

# Relatedness, Knowledge Diffusion, and the Evolution of Bilateral Trade

Bogang Jun<sup>a</sup>, Aamena Alshamsi<sup>b</sup>, Jian Gao<sup>a,c</sup>, César A. Hidalgo<sup>a,1</sup>

<sup>a</sup>*Collective Learning Group, MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

<sup>b</sup>*Masdar Institute of Science and Technology, Abu Dhabi, UAE*

<sup>c</sup>*ComplexLab, Web Sciences Center and Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, China.*

## Abstract

During the last decades two important contributions have reshaped our understanding of international trade. First, countries trade more with those with whom they share history, language, and culture, suggesting that trade is limited by information frictions. Second, countries are more likely to start exporting products that are similar to their current exports, suggesting that knowledge diffusion among related industries is a key constrain shaping the diversification of exports. But does knowledge about how to export to a destination also diffuses among related products and geographic neighbors? Do countries need to learn how to trade each product to each destination? Here, we use bilateral trade data from 2000 to 2015 to show that countries are more likely to increase their exports of a product to a destination when: (i) they export related products to it, (ii) they export the same product to the neighbor of a destination, (iii) they have neighbors who export the same product to that destination. Then, we explore the magnitude of these effects for new, nascent, and experienced exporters, (exporters with and without comparative advantage in a product) and also for groups of products with different level of technological sophistication. We find that the effects of product and geographic relatedness are stronger for new exporters, and also, that the effect of product relatedness is stronger for more technologically sophisticated products. These findings support the idea that international trade is shaped by information frictions that are reduced in the presence of related products and experienced geographic neighbors.

**Keywords:** International trade, Collective learning, Economic complexity, Knowledge Diffusion

## 1. Introduction

For more than a century, the literature on international trade explained global commerce as a consequence of differences in factor endowments Heckscher and Ohlin (1991), product quality, and product differentiation (Krugman, 1979, 1991; Anderson, 1979; Helpman, 1987). More recent streams of literature, however, have shown that there is more to international trade than endowments, costs, and distance, since countries need to learn how to produce and export each

\*Email address: hidalgo@mit.edu

product (Hidalgo et al., 2007; Hidalgo, 2015), and also, need to overcome important information frictions to enter each export destination (Rauch, 1999, 2001; Rauch and Trindade, 2002; Petropoulou, 2008; Portes and Rey, 2005; Casella and Rauch, 2002; Anderson and Marcouiller, 2002; Garmendia et al., 2012).

During the last two decades scholars have documented that volumes of bilateral trade decrease with the presence of borders (McCallum, 1995), and increase with migrants, shared language, and social networks (Rauch and Trindade, 2002; Rauch, 2001; Combes et al., 2005; Chaney, 2014; Bailey et al., 2017). In fact, using the random re-allocation of the Vietnamese boat people –a population of 1.4 million Vietnamese refugees reallocated in the U.S.–, Parsons and Vézina (2017) showed that states who received a 10% increase in their Vietnamese population experienced a growth in exports to Vietnam of between 4.5% and 14%.

But the evidence in favor of knowledge diffusion is not only expressed on aggregated trade flows, since scholars have also shown the effects of language, social networks, and informal institutions to be larger for differentiated products (Rauch, 1999, 2001; Rauch and Trindade, 2002). This suggests that factors limiting knowledge and information diffusion (from social networks to language) play a more important role in the diffusion of the knowledge and information needed to exchange more sophisticated goods.

A second stream of literature has focused on the supply side, in particular, on the process by which countries learn how to produce the products they export. This literature has shown that the ability of countries and regions to enter new export markets is limited by knowledge diffusion, since countries and regions are more likely to start exporting products when these are related to their current exports (Hidalgo et al., 2007; Hidalgo and Hausmann, 2009; Hausmann et al., 2014; Boschma et al., 2013), or when their geographic neighbors are already exporting them (Bahar et al., 2014). The importance of knowledge diffusion in the diversification of economic activities, however, is not limited only to the export of products. It has also been observed in the development of regional industries (Neffke et al., 2011; Gao et al., 2017), research activities (Guevara et al., 2016), and technologies (Kogler et al., 2013; Boschma et al., 2014), suggesting that relatedness between economic activities facilitates knowledge diffusion in general, not only in the context of international trade.

Together, these findings have given rise to a more nuanced picture of international trade. A picture where factor endowments and transportation costs do not determine trade fully, because information frictions and knowledge diffusion determine the knowledge a country has, and hence, the products it can produce and the partners it can trade with.

Here, we contribute to this literature by combining the stream of literature on knowledge and information frictions and that on relatedness by exploring the path dependencies affecting the evolution of the network of export destinations for each product. We use more than 15 years of bilateral trade data, disaggregated into more than 1,200 products, to construct a gravity model that validates previous findings and expands them. Looking at hundreds of thousands of bilateral trade links reveals that: (i) countries are more likely to increase their exports of a product to a destination when they already export related products to it; (ii) countries are more likely to increase their exports of a product to a destination when they already export to that destination's neighbors, and (iii) countries are more likely to increase the exports of a product to a destination when their neighbors export that product to that destination. Moreover, we find that sharing a colonial past, a language, a border, or bilingual speakers (when the two countries share a language), is also associated with an increase in the volume of trade.

Yet, only some of these findings are novel. The effects of common language, shared border, and shared colonial past that we reproduce here have been documented in the past (McCallum,

1995; Rauch, 1999). Also, we know that countries are more likely to start exporting to a destination when they export to that destination's neighbors (Chaney, 2014). What is novel, are (i) the effect of relatedness on bilateral trade volumes, (ii) the effect of having a neighbor export the same product to the same destination, and (iii) the effect of bilingual speakers. In particular, we find that the effects of relatedness—the finding that countries increase their exports of a product to a destination when they already export related products to it—are especially strong. In fact, a one standard deviation increase in relatedness is associated with a 20% increase in exports in a two year period. This effect is about 50% larger than the effect of exporting that product to a neighbor of the target destination (Chaney, 2014), and more than 170% larger than the effect of having a neighbor export the same product to the same destination. The effect of bilingual speakers, while significant after considering all controls, are much smaller than that of sharing a language.

