Establishment Performance. *American Economic Journal: Economic Policy*, **4** (3), 30–55.

ALVAREZ, F., BUERA, F. and LUCAS, R. (2012). Idea Flows, Economic Growth, and Trade.

ANDERSEN, T. B. and DALGAARD, C.-J. (2011). Flows of people, flows of ideas, and the inequality of nations. *Journal of Economic Growth*, **16** (1), 1–32.

ANDERSON, J. (1979). A Theoretical Foundation for the Gravity Equation. *American Economic Review*, **69** (1), 106–116.

— and VAN WINCOOP, E. (2001). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, **93** (1), 170–191.

ARROW, K. J. (1969). Classificatory Notes on the Production and Transmission of Technologcal Knowledge. *The American Economic Review*, **59** (2), 29–35.

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

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

BANK, W. (2013). World development indicators online.

BECKER, R., GRAY, W. and MARVAKOV, J. (2013). NBER-CES Manufacturing Industry Database: Technical Notes.

Bergstrand, J. (1985). The Gravity Equation in International Trade: some Microeconomic Foundation and Empirical Evidence. *The Review of Economics and Statistics*, **67** (3), 474–481.

— (1989). The generalized gravity equation, monopolistic competition, and the factor-proportions theory in international trade. *The Review of Economics and Statistics*, **71** (1), 143–153.

Bernard, A. B., Jensen, J. B. and Schott, P. K. (2006). Trade costs, firms and productivity. *Journal of Monetary Economics*, **53** (5), 917–937.

Black, S. and Strahan, P. (2002). Entrepreneurship and bank credit availability. *The Journal of Finance*, **LVII** (6), 2807–2833.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D. and Roberts, J. (2012). Does Management Matter? Evidence from India. *The Quarterly Journal of Economics*, **128** (1), 1–51.

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

Borjas, G. J. and Bratsberg, B. (1996). Who Leaves? The Outmigration of the Foreign-Born. *The Review of Economics and Statistics*, **78** (1), 165.

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

Brainard, S. (1993). A simple theory of multinational corporations and trade with trade-off between proximity and concentration. *NBER Working Paper Series*, (4269).

— (1997). An empirical assessment of the proximity-concentration tradeoff between multinational sales and trade. *The American Economic Review*, **87** (4), 520–544.

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

— and Krugman, P. (1983). A reciprocal dumping model of international trade. *Journal of International Economics*, **15** (1983), 313–321.

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

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

Carr, D., Markusen, J. and Maskus, K. (2001). Estimating the knowledge-capital model of the multinational enterprise. *The American Economic Review*, **91** (3), 693–708.

Caselli, F. (2005). Accounting for cross-country income differences. *Handbook of Economic Growth*, **1** (05), 679–741.

Caves, R. (1971). International Corporations: The Industrial Economics of Foreign Investment. *Economica*, **38** (149), 1–27.

Choudhury, P. (2014). Return Migration and Geography of Innovation in Mnes: A Natural Experiment of On-the-Job Learning of Knowledge Production by Local Workers Reporting to. *Harvard Business School Working Paper*, (14-078).

Coe, D. T. and Helpman, E. (1995). International R&D spillovers. *European Economic Review*, **2921** (94), 859–887.

—, — and Hoffmaister, A. W. (2009). International R&D spillovers and institutions. *European Economic Review*, **53** (7), 723–741.

Combes, P.-P., Lafourcade, M. and Mayer, T. (2005). The trade-creating effects of business and social networks: evidence from France. *Journal of International Economics*, **66** (1), 1–29.

Costinot, A., Oldenski, L. and Rauch, J. (2011). Adaptation and the boundary of multinational firms. *The Review of Economics and Statistics*, **93** (February), 298–308.

Coval, J. D. and Moskowitz, T. J. (2001). The Geography of Investment: Informed Trading and Asset Prices. *Journal of Political Economy*, **109** (4), 811–841.

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

Deardorff, A. (1998). Determinants of bilateral trade: does gravity work in a neoclassical world? In J. Frankel (ed.), *The Regionalization of the World Economy*, Chicago: University of Chicago Press.

Docquier, F., Marfouk, A., Ozden, C. and Parsons, C. (2010). Geographic, gender and skill structure of international migration. In *Report written for the Economic Research Forum*, pp. 1–27.

Eaton, J. and Kortum, S. (1997). Technology and bilateral trade. *NBER Working Paper Series*, (6253).

Fan, J. and Lang, L. (2000). The Measurement of Relatedness : An Application to Corporate Diversification. *The Journal of Business*, **73** (4), 629–660.

Feenstra, R. C., Lipsey, R. E., Deng, H., Ma, A. C. and Mo, H. (2005). World Trade Flows: 1962-2000.

—, Markusen, J. R. and Rose, A. K. (2001). Using the gravity equation to differentiate among alternative theories of trade. *Canadian Journal of Economics/Revue Canadienne d'Economique*, **34** (2), 430–447.

Felbermayr, G. J. and Jung, B. (2009). The pro-trade effect of the brain drain: Sorting out confounding factors. *Economics Letters*, **104** (2), 72–75.

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

Frankel, J. and Romer, D. (1999). Does Trade Cause Growth? *American Economic Review*, **89** (3), 379–399.

Frankel, J. A. (1997). *Regional Trading Blocs in the World Economic System*. Peterson Institute for International Economics.

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

Gould, D. (1994). Immigrant links to the home country: empirical implications for US bilateral trade flows. *The Review of Economics and Statistics*, **76** (2), 302–316.

Hall, R. E. and Jones, C. I. (1999). Why do Some Countries Produce So Much More Output Per Worker than Others? *The Quarterly Journal of Economics*, **114** (1), 83–116.

Harrison, A. E., Love, I. and McMillan, M. S. (2004). Global capital flows and financing constraints. *Journal of Development Economics*, **75** (1), 269–301.

Haskel, J., Pereira, S. and Slaughter, M. (2007). Does inward foreign direct investment boost the productivity of domestic firms? *The Review of Economics and Statistics*, **89** (3), 482–496.

Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Chung, S., Jímenez, J., Simoes, A. and Yildirim, M. A. (2011). *The Atlas of Economic Complexity: Mapping Paths to Prosperity*. Cambridge, MA.

— and Klinger, B. (2007). The Structure of the Product Space and the Evolution of Comparative Advantage. *CID Working Paper Series*, (146).

Helpman, E. (1984). A Simple Theory of International Trade with Multinational Corporations. *The Journal of Political Economy*, **92** (3), 451–471.

— and Krugman, P. (1985). *Market Structure and International Trade*. Cambridge, MA: MIT Press.

—, Melitz, M. J. and Yeaple, S. R. (2004). Export Versus FDI with Heterogeneous Firms. *American Economic Review*, **94** (1), 300–316.

Hidalgo, C. A., Klinger, B., Barabási, A. and Hausmann, R. (2007). The product space conditions the development of nations. *Science (New York, N.Y.)*, **317** (5837), 482–7.

Hippel, E. V. (1994). 'Sticky information' and the locus of problem solving: implications for innovation. *Management science*, **40** (4), 429–439.

Iranzo, S. and Peri, G. (2009). Migration and Trade: Theory with an Application to the Eastern-Western European Integration. *Journal of International Economics*, **79**, 1–19.

Irarrazabal, A., Moxnes, A. and Opromolla, L. (2013). The Margins of Multinational Production and the Role of Intrafirm Trade. *Journal of Political Economy*, **121** (1), 74–126.

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

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

—, Ozden, C., Spatareanu, M. and Neagu, C. (2011). Migrant networks and foreign direct investment. *Journal of Development Economics*, **94** (2), 231–241.

Kalnins, A. and Lafontaine, F. (2013). Too Far Away? The Effect of Distance to Headquarters on Business Establishment Performance. *American Economic Journal: Microeconomics*, **5** (3), 157–179.

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

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

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

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

— and — (2013). The Gravity of Knowledge. *American Economic Review*, **103** (4), 1414–1444.

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

Kogut, B. and Zander, U. (1992). Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology. *Organization science*, **3** (3), 383–397.

Kugler, M. and Rapoport, H. (2007). International labor and capital flows: Complements or substitutes? *Economics Letters*, **94** (2), 155–162.

— and — (2011). Migration, FDI, and the Margins of Trade. *CID Working Paper Series*, (222).

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

Leamer, E. and Levinsohn, J. (1995). International Trade Theory: The Evidence. In G. M. Grossman and K. Rogoff (eds.), *Handbook of International economics*, vol. III, *26*, Elsevier B.V., pp. 1341–1375.

Linder, S. B. (1961). *An essay on trade and transformation*. Stockholm: Almqvist & Wicksell.

Lipsey, R. (1978). The Creation of Microdata Sets for Enterprises and Establishments. *Annales de l'INSEE*, (30), 395–422.

Loungani, P., Mody, A. and Razin, A. (2002). The global disconnect: the role of transactional distance and scale economies in gravity equations. *Scottish Journal of Political Economy*, **49** (5), 526–543.

Markusen, J. (1997). Trade versus Investment Liberalization. *NBER Working Paper Series*, (6231).

—, Venables, A., Konan, D. and Zhang, K. (1996). A unified treatment of horizontal direct investment, vertical direct investment, and the pattern of trade in goods and services. *NBER Working Paper Series*, (5696).

Markusen, J. R. (1984). Multinationals, multi-plant economies, and the gains from trade. *Journal of international economics*, **16** (1984), 205–226.

— and Maskus, K. E. (2002). Discriminating among alternative theories of the multinational enterprise. *Review of International Economics*, **10** (4), 694–707.

Mayer, T. and Zignago, S. (2011). Notes on CEPII distances measures : The GeoDist database. *CEPII Working Paper*, (25).

Nunn, N. and Trefler, D. (2008). The boundaries of the multinational firm: an empirical analysis. In E. Helpman, D. Marin and T. Verdier (eds.), *The Organization of Firms in a Global Economy*, Cambridge, MA: Harvard University Press, pp. 55–83.

OECD (2012). International Direct Investment Statistics.

Pierce, J. and Schott, P. (2012). A concordance between ten-digit US Harmonized System Codes and SIC/NAICS product classes and industries. *Journal of Economic and Social Measurement*, (301).

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

— (1966). *The Tacit Dimension*. Chicago; London: University of Chicago Press, 2009th edn.

Portes, R. and Rey, H. (2005). The determinants of cross-border equity flows. *Journal of International Economics*, **65** (2), 269–296.

Pritchett, L. (1997). Divergence, big time. *The Journal of Economic Perspectives*, **11** (3), 3–17.

Rauch, J. E. and Trindade, V. (2002). Ethnic chinese networks in international trade. **84** (February), 116–130.

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

Saxenian, A. (2006). *The New Argonauts: Regional Advantage in a Global Economy*. Cambridge, MA: Harvard University Press.

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

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

Stock, J. and Yogo, M. (2002). Testing for Weak Instruments in Linear IV Regression. *NBER Technical Working Papers Series*, (284).

Szulanski, G. (1996). Exploring Internal Stickiness: Impediments to the Transfer of Best Practice Within the Firm. *Strategic management journal*, **17**, 27–43.

— (2002). *Sticky Knowledge: Barriers to Knowing in the Firm*. SAGE Strategy series, SAGE Publications.

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

Tong, S. (2005). Ethnic Networks in FDI and the Impact of Institutional Development. *Review of Development Economics*, **9** (4), 563–580.

United Nations (2010). COMTRADE database.

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

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

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

Appendix A

Appendix to Chapter 1

A.1 The Network of Export Similarity

The similarity in the export baskets of countries is strongly affected by variables that proxy for distance and is robust to the inclusion of institutional, income and factor endowment variables. In fact, one way to illustrate the strength of the similarity between neighboring countries is to represent the matrix of export similarity as a network, where each country is connected to the two other countries most similar to it. Figure A.1 presents 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 links is proportional to the similarity index and the color of the link indicates whether the similarity is driven by Primary and Resource Based (PRB) products (blue) or by NPRB products (red) (see Section A.4 of the web appendix for more details). We note that, in a large number of cases, the country with the most similar export structure is an immediate neighbor, such as in the case of France, Germany, Austria, the Czech Republic, Hungary and Slovakia or in the case of India, Pakistan, and Bangladesh. This visualization illustrates the strong association between proximity and export structure that characterizes the world economy.

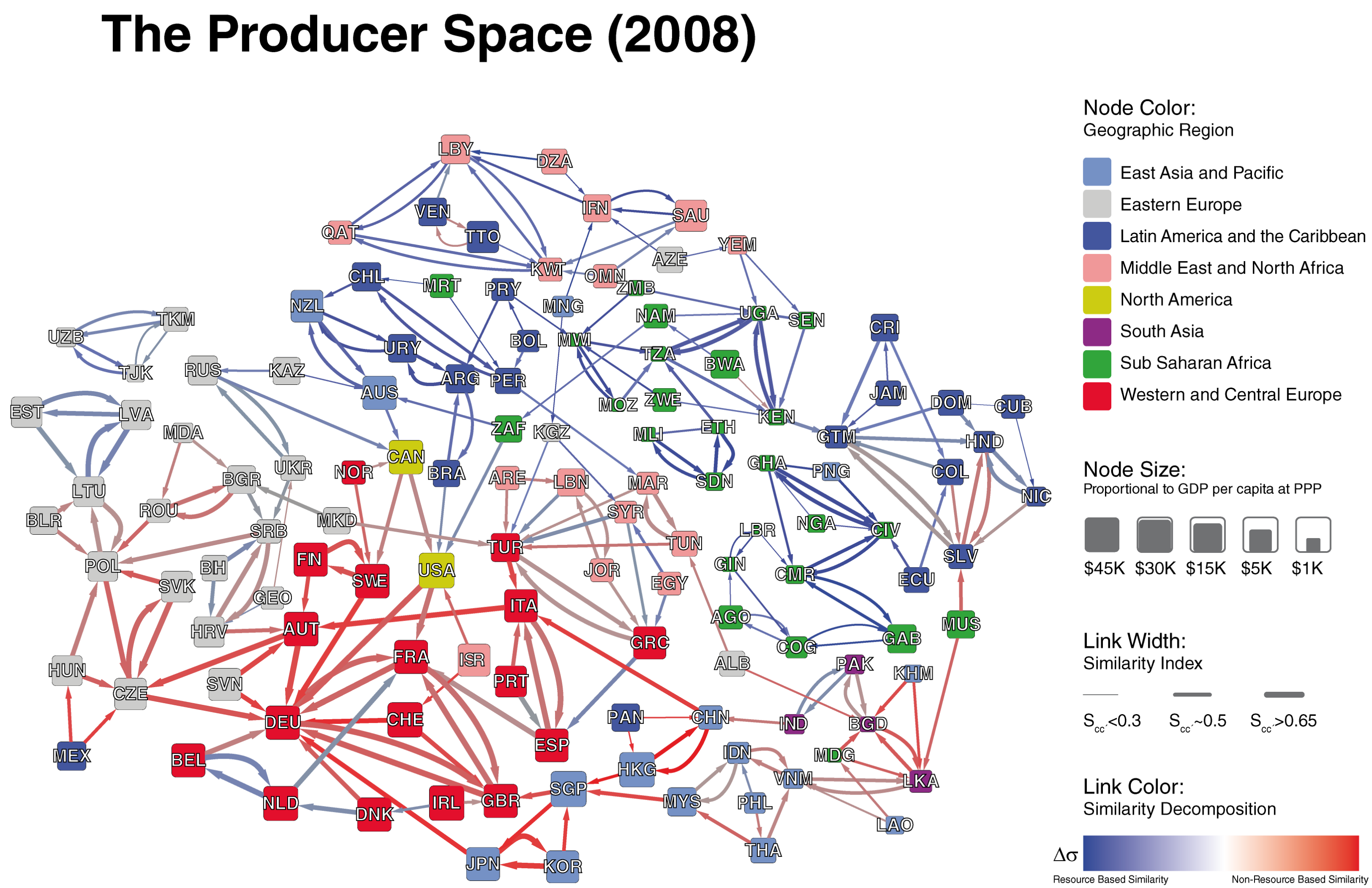
A.2 Bilateral Trade and Similarity in Exports

We pursue a more analytical approach to show that the similarity index among neighboring countries is mostly driven by their exports to the rest of the world, and not by the bilateral exports between themselves (see Figure 1.3). 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, identical to the way in which 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.

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 independent variables for year 2000. The results of such regression are in Table A.1.

The first two columns of Table A.1 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 for 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 repeat the exercise, but limit the dataset 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

Figure A.1: *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 A.1: *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, all products) as the dependent variable. Columns 1, 2 use all country-pairs in the sample, columns 3, 4 limit the sample to neighboring country pairs. Standard errors are clustered at the country-pair level.
$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

Similarity Index coefficient.

In all regressions, the similarity index as measured by exports to the ROW has a larger explanatory power. This hints that most of the similarity among countries and their neighbors is driven by exports to the ROW, and not by exports between themselves.

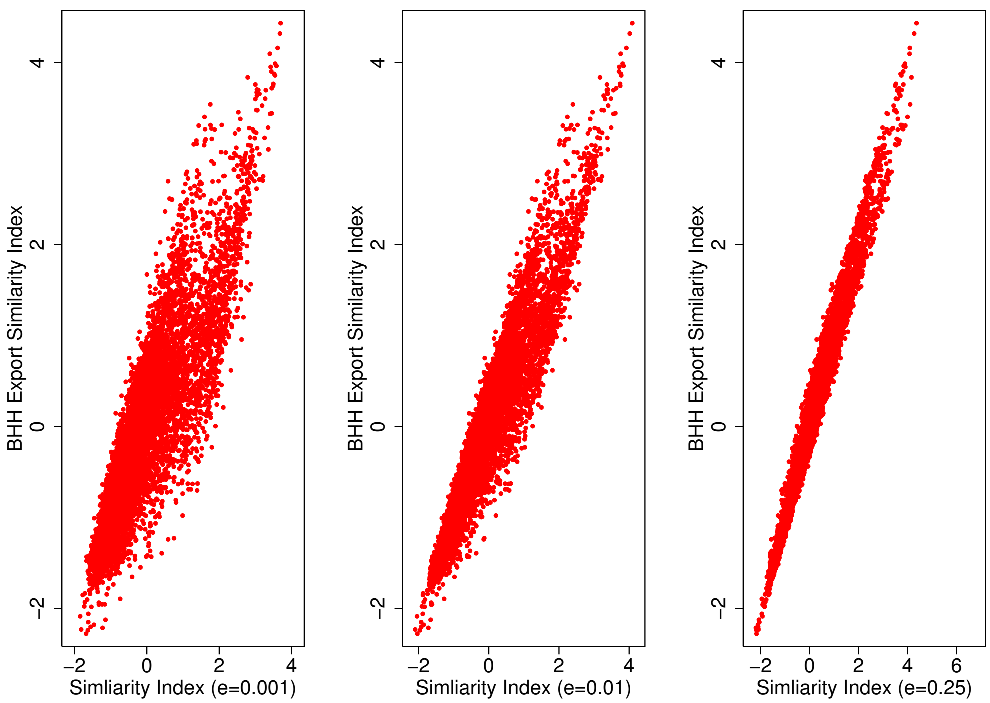
A.3 Robustness of the Stylized Facts

A.3.1 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 (1.2). Our original choice of $\varepsilon = 0.1$ allows 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 1.3. Therefore, we recalculated $S_{c,c'}$ defining ε as 0.001, 0.01 and 0.25. Figure A.2 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.

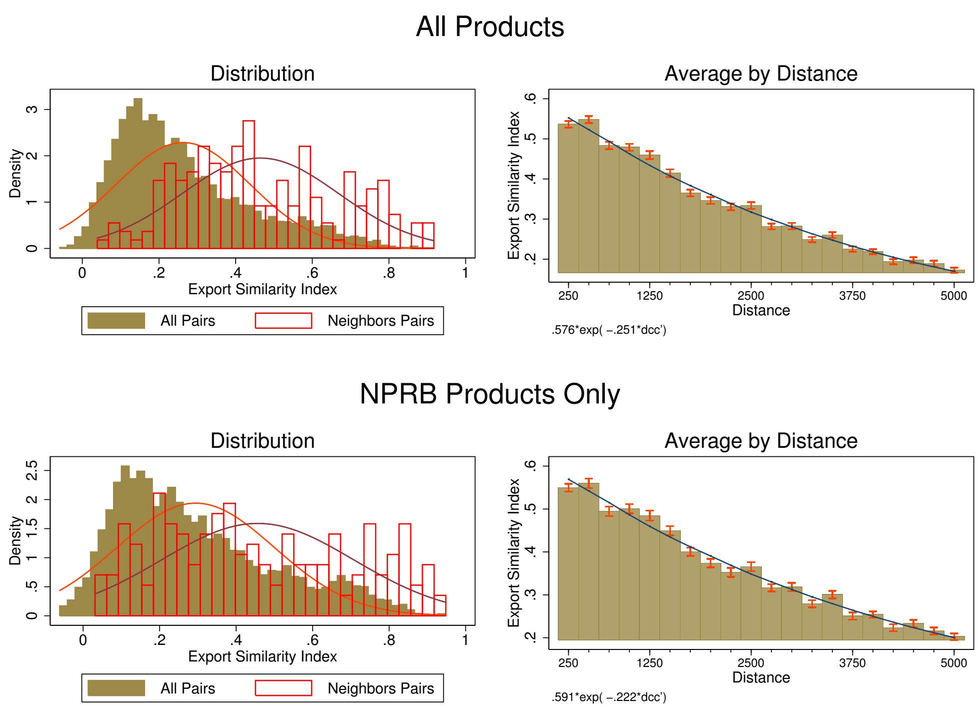
To convince the readers of the robustness of our choice of ε, we reproduce Figures 1.1

Figure A.2: *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 A.3: *Stylized Facts Similarity Index* ($\varepsilon = 0.001$)



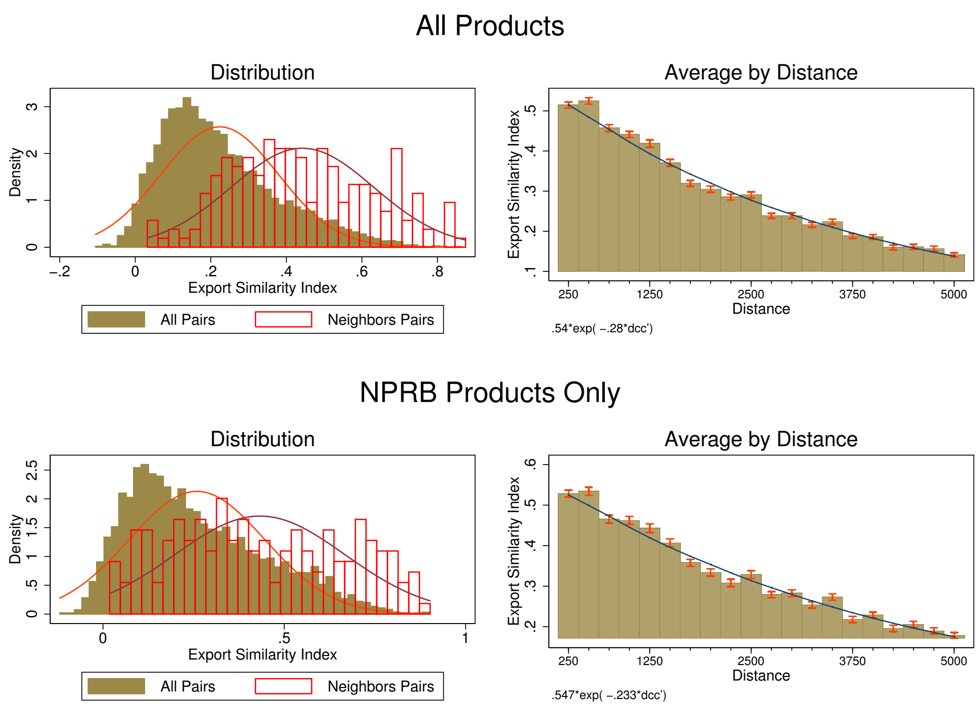
The left panel of the figure shows the distributions (in year 2000) of the Export Similarity Index (using $\varepsilon = 0.001$) for All (not neighbors) Country Pairs, and for Neighbors Pairs only. The right panel shows the average Export Similarity Index (using $\varepsilon = 0.001$) for country pairs 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.

and 1.2, along with Table 1.3, using this time $S_{c,c'}$ defined with each of the new ε values. Results are presented in Figures A.3-A.5, and Tables A.2-A.4. For all variations of ε, the results are qualitatively robust to our original measure.

A.3.2 Indexed RCA Export Similarity Index

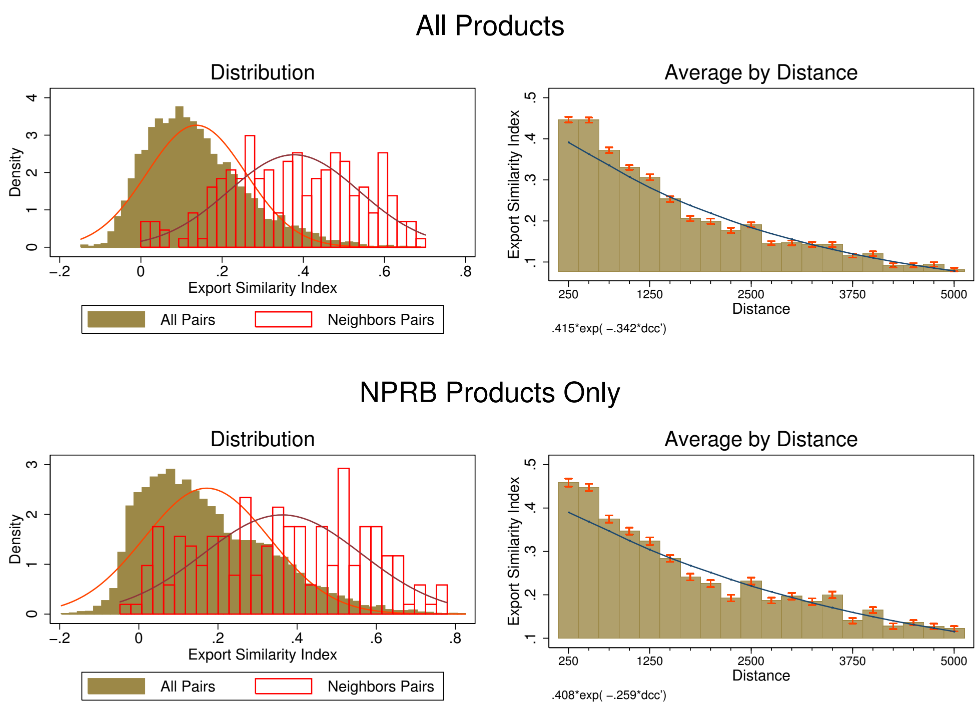
To erase concerns regarding the log-transformation of the RCA vectors in computing $S_{c,c'}$ based on equation (1.2), we constructed a variation of our similarity index, which we call the Indexed RCA Export Similarity Index, by substituting $r_{c,p}$ in equation (1.2), by:

Figure A.4: *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 A.5: *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 A.2: *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 (1.3) 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 A.3: *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 (1.3) 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.

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

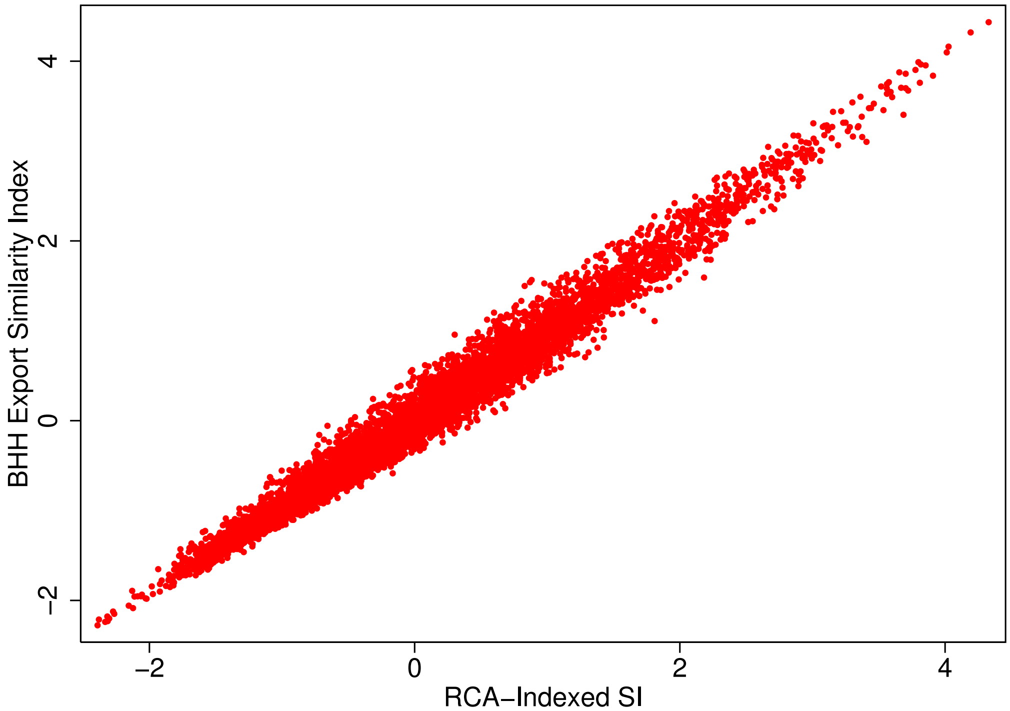
Table A.4: *Correlates of the Similarity Index, using $\varepsilon = 0.25$*

	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 (1.3) 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 A.6: *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.

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

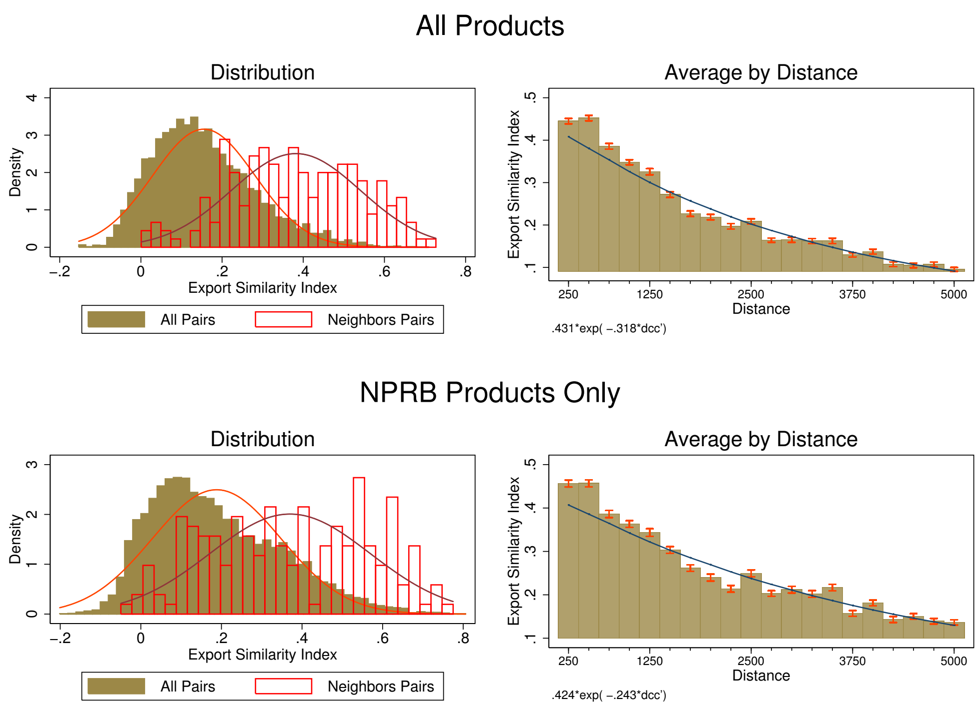
Under this definition, $r_{c,p} = 1$ if $RCA_{c,p} \to \infty$, and $r_{c,p} = -1$ if $RCA_{c,p} = 0$. This transformation also deals with fat tails in the original distribution of $RCA_{c,p}$ and hence eliminates the need to do a log-transformation.

Figure A.6 shows the high correlation between the original $S_{c,c'}$ and the Indexed RCA Export Similarity Index.

Figure A.7 uses the Indexed-RCA $S_{c,c'}$ to replicate Figures 1.1 and 1.2. Our original $S_{c,c'}$ is robust to this new transformation in terms of its correlation with distance.

We also estimated model (1.3) using the Indexed-RCA $S_{c,c'}$ (standardized with mean zero and unit standard deviation). The results are presented in Table A.5. Our static analysis

Figure A.7: *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.

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.3.3 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:

$$S^{F\&K}_{c,c'} = \sum_{p} min(s^{c}_{p}, s^{c'}_{p})$$

where p represents products, c and c' represent any two countries and s^{c}_{p} 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 $S^{F\&K}_{c,c'} = 1$.

Figure A.8 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 ($\rho = 0.65$). This implies that both indexes capture much of the same information.

Figure A.9 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 non neighbors, and the declining relationship of the index with distance for all products. The lower panel replicates the graphs using the F&K index computed with NPRB products only. We find that the results are robust to the ones presented in Figures 1.1 and 1.2.

We also replicated the analysis presented in Table 1.3, which estimates model (1.3). This time, we use a normalized version of the F&K Similarity Index as the dependent variable with mean zero and unit standard deviation. The results are presented in Table A.6. Our static analysis is robust to using this way of measuring similarity in export baskets.

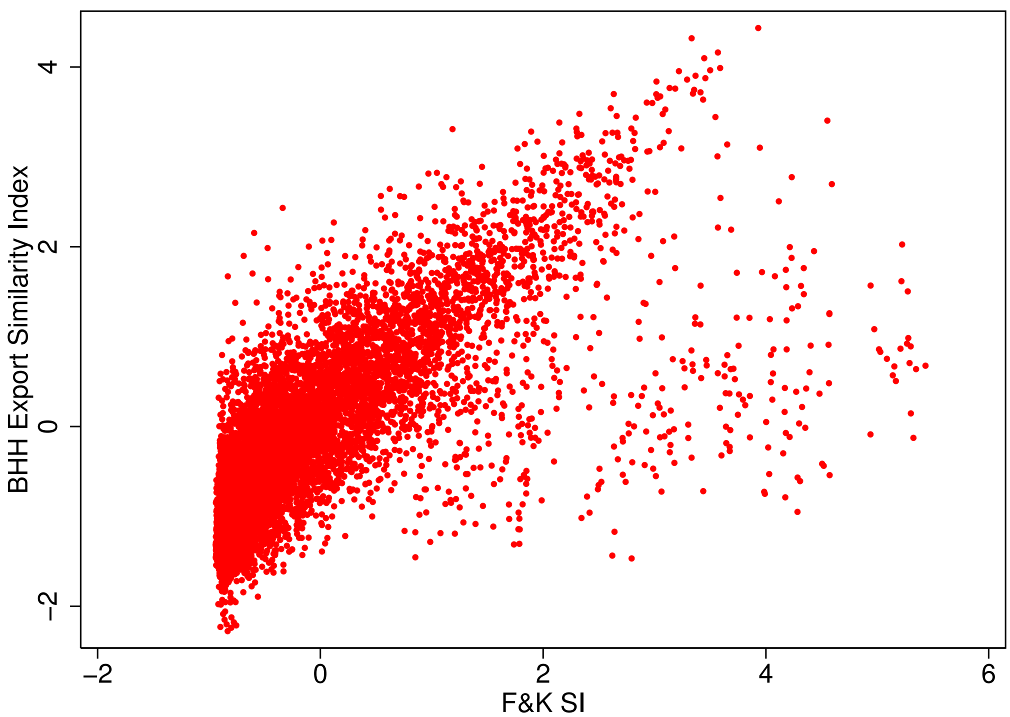
Table A.5: *Correlates of the Indexed RCA Similarity Index*

	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 (1.3) 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.

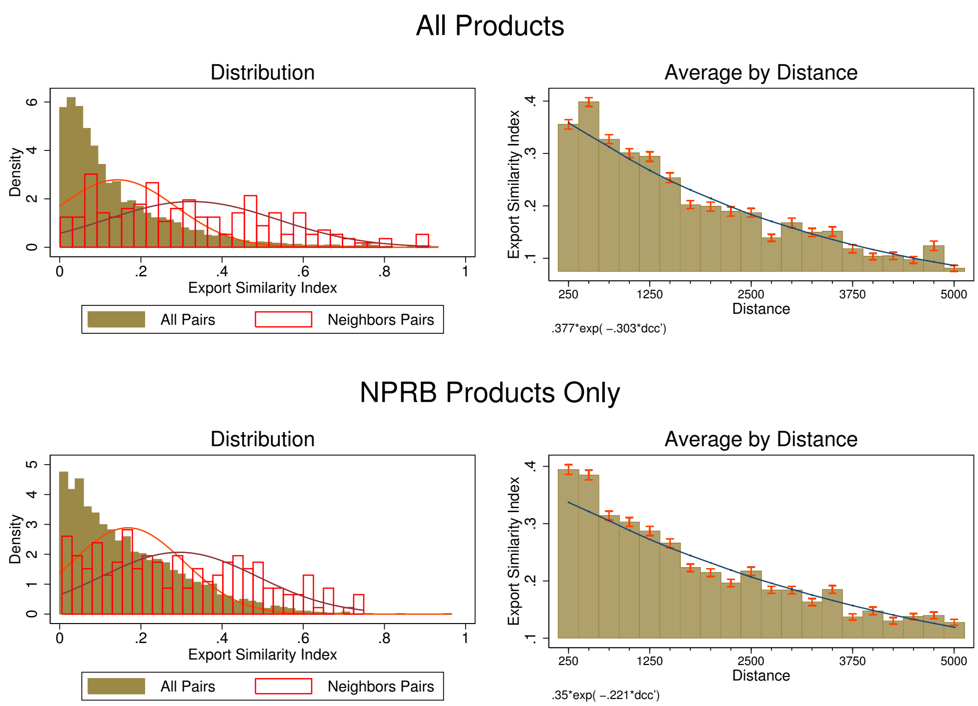
$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

Figure A.8: *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.

Figure A.9: *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 A.6: *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 (1.3) 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.

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

A.3.4 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 also being weighted by the proximity matrix ϕ. In other words, the index would be as follows:

$$S_{c,c'}^{PROX} \equiv \frac{\sum_p (r_{c,p} - \bar{r_c})(r_{c',p} - \bar{r_{c'}})}{\sqrt{\sum_p (r_{c,p} - \bar{r_c})^2 \sum_p (r_{c',p} - \bar{r_{c'}})^2}} \tag{A.1}$$

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,

$$RCA_{c,p}^{PROX} = \frac{\sum_{p'} RCA_{c,p} \times \phi_{p,p'}}{\sum_{p'} \phi_{p,p'}} \tag{A.2}$$

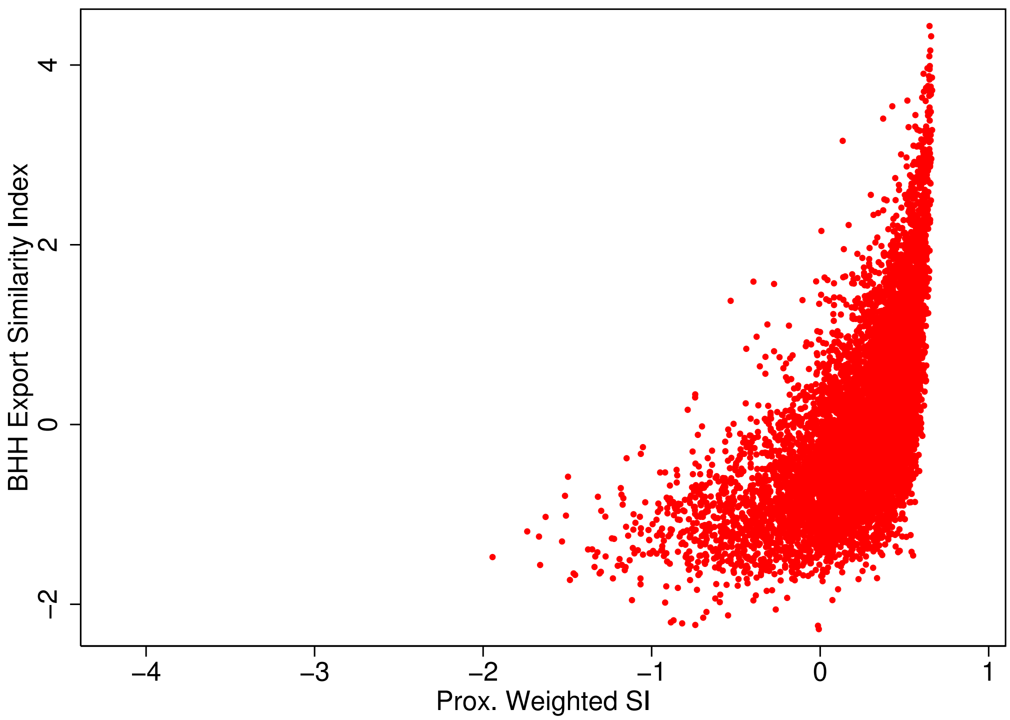
$RCA_{c,p}^{PROX}$ is basically a proximity-weighted RCA measure (similar to the "density" measure developed by Hausmann and Klinger, 2007 and Hidalgo et. al. 2007). Proximity ($\phi_{p,p'}$), a product-product variable, measures the minimum conditional probability of two products being co-exported by 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 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 A.10

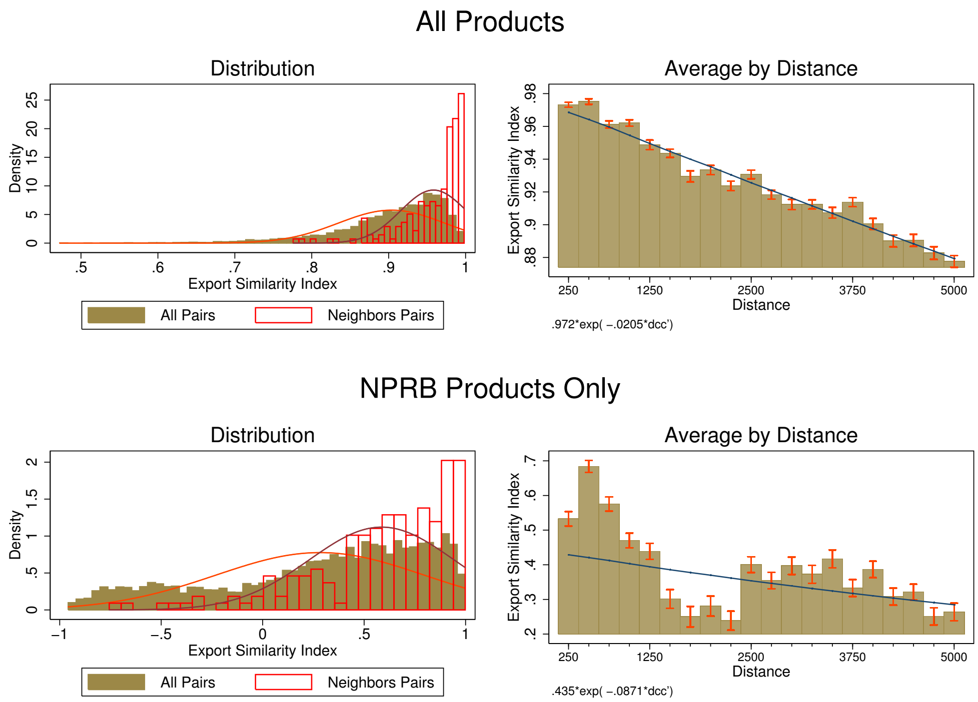
Figure A.11 replicates Figures 1.1 and 1.2 using the proximity weighted similarity index. The upper panel uses all products, while the bottom panel uses NPRB products only. In both

Figure A.10: *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.

Figure A.11: *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.

we can see how the distribution of the proximity weighted similarity index for neighboring countries is shifted to the right. This 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 some discontinuity in the declining relation with distance, but is declining overall.

We turn to study this more in detail by replicating model (1.3), using a normalized version of the Proximity Weighted Similarity Index on the LHS (with mean zero and unit standard deviation). The results are presented in Table A.7. Consistent 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.4 Decomposing Similarity

The similarity observed in the network in Figure A.1 is based on the correlated export of resource-based products for some country-pairs (blue links) and of non-resource-based products for others (red links).

We created a measure to determine whether 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 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:

$$\Delta\sigma_{c,c'} = \sigma_{c,c'}^{NPRB} - \sigma_{c,c'}^{PRB} \tag{A.3}$$

where

$$\sigma_{c,c'}^{NPRB} = \frac{1}{N_{NPRB}} \sum_{p \epsilon NPRB} \delta_{c,c',p} \tag{A.4}$$

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

$$\delta_{c,c',p} = \begin{cases} 1 & \text{if } RCA_{c,p} \geq \overline{RCA_c} \text{ and } RCA_{c',p} \geq \overline{RCA_{c'}} \\ 0 & \text{otherwise} \end{cases} \tag{A.5}$$

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

Table A.7: *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.

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

The definition for $\sigma^{PRB}_{c,c'}$ can be obtained by changing NPRB for PRB in (A.4).

From equation (A.3), $\Delta\sigma_{c,c'} > 0$ 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. For example, Figure A.12 plots Japan and Koreas' RCA in all products in 2008 and shows NRBP products in red and PRB products in blue. The horizontal flat line represents the average RCA for all products for Korea, while the vertical flat line does so for Japan. In this case $\sigma^{NPRB}_{c,c'c} = 0.6517$, $\sigma^{PRB}_{c,c'} = 0.3471$ 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 we are measuring comes mainly from their correlated export of NPRB products.

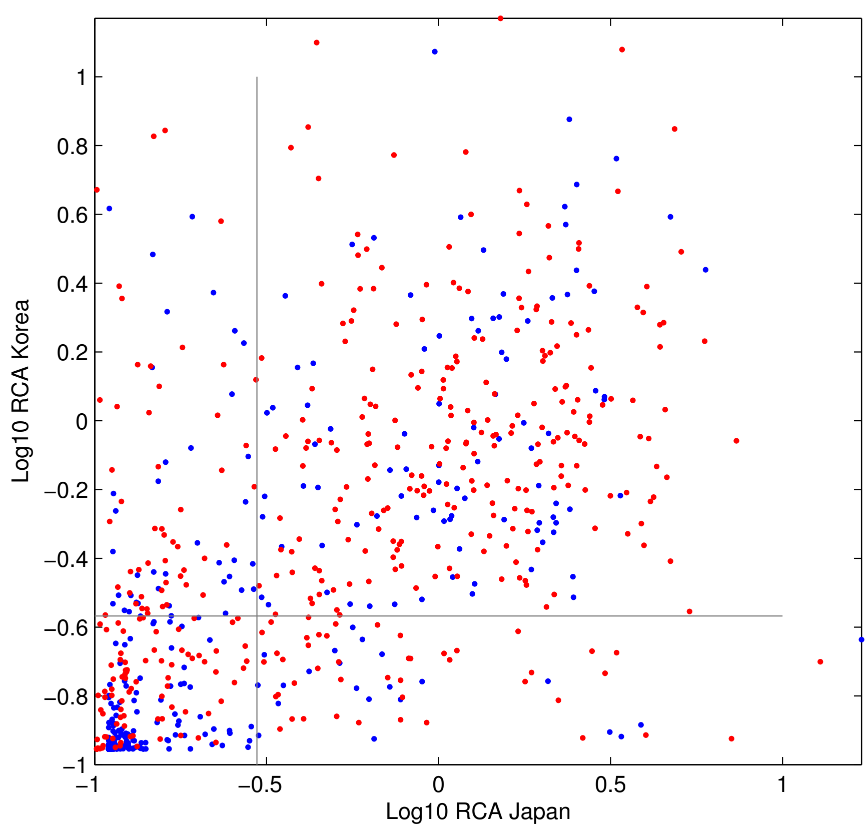
By using these measures we are able to document for any pair of countries whether their export similarity is driven by NPRB or by PRB products. Not all country similarity is driven by the same kind of products. Figure (A.13) summarizes this information by showing, within each region of the World, what proportion of country-pairs are similar due to NPRB products or PRB products.

A.5 Robustness Tests: Dynamics of Export Similarity

The results for the dynamics of export similarity are robust to a number of different specifications of model (1.4). 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 is not statistically different from zero when using the control sample with a random set of neighbors.

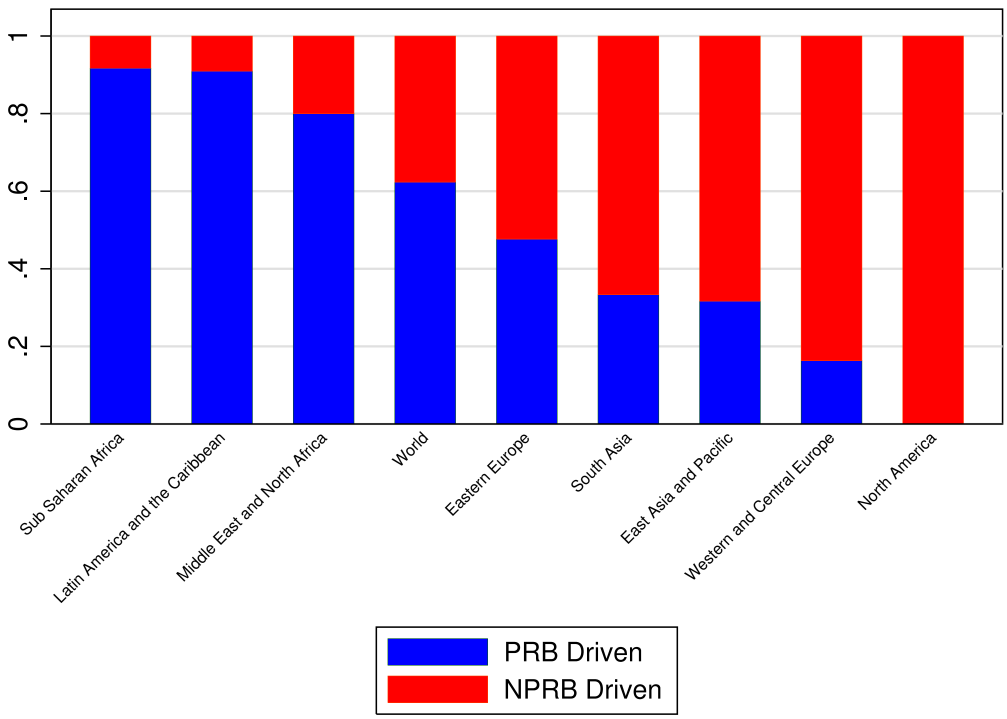
Table A.8 is analogous to Table 1.9 in the main body of the paper, but uses growth in export value as the dependent variable, instead of growth in RCA value. Having a neighbor that exports a product is correlated with an increase in the annual growth of exports for that product 4% to 5%. The results are larger in magnitude (given that the export value is

Figure A.12: *Decomposition of Similarity Index for Korea and Japan in 2008*



The figure shows a scatterplot in which 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 blue otherwise. The data is from year 2008.

Figure A.13: *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.

nominal) but are qualitatively the same.

Table A.9 replicates the results excluding the period 1980-1990 from the sample in order to ensure our results are robust to the changes in the SITC classification in the year 1985.

Our results are also robust to our definition of "jumps": Table A.10 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 A.11 and A.12 redefine the left-hand side of model (1.4): 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 A.13 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 A.14 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 in this classification we also 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 A.8: *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 amount of neighbors per country. All regressions include country-neighbor-by-year and product-by-year fixed effects. Standard errors are clustered at the country-neighbor level.

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

Table A.9: *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 amount of neighbors per country. All regressions include country-neighbor-by-year and product-by-year fixed effects. Standard errors are clustered at the country-neighbor level.

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

Table A.10: *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 amount of neighbors per country. All regressions include country-neighbor-by-year and product-by-year fixed effects. Standard errors are clustered at the country-neighbor level.

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

Table A.11: *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 amount of neighbors per country. All regressions include country-neighbor-by-year and product-by-year fixed effects. Standard errors are clustered at the country-neighbor level.

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

Table A.12: *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 amount of neighbors per country. All regressions include country-neighbor-by-year and product-by-year fixed effects. Standard errors are clustered at the country-neighbor level.

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

Table A.13: *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 amount of neighbors per country. All regressions include country-neighbor-by-year and product-by-year fixed effects. Standard errors are clustered at the country-neighbor level.

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

Table A.14: *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 amount of neighbors per country. All regressions include country-neighbor-by-year and product-by-year fixed effects. Standard errors are clustered at the country-neighbor level.

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

Appendix B

Appendix to Chapter 2

B.1 A Conceptual Framework

Consider a small open economy in a world with a fixed set ω of goods, each one, indexed by i, with the same production function given by:

$$q_i = \varphi_i L_I^{\alpha}$$

Where $0 < \alpha < 1$, L is units of labor, the only factor of production (which is inelastically supplied) and φ_i is product specific productivity. Prices are exogenously determined in world markets, and defined by p_i. Product specific productivity, φ_i, is an increasing function of product-specific tacit knowledge which is equals the amount of non-workers people in the economy, η_i, that have such product specific knowledge. Each economy has an initial endowed vector of η_i (for each product in the set ω) determined exogenously.

Assume the following functional form:

$$\varphi_i = \eta_i^{\rho} \text{ where } 0 < \rho < \frac{1-\alpha}{1+\alpha} < 1 \tag{B.1}$$

Each firm has to pay a fixed cost $C(\varphi) = \frac{1}{\varphi_i^{\frac{1}{1-\alpha}}}$ to enter the market. Hence, the total cost of production is:

$$TC_i = w \left[\frac{q_i}{\varphi_i} \right]^{1/\alpha} + C(\varphi_i)$$

Where w is wage, which is set to 1 for simplicity. Hence, a firm will produce the product only if it can achieve positive profits:

$$\Pi_i = p_i q_i - \left[\frac{q_i}{\varphi_i} \right]^{1/\alpha} - C(\varphi_i) > 0$$

Hence, to have $\Pi_i > 0$ we must have that:

$$\varphi_i > \left[\frac{q_i^{1/\alpha} + 1}{p_i q_i} \right]^{\alpha} \tag{B.2}$$

Assuming equation (B.2) holds, we can compute the optimal amount q_i that a firm will produce given the exogenous price p_i (by equating price to MC):

$$p_i - \frac{1}{\alpha \varphi_i} \left[\frac{q_i}{\varphi_i} \right]^{\frac{1-\alpha}{\alpha}} = 0$$

Solving this results in:

$$q_i^* = (\alpha p_i)^{\frac{\alpha}{1-\alpha}} \varphi_i^{\frac{1+\alpha}{1-\alpha}} = (\alpha p_i)^{\frac{\alpha}{1-\alpha}} \eta_i^{\frac{\rho(1+\alpha)}{1-\alpha}} \tag{B.3}$$

It is easy to see that $\frac{\partial q_i}{\partial \eta_i} > 0$, and $\frac{\partial^2 q_i}{\partial \eta_i^2} < 0$ given the assumptions in (B.1). This brings us to Proposition 1:

Proposition 3 *A firm's decision on the quantity of the product it produces, once such enters the market, is increasing in the amount of knowledge specific to that product available in the country.*

Thus, one testable implication based on Proposition 1, at the country level, is that a country will increase the level of production/exports of certain product whenever it has acquired more knowledge specific to that product.

With the equilibrium amount q_i^* from (B.3), we can find the required amount of knowledge needed for a firm in order to enter the market. Hence, plugging q_i^* into (B.2) we have:

$$\Gamma(\eta_i) = \eta_i - \left[\frac{\eta_i^{\frac{\rho}{1-\alpha}} (\alpha p_i)^{\frac{\alpha}{1-\alpha}} \left[1 + \left(\eta_i^{\frac{\rho}{1-\alpha}} (\alpha p_i)^{\frac{\alpha}{1-\alpha}} \right)^{\frac{1}{\alpha}} \right]}{p_i} \right]^{\frac{\alpha}{\rho}} > 0$$

Therefore, it can be inferred:

$$q_i > 0 \; if \; \Gamma(\eta_i) > 0, \text{ with } \frac{\partial \Gamma(\eta_i)}{\partial \eta_i} > 0$$

Proposition 4 *A product will be produced by a firm in the country if the country has the minimum required amount of specific product knowledge to make it profitable to the firm. The probability of a product being produced by a firm in a country is an increasing function of the number of people in the economy that have the product specific knowledge.*

A second testable implication based on Proposition 2, at the country level, implies that an increase in the amount of specific knowledge on certain product should be positively correlated with the ability of that country to start producing/exporting such product.

$$\frac{\partial q_i}{\partial \eta_i} = \frac{\partial q_i}{\partial \varphi_i} \frac{\partial \varphi_i}{\partial \eta_i} \tag{B.4}$$

The empirical section estimates changes in q_i as a result of variation in η_i. Changes in η_i specific to each product are proxied by exploiting the variation on bilateral relationships (i.e. migration, trade and FDI) from and to countries that export intensively product i.

B.2 Robustness Tests

B.2.1 Excluding products pinned down from geology or climate

In order to limit the sample to products that are not pinned down by geology or climate, we follow the classification provided by Lall (2000), shown in Table B.1. 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

Table B.1: *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

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

The results, presented in Tables B.2, are consistent with the ones shown in the main body of the paper. In the Table, the upper panel shows the results for the extensive margin while the lower panel studies the intensive margin of trade. The results are qualitatively similar to those in Table 2.4. While not robust to all specifications, in most of them immigrants and emigrants correlate to the ability of countries to export new NPRB products. In the intensive margin, immigrants and emigrants appear to have positive and statistically significant coefficients when using the $RCA \geq 5$ threshold, while the evidence is mixed for the $RCA \geq 1$ threshold.

B.2.2 Excluding Bilateral Exports

Kugler and Rapoport (2011) find evidence that migrants reduce transaction costs inducing bilateral trade and capital flows. This section expand on the test presented in the main body of the paper which discard the results being driven by this pattern, instead of a gain in productivity.

As explained in the main body of the paper, we reconstruct the dataset such that the

Table B.2: *Fixed Effects, NPRB products*

	R1		R5	
Panel A: Extensive Margin				
	All	Skilled	All	Skilled
Ln Immigrants	0.0010	0.0019	-0.0004	0.0013
	(0.000)**	(0.001)***	(0.001)	(0.001)
Ln Emigrants	0.0029	0.0028	0.0022	0.0025
	(0.001)**	(0.001)*	(0.001)***	(0.001)***
Ln FDI, total	0.0017	0.0016	0.0011	0.0009
	(0.001)	(0.001)	(0.001)*	(0.001)
Ln Trade, total	-0.0040	-0.0045	0.0007	0.0007
	(0.004)	(0.004)	(0.001)	(0.001)
N	41215	41215	41215	41215
r2	0.21	0.21	0.21	0.21
Panel B: Intensive Margin				
	R1		R5	
	All	Skilled	All	Skilled
Ln Immigrants	0.0014	0.0041	0.0018	0.0017
	(0.001)	(0.001)***	(0.001)**	(0.001)*
Ln Emigrants	0.0048	-0.0031	0.0037	0.0036
	(0.002)**	(0.002)	(0.001)***	(0.001)***
Ln FDI, total	-0.0016	-0.0013	-0.0001	-0.0001
	(0.001)***	(0.001)**	(0.000)	(0.000)
Ln Trade, total	0.0175	0.0227	0.0009	0.0014
	(0.006)***	(0.006)***	(0.001)	(0.001)*
Baseline Log Exports	-0.0446	-0.0446	-0.0436	-0.0433
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
Previous Exports Growth	-0.0853	-0.0850	-0.0885	-0.0896
	(0.016)***	(0.016)***	(0.017)***	(0.017)***
Previous Exports Zero	-0.0310	-0.0310	-0.0266	-0.0258
	(0.016)**	(0.015)**	(0.016)*	(0.016)*
N	42737	42737	42737	42737
r2	0.40	0.40	0.40	0.39

All specifications include country-by-year and product-by-year fixed effects. SE clustered at the country level presented in parenthesis

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

export value for each product and country to the rest of the world excludes exports to nations where migrants are from/in. The number of minimum number of migrants (immigrants + emigrants) used to exclude those bilateral flows are defined in thresholds. Tables B.3 to B.5 present results using as thresholds 1000, 2500 and 5000 migrants, to complement the result in the main body of the paper that presented the 500 threshold (i.e., if there are more than 500 migrants between any two given countries, we exclude all bilateral trade between those two countries, to construct the dependent variable).

Table B.3: *FE, excluding bilateral exports (1000 migrants threshold)*

Panel A: Extensive Margin				
	R1		R5	
	All	Skilled	All	Skilled
Ln Immigrants	0.0017	0.0028	0.0011	0.0012
	(0.000)***	(0.001)***	(0.000)***	(0.001)*
Ln Emigrants	0.0032	0.0024	0.0013	0.0008
	(0.001)***	(0.001)**	(0.000)***	(0.000)*
Ln FDI, total	0.0008	0.0007	-0.0001	-0.0001
	(0.001)	(0.001)	(0.000)	(0.000)
Ln Trade, total	-0.0061	-0.0063	0.0027	0.0028
	(0.004)	(0.004)	(0.001)**	(0.001)**
N	114227	114227	114227	114227
r2	0.11	0.11	0.11	0.11
Panel B: Intensive Margin				
	R1		R5	
	All	Skilled	All	Skilled
Ln Immigrants	0.0005	0.0034	0.0020	0.0015
	(0.001)	(0.002)**	(0.001)**	(0.001)
Ln Emigrants	0.0072	-0.0007	0.0031	0.0030
	(0.002)***	(0.003)	(0.001)***	(0.001)**
Ln FDI, total	-0.0009	-0.0007	0.0010	0.0010
	(0.001)	(0.001)	(0.000)***	(0.000)***
Ln Trade, total	0.0124	0.0175	0.0016	0.0022
	(0.006)**	(0.006)***	(0.001)*	(0.001)**
Baseline Log Exports	-0.0072	-0.0072	-0.0062	-0.0058
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
Previous Exports Growth	-0.1045	-0.1041	-0.1077	-0.1088
	(0.017)***	(0.017)***	(0.017)***	(0.017)***
Previous Exports Zero	-0.0691	-0.0695	-0.0639	-0.0635
	(0.016)***	(0.016)***	(0.016)***	(0.016)***
N	57956	57956	57956	57956
r2	0.32	0.32	0.32	0.32

All specifications include country-by-year and product-by-year fixed effects. The dependent variable in all specifications is constructed using exports of country c to the whole world excluding to countries c' where total migration between c and c' exceeds 1000 people.

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

Table B.4: *FE, excluding bilateral exports (2500 migrants threshold)*

Panel A: Extensive Margin				
	R1		R5	
	All	Skilled	All	Skilled
Ln Immigrants	0.0018	0.0028	0.0009	0.0009
	(0.000)***	(0.001)***	(0.000)**	(0.001)
Ln Emigrants	0.0029	0.0024	0.0013	0.0012
	(0.001)***	(0.001)**	(0.000)***	(0.001)**
Ln FDI, total	0.0009	0.0008	-0.0002	-0.0002
	(0.001)	(0.001)	(0.000)	(0.000)
Ln Trade, total	-0.0063	-0.0065	0.0023	0.0023
	(0.004)	(0.004)	(0.001)*	(0.001)*
N	106881	106881	106881	106881
r2	0.11	0.11	0.11	0.11
Panel B: Intensive Margin				
	R1		R5	
	All	Skilled	All	Skilled
Ln Immigrants	0.0012	0.0029	0.0020	0.0014
	(0.001)	(0.001)**	(0.001)**	(0.001)
Ln Emigrants	0.0067	-0.0009	0.0034	0.0030
	(0.002)***	(0.003)	(0.001)***	(0.001)**
Ln FDI, total	-0.0006	-0.0004	0.0007	0.0008
	(0.001)	(0.001)	(0.000)*	(0.000)**
Ln Trade, total	0.0094	0.0157	0.0017	0.0024
	(0.006)*	(0.006)***	(0.001)**	(0.001)***
Baseline Log Exports	-0.0126	-0.0124	-0.0118	-0.0114
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
Previous Exports Growth	-0.1070	-0.1072	-0.1096	-0.1109
	(0.016)***	(0.016)***	(0.016)***	(0.016)***
Previous Exports Zero	-0.0610	-0.0611	-0.0567	-0.0561
	(0.015)***	(0.015)***	(0.015)***	(0.015)***
N	61476	61476	61476	61476
r2	0.31	0.31	0.31	0.31

All specifications include country-by-year and product-by-year fixed effects. The dependent variable in all specifications is constructed using exports of country c to the whole world excluding to countries c' where total migration between c and c' exceeds 2500 people.

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

Table B.5: *FE, excluding bilateral exports (5000 migrants threshold)*

Panel A: Extensive Margin				
	R1		R5	
	All	Skilled	All	Skilled
Ln Immigrants	0.0020	0.0026	0.0009	0.0009
	(0.000)***	(0.001)***	(0.000)**	(0.001)
Ln Emigrants	0.0026	0.0023	0.0016	0.0015
	(0.001)***	(0.001)**	(0.000)***	(0.001)***
Ln FDI, total	0.0009	0.0009	-0.0000	-0.0000
	(0.001)	(0.001)	(0.000)	(0.000)
Ln Trade, total	-0.0068	-0.0069	0.0020	0.0021
	(0.004)	(0.004)	(0.001)	(0.001)
N	100988	100988	100988	100988
r2	0.12	0.12	0.12	0.12
Panel B: Intensive Margin				
	R1		R5	
	All	Skilled	All	Skilled
Ln Immigrants	0.0017	0.0040	0.0020	0.0015
	(0.001)	(0.001)***	(0.001)***	(0.001)
Ln Emigrants	0.0062	-0.0025	0.0039	0.0035
	(0.002)***	(0.003)	(0.001)***	(0.001)***
Ln FDI, total	-0.0009	-0.0007	0.0005	0.0005
	(0.000)*	(0.000)	(0.000)	(0.000)
Ln Trade, total	0.0087	0.0152	0.0017	0.0024
	(0.005)	(0.005)***	(0.001)**	(0.001)***
Baseline Log Exports	-0.0165	-0.0163	-0.0158	-0.0153
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
Previous Exports Growth	-0.1060	-0.1060	-0.1083	-0.1097
	(0.016)***	(0.016)***	(0.016)***	(0.016)***
Previous Exports Zero	-0.0549	-0.0551	-0.0508	-0.0503
	(0.015)***	(0.015)***	(0.015)***	(0.015)***
N	64343	64343	64343	64343
r2	0.32	0.32	0.32	0.32

All specifications include country-by-year and product-by-year fixed effects. The dependent variable in all specifications is constructed using exports of country c to the whole world excluding to countries c' where total migration between c and c' exceeds 5000 people.

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

Appendix C

Appendix to Chapter 3

C.1 Condition for $\partial\phi(a_I)/\partial d > 0$

Horizontal FDI will be less profitable at longer distances if $\partial\phi(a_I)/\partial d > 0$ (where $\phi(a_I) = a_I^{1-\varepsilon}$). For simplicity, I compute the conditions for which $\partial log(\phi(a_I))/\partial d > 0$ instead:

$$\frac{\varepsilon - 1}{\kappa(k,d)^{1-\varepsilon} - \tau(t,d)^{1-\varepsilon}} \cdot \left[\kappa(k,d)^{-\varepsilon}\frac{\partial\kappa}{\partial d} - \tau(t,d)^{-\varepsilon}\frac{\partial\tau}{\partial d}\right] > 0$$

Given that the left term will always be positive (given the assumption that $\tau(t,d) > \kappa(k,d)$ and $\varepsilon > 1$, the conditions for the inequality to hold are derived from the right term only. Hence, we have:

$$\begin{aligned} \kappa(k,d)^{-\varepsilon} \cdot \frac{\partial\kappa}{\partial d} &> \tau(t,d)^{-\varepsilon} \cdot \frac{\partial\tau}{\partial d} \\ \frac{1}{\kappa(k,d)^{\varepsilon-1}} \cdot \epsilon_{\kappa,d} &> \frac{1}{\tau(t,d)^{1-\varepsilon}} \cdot \epsilon_{\tau,d} \end{aligned}$$

Where $\epsilon_{\kappa,d}$ and $\epsilon_{\tau,d}$ are the elasticity of κ and τ with respect to distance d, respectively.

Hence for the condition to hold it must be that:

$$\left[\frac{\tau(t,d)}{\kappa(k,d)}\right]^{\varepsilon-1} > \frac{\epsilon_{\tau,d}}{\epsilon_{\kappa,d}}$$

There is no reason to believe that this condition is not economically feasible.

C.2 Heterogeneity in the number of reported SIC industries in the dataset

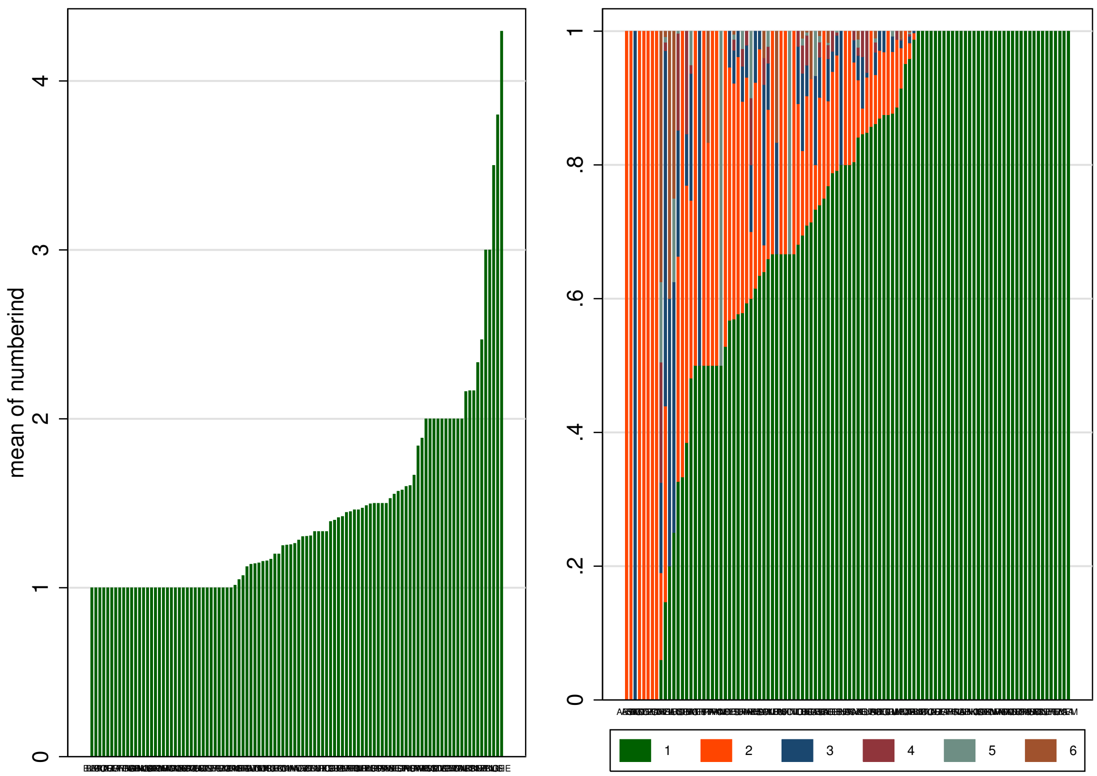
While the dataset has information on up to six industries per plant (a main one plus five other) the number of establishments that report more than one activity varies dramatically per country. The left panel of Figure C.1 shows the average number of reported industries across all subsidiaries per country, while the right panel shows, per country, the percentage of firms reporting one, two, three, four, five or six industries. In most countries, the average number of reported firms is below two; and the majority of firms in more than half the countries report only one SIC code.

C.3 Using the input-output table to define vertical relationships

In order to filter out from the definition of horizontal those links that could also be defined as vertical, either upstream or downstream, I use the US input-output provided by Fan and Lang (2000). I follow the methodology suggested by Alfaro and Charlton (2009) and Acemoglu et. al. (2009) to define vertical relationships.

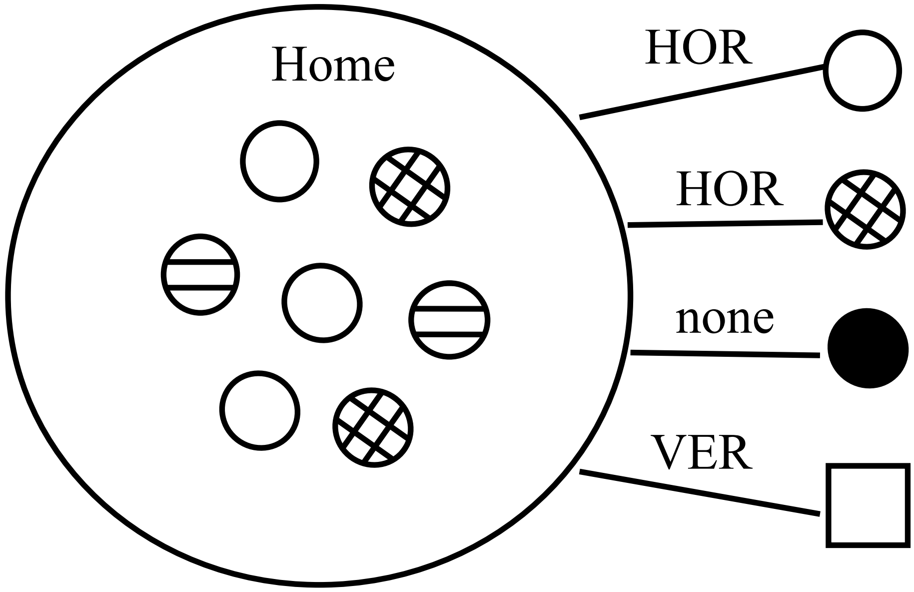
More in general, the diagram in Figure C.2 is useful to understand how horizontal and vertical links are defined in the dataset. Within a single MNC firm, an horizontal link is defined as a foreign subsidiary that is classified under the same SIC code as any of its domestic subsidiaries. Then I use the US I/O table by Fan & Lang (2000) to define vertical relationships, both downstream and upstream. A subsidiary is defined as upstream vertical if its main economic activity is an input of \$0.05 or more per each dollar of output of any

Figure C.1: *Distribution of reported SIC codes by plant, per country*



The figure describe the distribution of number of industries reported by establishment in the sample. The left panel shows the average number of reported industries across all subsidiaries per country, while the right panel shows, per country, the percentage of firms reporting one, two, three, four, five or six industries.

Figure C.2: *Definition of Horizontal and Vertical*



The diagram describes the methodology used to classify foreign subsidiaries as horizontal expansions based on their reported economic activity vis-a-vis the economic activity of the MNC in its home country.

of the domestic subsidiaries of the firm. Similarly, a subsidiary is defined as downstream vertical if any of the domestic subsidiary provides an input to it of \$0.05 or more per each dollar of output.

After such classification, those subsidiaries that fall into both categories (horizontal and vertical) are filtered out from the horizontal classification. This implies that the sample classifies as horizontal only final goods, which is the matter of study of the theoretical framework presented.

Appendix Section C.6.1 presents robustness tests of all tables using alternative thresholds (0.01 and 0.10). The use of \$0.05 in the main body of the paper follows the precedent set by Alfaro & Charlton (2009).

A limitation of this methodology is that technologies might vary across countries, and hence, the US I/O table would loss some validity in defining upstream or downstream relationships. While acknowledging this limitation I assume that the US I/O table is a good proxy for measuring vertical links, regardless of the country, in line with the previous literature.

C.4 Limitations of the R&D intensity measures

Nunn & Trefler (2008) and Keller & Yeaple (2013) use the average R&D share of firms' sales as their measure of knowledge intensity. Nunn & Trefler use firm-level data from Orbis, while Keller & Yeaple use data from COMPUSTAT.

The two measures are skewed towards the few industries with large R&D investment, while the zeros or very small values are highly abundant (see Figure C.3). In fact, for Nunn & Trefler half of the industries have an R&D intensity measure below 0.2%, while the largest industries have a value of 190%. In Keller & Yeaple's measure the median is 0.7% while the most knowledge intensive industry has a share of R&D over sales of over 1000%.

C.5 O*NET knowledge intensity measures

Figure C.4 presents the distribution of the knowledge intensity measure used in the paper: experience plus training (based on experience plus on-site and on-the-job training figures of workers in each industry). As opposed to the R&D investment based variables used in the literature (see Section C.4), the distribution of the O*NET based variables is smoother, and behaves more like a normal probability density function. Figure C.5 presents the same graphs limiting the sample to manufacturing industries only.

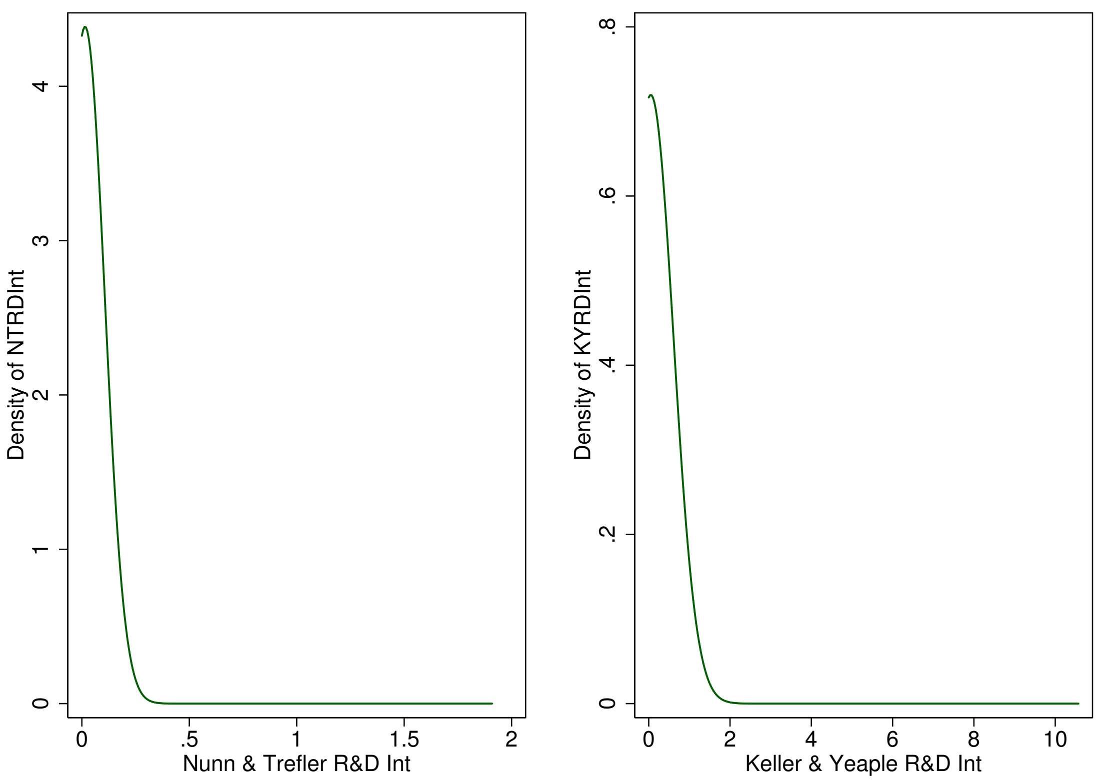
Tables C.1 presents the top and bottom ten products in the manufacturing division (SIC codes 2000 to 3999) ranked by the knowledge intensity measure.

C.6 Robustness Tests

C.6.1 Varying thresholds in the definition of horizontal subsidiaries

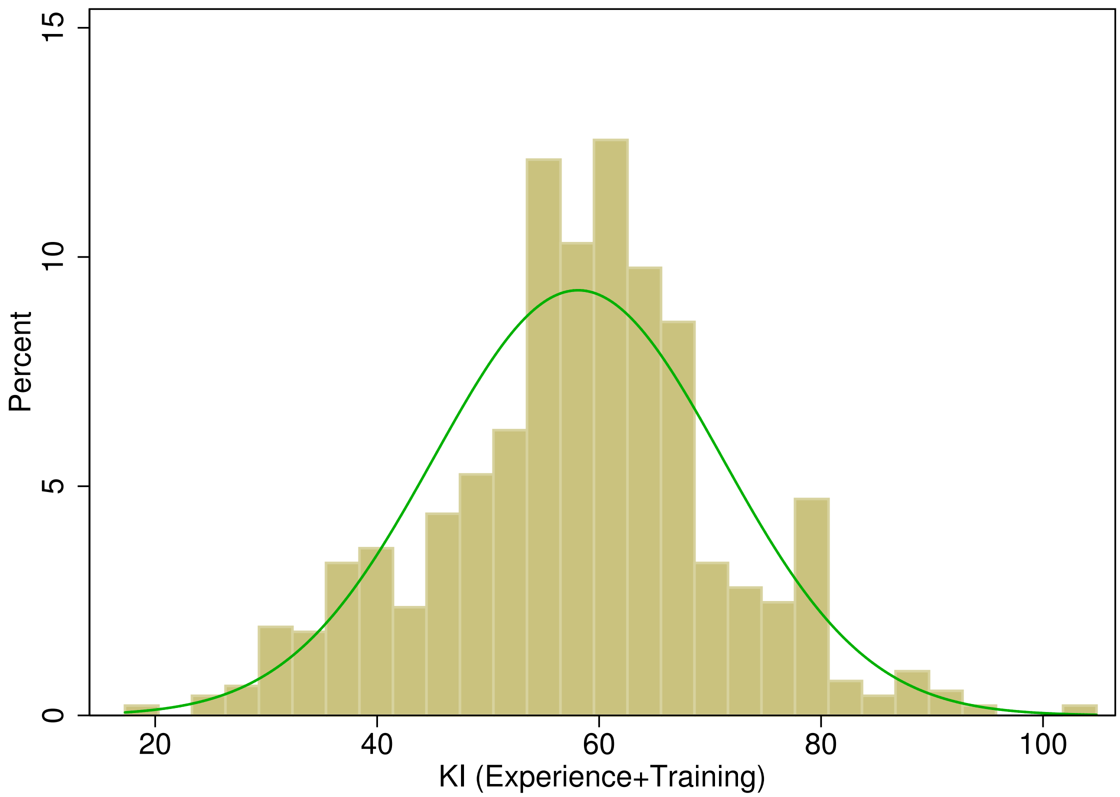
As explained in Section C.3, subsidiaries that classify both as horizontal and vertical (according to the I/O table) are not considered horizontal. The intuition for such approach

Figure C.3: *Fitted distribution of R&D Measures*



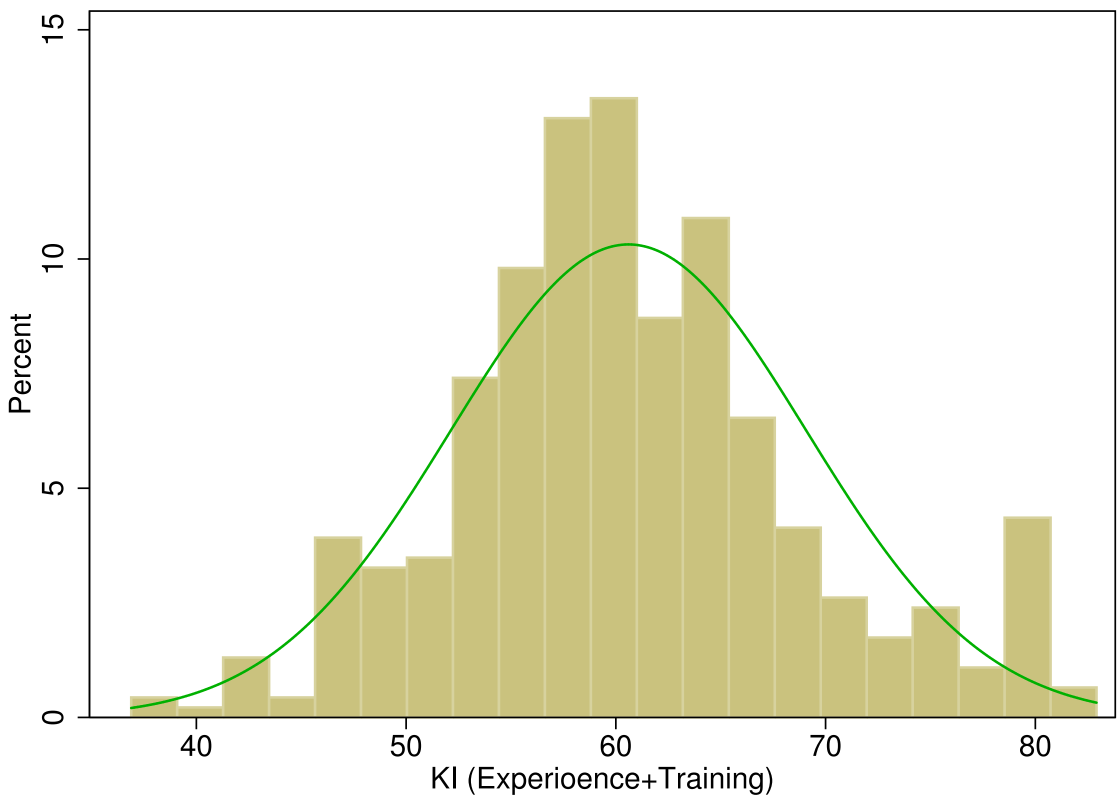
The figure shows the fitted distribution for the industry level R&D investment as share of sales, compiled from firm level datasets by Nunn & Trefler (2008) and Keller & Yeaple (2013) in the left and right panel respectively.

Figure C.4: *Histogram O*NET-based KI (All Industries)*



The figure shows the fitted distribution for the computed "experience plus training" O*NET-based knowledge intensity measures for all industries. Industries are defined in SIC 1987 4-digit industries.

Figure C.5: *Histogram O*NET-based KI (Manufacturing Only)*



The figure shows the fitted distribution for the computed "experience plus training" O*NET-based knowledge intensity measures for manufacturing industries only. Industries are defined in SIC 1987 4-digit industries.

Table C.1: *Top and bottom 10 manufacturing products, ranked by KI*

Rank	SIC	Name	Value
		Ranking by Experience + Training, Top 10	
1	3669	Communications Equipment, NEC	82.92
2	3663	Radio and Television Broadcasting and Communications Equipment	82.92
3	3661	Telephone and Telegraph Apparatus (except consumer external modems)	81.45
4	3677	Electronic Coils, Transformers, and Other Inductors	79.97
5	3676	Electronic Resistors	79.97
6	3678	Electronic Connectors	79.97
7	3675	Electronic Capacitors	79.97
8	3671	Electron Tubes	79.97
9	3672	Printed Circuit Boards	79.97
10	3674	Semiconductors and Related Devices	79.97
		Ranking by Experience + Training, Bottom 10	
459	2013	Sausages and Other Prepared Meat Products (except lard made from purchased materials)	36.89
458	2011	Meat Packing Plants	36.89
457	2411	Logging	39.79
456	2077	Animal and Marine Fats and Oils (animal fats and oils)	41.39
455	2053	Frozen Bakery Products, Except Bread	41.53
454	2045	Prepared Flour Mixes and Doughs	41.53
453	2098	Macaroni, Spaghetti, Vermicelli and Noodles	41.53
452	2051	Bread and Other Bakery Products, Except Cookies and Crackers	41.53
451	2015	Poultry Slaughtering and Processing (poultry slaughtering and processing)	42.72
450	2052	Cookies and Crackers (unleavened bread and soft pretzels)	45.04

The table presents the top and bottom 10 manufacturing sectors ranked by the "experience plus training" O*NET based knowledge intensity measure.

is to limit the analysis of horizontal to final goods only.

To do so, a threshold of \$0.05 per each \$ of output, was selected in order to define vertical relationships. This section presents the robustness test varying such threshold, for all tables in the main body of the paper.

Tables C.2-C.4 replicate all results using threshold 0.01, while tables C.5-C.7 replicate all results using the threshold 0.1.

Varying the input-output threshold is robust to the results presented in the main body of the paper.

C.6.2 Additional measures of knowledge intensity

In the main body of the paper I perform the analysis using one constructed measure of knowledge intensity denominated *experience plus training.* In this section I use instead a modification of such measure which only takes into account the accumulated experience of the workers in the industry (excluding the on-site and on-the-job training component). The results are robust to this other measure as can be seen in Tables C.8, C.9 and C.10, as well as in Figure C.6.

C.6.3 Non-linear effects of distance

Is the negative relationship between distance and the likelihood of a foreign subsidiary being horizontal linear? I test for that substituting in the estimation of specification 3.8 the continuos measure of distance ($log(d)$) by a set of dummies, each one representing a 500km interval in the distance between the headquarters and the foreign subsidiary. The results are presented in table C.11. As it can be seen, the negative correlation becomes larger in magnitude the further away the headquarters is from the location of the foreign subsidiary.

The results suggest that up to 8000Km the correlation between distance and the existence of an horizontal foreign subsidiary is negative and increasing in magnitude (besides the 3000-4000km bucket, which present a positive, though non-statistically significant coefficient). Only after 8000 km the coefficients are reduced in terms of magnitude, while still negative.

Table C.2: *Determinants of Foreign Replication of Production (threshold 0.01)*

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0416	-0.0233	-0.0162	-0.0405
	(0.016)**	(0.016)	(0.012)	(0.013)***
log(t)	-0.0553			-0.0512
	(0.020)***			(0.017)***
GDP per capita ratio	-0.4123		-0.3943	-0.0361
	(0.177)**		(0.173)**	(0.061)
Population ratio	0.0621		0.0635	0.0641
	(0.025)**		(0.024)***	(0.016)***
Capital per worker ratio	0.2818		0.2769	0.0482
	(0.076)***		(0.075)***	(0.035)
Human Capital ratio	1.1475		1.1184	0.0000
	(0.237)***		(0.230)***	(.)
Land per worker ratio	-0.1017		-0.0985	0.1951
	(0.022)***		(0.022)***	(0.039)***
Constant	0.1114	0.2419	0.2133	0.7139
	(0.038)***	(0.008)***	(0.007)***	(0.070)***
N	47657	50146	50096	47657
R-squared	0.52	0.41	0.52	0.57
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	Y

The table presents results for the estimation of Specification (3.7) using a sample of domestic and foreign subsidiaries that replicate home production. The left hand side variable is a binary variable that takes the value 1 if the subsidiary is foreign. The variables in the right hand side include the unit shipping cost associated with the industry, knowledge intensity measures (in standard deviations from the mean) and other controls. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parentheses.

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

Table C.3: *Determinants of Horizontal FDI (threshold 0.01)*

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0859	-0.0855		-0.0870
	(0.030)***	(0.030)***		(0.030)***
log(d)	-0.0202		-0.0199	-0.0190
	(0.006)***		(0.006)***	(0.006)***
log(t)	-0.0308	-0.0302	0.0063	-0.0306
	(0.043)	(0.043)	(0.037)	(0.043)
GDP per capita ratio	0.0909	0.0920	0.0894	0.7052
	(0.050)*	(0.051)*	(0.050)*	(0.196)***
Population ratio	-0.0072	-0.0004	-0.0058	0.2853
	(0.005)	(0.005)	(0.006)	(0.072)***
Capital per worker ratio	-0.0733	-0.0876	-0.0693	-0.3497
	(0.040)*	(0.042)**	(0.041)*	(0.117)***
Human Capital ratio	-0.0150	0.0202	-0.0195	-0.7377
	(0.055)	(0.053)	(0.055)	(0.303)**
Land per worker ratio	-0.0177	-0.0157	-0.0178	0.0634
	(0.007)***	(0.007)**	(0.007)***	(0.035)*
Constant	0.3534	0.2002	0.4005	0.2565
	(0.092)***	(0.080)**	(0.085)***	(0.151)*
N	55136	55137	55136	55136
R-squared	0.48	0.48	0.48	0.49
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	Y

The table presents results for the estimation of Specification (3.8) using a sample of foreign subsidiaries of MNCs. The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as an horizontal expansion. The variables in the right hand side include the distance from the MNC headquarters to the foreign subsidiary, the unit shipping cost, knowledge intensity measures (in standard deviations from the mean) and other controls. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parenthesis.

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

Table C.4: *Ease of Communication (threshold 0.01)*

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0859	-0.0861	-0.0858	-0.0857
	(0.030)***	(0.030)***	(0.030)***	(0.030)***
log(d)	-0.0202	-0.0219	-0.0177	-0.0146
	(0.006)***	(0.006)***	(0.010)*	(0.005)***
log(t)	-0.0308	-0.0310	-0.0307	-0.0308
	(0.043)	(0.043)	(0.043)	(0.043)
Non-stop Flight		-0.0122		
		(0.009)		
Working hours overlap			0.0013	
			(0.004)	
Common Language				0.0500
				(0.024)**
GDP per capita ratio	0.0909	0.0916	0.0905	0.0890
	(0.050)*	(0.050)*	(0.051)*	(0.049)*
Population ratio	-0.0072	-0.0068	-0.0073	-0.0062
	(0.005)	(0.005)	(0.005)	(0.005)
Capital per worker ratio	-0.0733	-0.0745	-0.0742	-0.0700
	(0.040)*	(0.040)*	(0.039)*	(0.038)*
Human Capital ratio	-0.0150	-0.0108	-0.0147	-0.0205
	(0.055)	(0.054)	(0.055)	(0.055)
Land per worker ratio	-0.0177	-0.0167	-0.0170	-0.0121
	(0.007)***	(0.007)**	(0.006)***	(0.007)*
Constant	0.3534	0.3705	0.3249	0.2985
	(0.092)***	(0.093)***	(0.138)**	(0.086)***
N	55136	55136	55136	55132
R-squared	0.48	0.48	0.48	0.49
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	N

The table presents results for the estimation of Specification (3.8) using a sample of foreign subsidiaries of MNCs. The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as an horizontal expansion. The variables in the right hand side include the distance from the MNC headquarters to the foreign subsidiary, the unit shipping cost, knowledge intensity measures (in standard deviations from the mean) and other controls. The right hand side also includes variables measuring the ease of communication between a headquarters and its subsidiaries. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parenthesis.

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

Table C.5: *Determinants of Foreign Replication of Production (threshold 0.1)*

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0334	-0.0416	-0.0255	-0.0324
	(0.016)**	(0.015)***	(0.012)**	(0.014)**
log(t)	-0.0111			-0.0077
	(0.021)			(0.020)
GDP per capita ratio	-0.3745		-0.3818	0.4705
	(0.124)***		(0.122)***	(0.124)***
Population ratio	0.0894		0.0914	-0.0498
	(0.018)***		(0.017)***	(0.025)**
Capital per worker ratio	0.3138		0.3202	-0.2595
	(0.077)***		(0.075)***	(0.066)***
Human Capital ratio	0.9593		0.9491	-0.0210
	(0.167)***		(0.163)***	(0.069)
Land per worker ratio	-0.0994		-0.0968	0.1379
	(0.017)***		(0.017)***	(0.041)***
Constant	0.2536	0.3079	0.2695	0.8667
	(0.041)***	(0.008)***	(0.007)***	(0.099)***
N	65058	68293	68207	65058
R-squared	0.52	0.39	0.51	0.55
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	Y

The table presents results for the estimation of Specification (3.7) using a sample of domestic and foreign subsidiaries that replicate home production. The left hand side variable is a binary variable that takes the value 1 if the subsidiary is foreign. The variables in the right hand side include the unit shipping cost associated with the industry, knowledge intensity measures (in standard deviations from the mean) and other controls. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parentheses.

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

Table C.6: *Determinants of Horizontal FDI (threshold 0.1)*

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0937	-0.0932		-0.0957
	(0.042)**	(0.042)**		(0.042)**
log(d)	-0.0244		-0.0241	-0.0216
	(0.009)***		(0.009)***	(0.010)**
log(t)	0.0489	0.0496	0.0893	0.0484
	(0.062)	(0.062)	(0.052)*	(0.061)
GDP per capita ratio	0.1267	0.1281	0.1250	0.9220
	(0.055)**	(0.056)**	(0.056)**	(0.246)***
Population ratio	0.0153	0.0236	0.0169	0.3181
	(0.007)**	(0.007)***	(0.007)**	(0.078)***
Capital per worker ratio	-0.0837	-0.1010	-0.0794	-0.6704
	(0.045)*	(0.046)**	(0.046)*	(0.187)***
Human Capital ratio	0.0068	0.0494	0.0019	-0.4480
	(0.062)	(0.062)	(0.063)	(0.285)
Land per worker ratio	-0.0169	-0.0144	-0.0169	0.0110
	(0.007)**	(0.007)**	(0.007)**	(0.055)
Constant	0.7121	0.5271	0.7636	0.8320
	(0.151)***	(0.127)***	(0.138)***	(0.222)***
N	55136	55137	55136	55136
R-squared	0.48	0.48	0.48	0.48
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	Y

The table presents results for the estimation of Specification (3.8) using a sample of foreign subsidiaries of MNCs. The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as an horizontal expansion. The variables in the right hand side include the distance from the MNC headquarters to the foreign subsidiary, the unit shipping cost, knowledge intensity measures (in standard deviations from the mean) and other controls. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parenthesis.

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

Table C.7: *Ease of Communication (threshold 0.1)*

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0937	-0.0938	-0.0928	-0.0935
	(0.042)**	(0.042)**	(0.042)**	(0.042)**
log(d)	-0.0244	-0.0253	-0.0077	-0.0197
	(0.009)***	(0.009)***	(0.014)	(0.008)**
log(t)	0.0489	0.0488	0.0495	0.0489
	(0.062)	(0.062)	(0.062)	(0.062)
Non-stop Flight		-0.0063		
		(0.010)		
Working hours overlap			0.0092	
			(0.005)**	
Common Language				0.0420
				(0.024)*
GDP per capita ratio	0.1267	0.1270	0.1240	0.1251
	(0.055)**	(0.056)**	(0.055)**	(0.054)**
Population ratio	0.0153	0.0155	0.0149	0.0162
	(0.007)**	(0.007)**	(0.007)**	(0.007)**
Capital per worker ratio	-0.0837	-0.0843	-0.0896	-0.0809
	(0.045)*	(0.045)*	(0.044)**	(0.044)*
Human Capital ratio	0.0068	0.0089	0.0087	0.0021
	(0.062)	(0.061)	(0.062)	(0.063)
Land per worker ratio	-0.0169	-0.0164	-0.0117	-0.0121
	(0.007)**	(0.007)**	(0.007)*	(0.007)*
Constant	0.7121	0.7210	0.5171	0.6661
	(0.151)***	(0.150)***	(0.206)**	(0.149)***
N	55136	55136	55136	55132
R-squared	0.48	0.48	0.48	0.48
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	N

The table presents results for the estimation of Specification (3.8) using a sample of foreign subsidiaries of MNCs. The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as an horizontal expansion. The variables in the right hand side include the distance from the MNC headquarters to the foreign subsidiary, the unit shipping cost, knowledge intensity measures (in standard deviations from the mean) and other controls. The right hand side also includes variables measuring the ease of communication between a headquarters and its subsidiaries. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parenthesis.

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

Table C.8: *Determinants of Foreign Replication of Production (KI: experience)*

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0300	-0.0291	-0.0186	-0.0291
	(0.014)**	(0.012)**	(0.010)*	(0.012)**
log(t)	-0.0244			-0.0206
	(0.024)			(0.022)
GDP per capita ratio	-0.3948		-0.4005	0.3856
	(0.131)***		(0.128)***	(0.131)***
Population ratio	0.0849		0.0866	-0.0673
	(0.019)***		(0.019)***	(0.028)**
Capital per worker ratio	0.3295		0.3323	-0.2326
	(0.080)***		(0.078)***	(0.069)***
Human Capital ratio	0.9535		0.9500	0.0290
	(0.180)***		(0.176)***	(0.072)
Land per worker ratio	-0.1030		-0.0995	0.1095
	(0.018)***		(0.018)***	(0.045)**
Constant	0.2214	0.2972	0.2620	0.9514
	(0.046)***	(0.009)***	(0.007)***	(0.103)***
N	61410	64462	64389	61410
R-squared	0.52	0.40	0.51	0.56
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	Y

The table presents results for the estimation of Specification (3.7) using a sample of domestic and foreign subsidiaries that replicate home production. It uses an O*NET-based indicator for knowledge intensity based on workers' accumulated experience (excluding training). The left hand side variable is a binary variable that takes the value 1 if the subsidiary is foreign. The variables in the right hand side include the unit shipping cost associated with the industry, knowledge intensity measures (in standard deviations from the mean) and other controls. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parentheses.

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

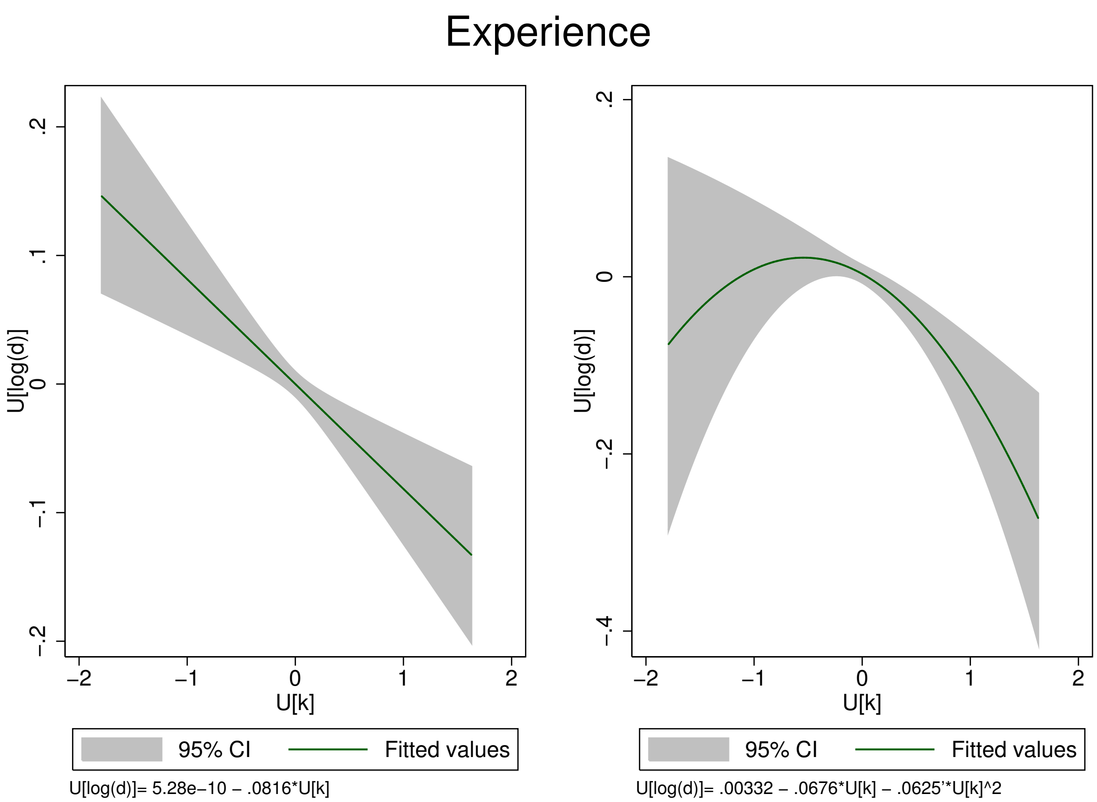
Table C.9: *Determinants of Horizontal FDI (KI: experience)*

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0628	-0.0625		-0.0645
	(0.037)*	(0.037)*		(0.037)*
log(d)	-0.0241		-0.0239	-0.0230
	(0.009)**		(0.009)**	(0.010)**
log(t)	0.0275	0.0281	0.0611	0.0271
	(0.067)	(0.067)	(0.056)	(0.066)
GDP per capita ratio	0.1309	0.1323	0.1282	0.8952
	(0.056)**	(0.057)**	(0.056)**	(0.241)***
Population ratio	0.0131	0.0212	0.0142	0.3121
	(0.007)*	(0.007)***	(0.007)**	(0.077)***
Capital per worker ratio	-0.0836	-0.1007	-0.0792	-0.6494
	(0.046)*	(0.047)**	(0.046)*	(0.180)***
Human Capital ratio	-0.0071	0.0350	-0.0098	-0.4310
	(0.063)	(0.063)	(0.064)	(0.283)
Land per worker ratio	-0.0132	-0.0107	-0.0131	0.0122
	(0.007)*	(0.008)	(0.007)*	(0.053)
Constant	0.6318	0.4487	0.6745	0.7556
	(0.162)***	(0.136)***	(0.148)***	(0.227)***
N	55136	55137	55136	55136
R-squared	0.47	0.47	0.47	0.47
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	Y

The table presents results for the estimation of Specification (3.8) using a sample of foreign subsidiaries of MNCs. It uses an O*NET-based indicator for knowledge intensity based on workers' accumulated experience (excluding training). The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as an horizontal expansion. The variables in the right hand side include the distance from the MNC headquarters to the foreign subsidiary, the unit shipping cost, knowledge intensity measures (in standard deviations from the mean) and other controls. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parenthesis.

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

Figure C.6: $U[k]$ *vs.* $U[log(d)]$, *(KI: experience)*



The figure presents the empirical fit for the relationship between d and k (the latter proxied by the *experience* measure). The left column performs a linear fit between k and d while the right column performs a quadratic firm between the two. The grey area represents the 95% confidence interval for the estimated relationship. The sample excludes foreign subsidiaries located in Europe owned by a European firm.

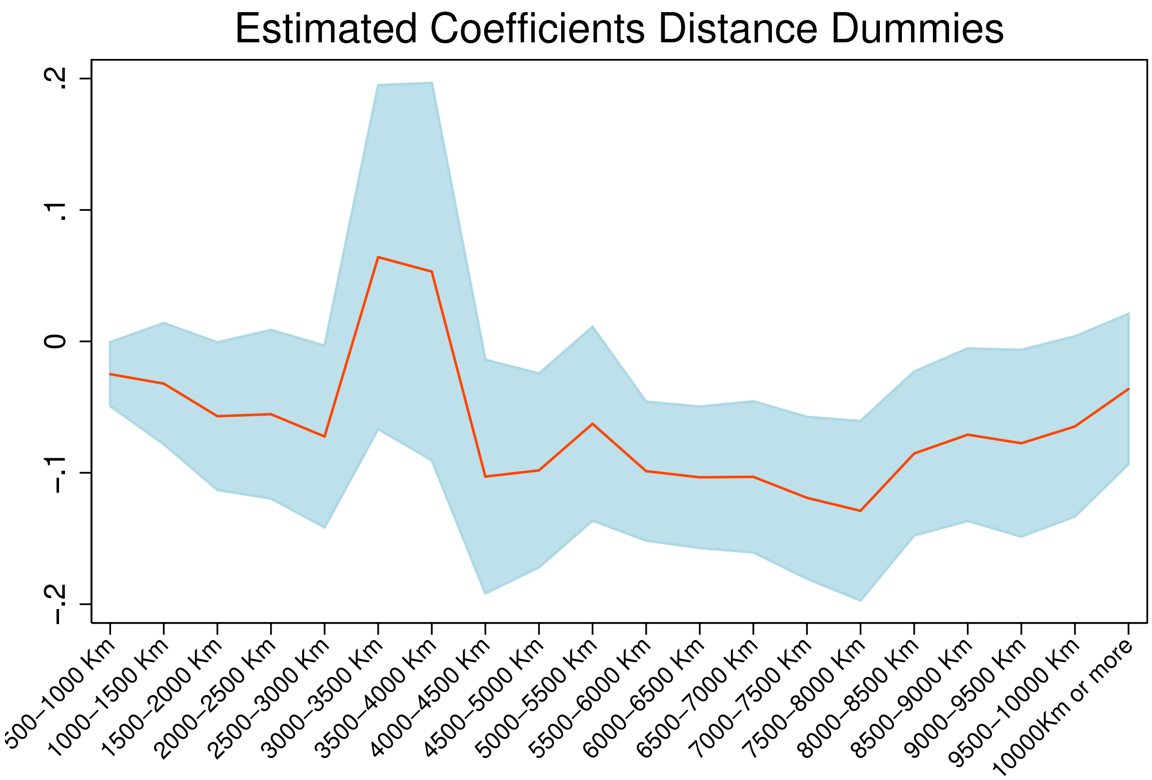
Table C.10: *Ease of Communication (KI: experience)*

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0628	-0.0629	-0.0621	-0.0629
	(0.037)*	(0.037)*	(0.037)*	(0.037)*
log(d)	-0.0241	-0.0252	-0.0072	-0.0185
	(0.009)**	(0.009)***	(0.014)	(0.008)**
log(t)	0.0275	0.0274	0.0281	0.0273
	(0.067)	(0.067)	(0.067)	(0.067)
Non-stop Flight		-0.0080		
		(0.010)		
Working hours overlap			0.0093	
			(0.005)**	
Common Language				0.0503
				(0.024)**
GDP per capita ratio	0.1309	0.1314	0.1282	0.1290
	(0.056)**	(0.056)**	(0.056)**	(0.055)**
Population ratio	0.0131	0.0133	0.0126	0.0142
	(0.007)*	(0.007)*	(0.007)*	(0.007)**
Capital per worker ratio	-0.0836	-0.0844	-0.0895	-0.0803
	(0.046)*	(0.046)*	(0.044)**	(0.044)*
Human Capital ratio	-0.0071	-0.0043	-0.0051	-0.0126
	(0.063)	(0.062)	(0.063)	(0.064)
Land per worker ratio	-0.0132	-0.0126	-0.0079	-0.0075
	(0.007)*	(0.007)*	(0.007)	(0.008)
Constant	0.6318	0.6431	0.4340	0.5765
	(0.162)***	(0.161)***	(0.219)**	(0.160)***
N	55136	55136	55136	55132
R-squared	0.47	0.47	0.47	0.47
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	N

The table presents results for the estimation of Specification (3.8) using a sample of foreign subsidiaries of MNCs. It uses an O*NET-based indicator for knowledge intensity based on workers' accumulated experience (excluding training). The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as an horizontal expansion. The variables in the right hand side include the distance from the MNC headquarters to the foreign subsidiary, the unit shipping cost, knowledge intensity measures (in standard deviations from the mean) and other controls. The right hand side also includes variables measuring the ease of communication between a headquarters and its subsidiaries. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parenthesis.

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

Figure C.7: *Distance Intervals Estimators*



The figure presents the empirical estimation for the distance intervals coefficients from Table C.11. The grey area represents 95% confidence intervals.

Figure C.7 looks at the non-linearity of the distance effect. The Figure reflects a monotonically decreasing relationship between distance and the likelihood of horizontal foreign subsidiaries, in general, up to 8000Km. This is consistent with the linear fit shown in the main body of the paper. Given the standard errors, however, there is little we can say about a U-shaped non-linear form. Yet, for longer distances, the coefficients are strictly negative. It is important to note that after 8000Km there are considerably less observations in each one of those buckets.

Table C.11: *Determinants of Horizontal FDI, Distance Dummies*

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable			
	(1)	(2)	(3)
k	-0.0868		-0.0894
	(0.043)**		(0.043)**
log(t)	0.0225	0.0598	0.0217
	(0.066)	(0.056)	(0.065)
500-1000Km	-0.0249	-0.0246	-0.0234
	(0.012)**	(0.012)**	(0.013)*
1000-1500Km	-0.0321	-0.0310	-0.0230
	(0.023)	(0.024)	(0.021)
1500-2000Km	-0.0569	-0.0570	-0.0411
	(0.029)**	(0.028)**	(0.024)*
2000-2500Km	-0.0554	-0.0546	-0.0309
	(0.033)*	(0.033)*	(0.029)
2500-3000Km	-0.0724	-0.0728	-0.0391
	(0.035)**	(0.035)**	(0.031)
3000-3500Km	0.0641	0.0636	0.0907
	(0.067)	(0.066)	(0.060)
3500-4000Km	0.0531	0.0502	0.0830
	(0.073)	(0.073)	(0.067)
4000-4500Km	-0.1029	-0.1031	-0.0707
	(0.045)**	(0.045)**	(0.048)
4500-5000Km	-0.0982	-0.1009	-0.0867
	(0.038)***	(0.037)***	(0.036)**
5000-5500Km	-0.0627	-0.0661	-0.0604
	(0.037)*	(0.037)*	(0.036)*
5500-6000Km	-0.0987	-0.1008	-0.0951
	(0.027)***	(0.026)***	(0.027)***
6000-6500Km	-0.1034	-0.1037	-0.0994
	(0.027)***	(0.027)***	(0.027)***
6500-7000Km	-0.1031	-0.1037	-0.0957
	(0.029)***	(0.029)***	(0.030)***
7000-7500Km	-0.1191	-0.1190	-0.1068
	(0.031)***	(0.031)***	(0.031)***
7500-8000Km	-0.1289	-0.1284	-0.1097
	(0.035)***	(0.034)***	(0.033)***
8000-8500Km	-0.0853	-0.0863	-0.0653
	(0.032)***	(0.031)***	(0.034)*
8500-9000Km	-0.0710	-0.0705	-0.0475
	(0.033)**	(0.033)**	(0.037)
9000-9500Km	-0.0775	-0.0766	-0.0528
	(0.036)**	(0.036)**	(0.037)
9500-10000Km	-0.0647	-0.0596	-0.0350
	(0.035)*	(0.036)*	(0.036)
10000Km+	-0.0362	-0.0349	-0.0087
	(0.029)	(0.029)	(0.033)
Constant	0.4849	0.5342	0.5921
	(0.140)***	(0.127)***	(0.218)***
N	55137	55137	55137
R-squared	0.47	0.47	0.47
MNC FE	Y	Y	Y
Host Cntry FE	N	N	Y

The table presents results for the estimation of Specification (3.8) using a sample of foreign subsidiaries of MNCs. The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as an horizontal expansion. The variables in the right hand side include the distance from the MNC headquarters to the foreign subsidiary in dummies each representing a 500km interval, the unit shipping cost, knowledge intensity measures. All specifications include a vector of controls which include the ratio of GDP per capita, population, human capital, physical capital and land between the home and recipient country of the investment. All columns also include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parenthesis.

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

C.6.4 Intellectual Property Rights: Excluding China

To alleviate concerns that the results are driven by the lack of intellectual property rights in China, I replicate Table 3.5 excluding China from the sample. When excluding China from the sample, however, the results are robust to the ones presented in the main body of the paper, as can be seen in Table C.12.

C.6.5 Excluding European Firms

Given the large number of European firms in the sample, and the short distances in the continent, this raises concerns about the validity of the analysis in terms of the tradeoff MNC face in locating their knowledge intensive subsidiaries at shorter distances. Hence, I repeat the corresponding analysis excluding all foreign subsidiaries located in Western Europe that belong to a European MNC (i.e., for which its headquarters is located in Western Europe). The results can be seen in in Figure C.8. As can be seen, the results are robust to the exclusion of these observations from the sample.

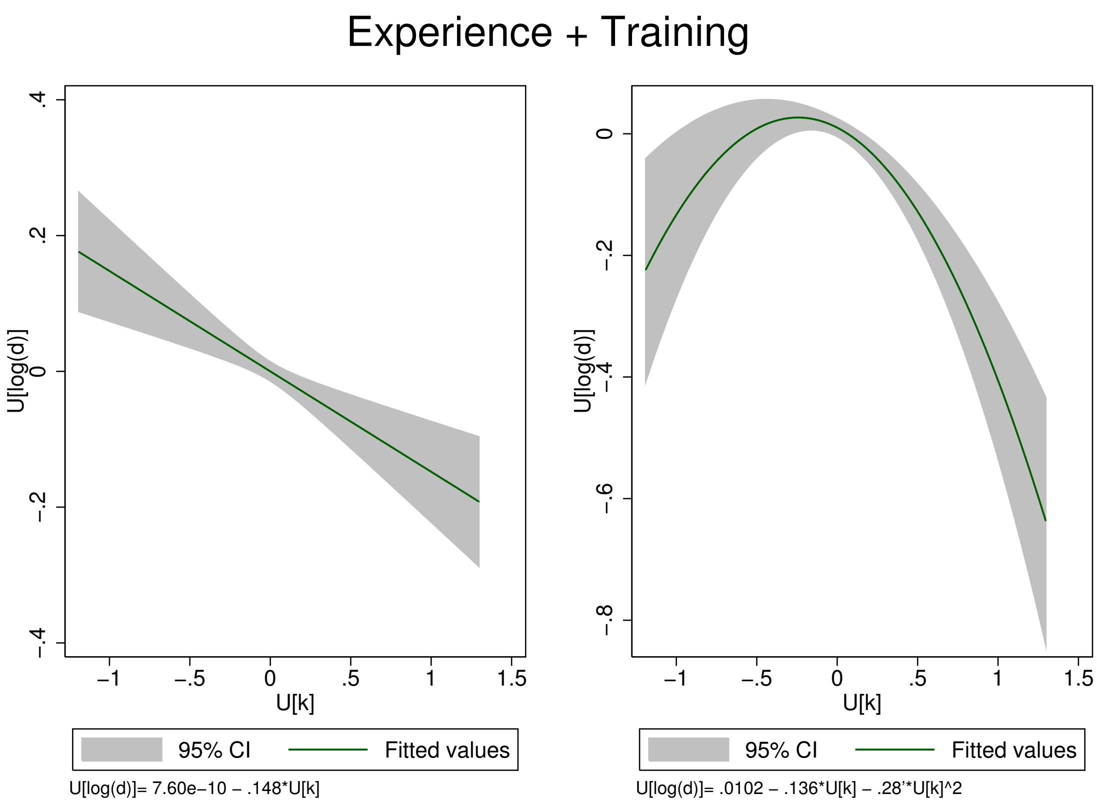
Table C.12: *Determinants of Horizontal FDI, excluding China*

Dependent Variable: Horizontal Foreign Subsidiary Binary Variable				
	(1)	(2)	(3)	(4)
k	-0.0869	-0.0864		-0.0890
	(0.044)**	(0.044)**		(0.044)**
log(d)	-0.0257		-0.0254	-0.0249
	(0.010)***		(0.010)***	(0.011)**
log(t)	0.0256	0.0261	0.0625	0.0253
	(0.066)	(0.067)	(0.057)	(0.065)
GDP per capita ratio	0.1303	0.1269	0.1287	0.9143
	(0.055)**	(0.055)**	(0.055)**	(0.243)***
Population ratio	0.0127	0.0219	0.0141	0.3109
	(0.007)*	(0.007)***	(0.007)**	(0.077)***
Capital per worker ratio	-0.0834	-0.1030	-0.0793	-0.6143
	(0.047)*	(0.048)**	(0.048)*	(0.182)***
Human Capital ratio	-0.0064	0.0531	-0.0106	-0.5025
	(0.066)	(0.065)	(0.067)	(0.292)*
Land per worker ratio	-0.0131	-0.0120	-0.0132	-0.0179
	(0.007)*	(0.008)	(0.007)*	(0.052)
Constant	0.6420	0.4466	0.6891	0.6876
	(0.161)***	(0.135)***	(0.149)***	(0.214)***
N	54259	54260	54259	54259
R-squared	0.47	0.47	0.47	0.48
MNC FE	Y	Y	Y	Y
Host Cntry FE	N	N	N	Y

The table presents results for the estimation of Specification (3.8) using a sample of foreign subsidiaries of MNCs, excluding subsidiaries in China. The left hand side variable is a binary variable that takes the value 1 if the foreign subsidiary is classified as an horizontal expansion. The variables in the right hand side include the distance from the MNC headquarters to the foreign subsidiary, the unit shipping cost, knowledge intensity measures and other controls. All specifications include MNC fixed effects. Robust standard errors clustered at the industry level are presented in parenthesis.

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

Figure C.8: $U[k]$ *vs.* $U[log(d)]$*, excluding Europe*



The figure presents the empirical fit for the relationship between d and k (the latter proxied by the *experience plus training* measure). The left column performs a linear fit between k and d while the right column performs a quadratic firm between the two. The grey area represents the 95% confidence interval for the estimated relationship. The sample excludes foreign subsidiaries located in Europe owned by a European firm.