We also study these effects by separating exporters into new exporters, nascent, and experienced exporters, by considering as new exporters those without comparative advantage in a product. We find the effects of product relatedness, and especially those of geographic relatedness among exporters, to be stronger for new exporters than for experienced exporters. Also, we test the hypothesis that knowledge diffusion should affect more strongly products that are knowledge intensive (Rauch, 1999, 2001; Rauch and Trindade, 2002) by dividing products into the five technological categories suggested by Lall (Lall, 2000): primary, resource-based manufactures, low-tech, medium-tech, and high-tech products. We find that exporting related products has a stronger effect on the increase of exports for technological sophisticated products, suggesting that knowledge diffusion is more important for knowledge intense products. Surprisingly, we find no effect of technological sophistication on both of our measures of regional relatedness. That is, neither exporting to a neighbor, nor having a neighbor export the same product, appears to have an effect on future exports that either increases or decreases with technological sophistication. Also, we find that sharing a language and a colonial past has a larger effect on the increase of exports for more technologically sophisticated products, providing further evidence that the effects are due to information and knowledge frictions. Moreover, we find the negative effect of distance—which are correlated with social network connections (Breschi and Lissoni, 2009; Singh, 2005)—to be larger for technologically sophisticated products. These findings support the idea that establishing trade relationships requires overcoming information and knowledge frictions, and that product and geographic relatedness help reduce these frictions.

## 2. Data

We use bilateral trade data from 2000 to 2015 from MIT's Observatory of Economic Complexity (Simoes and Hidalgo, 2011). The data is disaggregated into the Harmonized System (HS rev 1992, four-digit level) and consists of imports and exports between countries. Because both exporter and importer report their trade information, we clean the data by comparing the data reported by exporters and importers following the work of Feenstra et al. (2005). Also, we exclude countries that have population less than 1.2 million or have a trade volume in 2008 that is below 1 billion in US dollar. Also, we exclude data from Iraq, Chad and Macau.

Macroeconomic data (GDP at market prices in current US dollar and population) comes from the World Bank's World Development Indicators. Data on geographical and cultural distance (shared language, physical distance between most populated cities, sharing a border, and shared colonial past) comes from GEODIST data from CEPII (Mayer and Zignago, 2011). For language proximity, we use one of the global language networks of Ronen et al. (2014): the

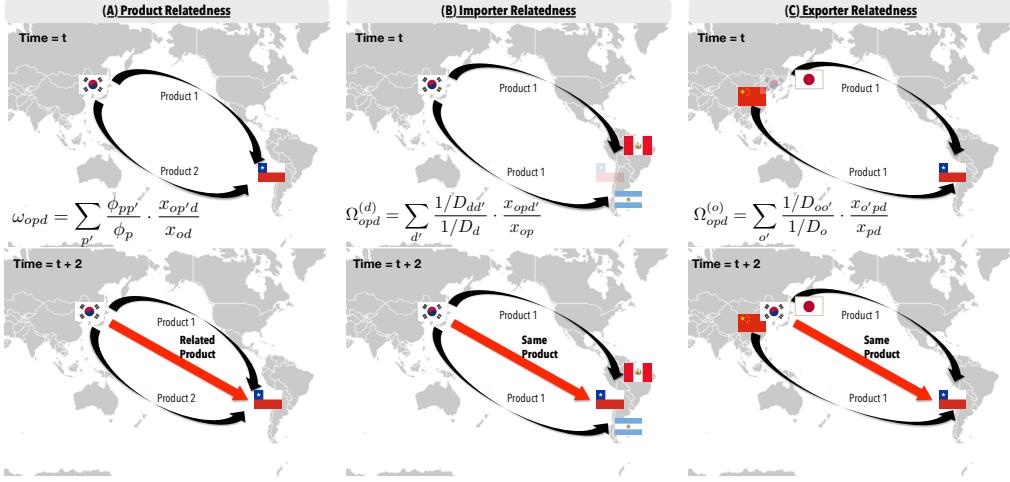


Figure 1: Relatedness among products, exporters, and importers. (A) Product Relatedness: the similarity between a product and the other products that a country already exports to a destination, (B) Importer Relatedness: the fraction of the geographic neighbors of a country that import a product from the same origin, and (C) Exporter Relatedness: the fraction of neighbors of a country that export a product to the same destination.

one considering the number of books translated from one language to another as a proxy for the number of translators, or bilingual speakers, between two languages.

### 3. Results

Does relatedness among products or geographic neighbors help facilitate the knowledge flows needed to increase bilateral trade flows?

To explore this question we introduce three measures of relatedness. We use these to estimate: (i) the similarity between a product and the other products that a country already exports to a destination (Product Relatedness), (ii) the fraction of the geographic neighbors of a country that import a product from the same origin (Importer Relatedness), and (iii) the fraction of neighbors of a country that export a product to the same destination (Exporter Relatedness). Product Relatedness should help us capture information about knowledge flows between products (which range from knowledge flows among industries to knowledge flows among product lines within a firm). Figure 1(A) illustrates Product Relatedness in the context of Korea and Chile. In the example, Korea exports Products I and II to Chile (Shirts and Pants), and this may affect the future exports of Product III (Coats) to Chile, when Product III (Coats) is highly related to Products I and II (Shirts and Pants). Our hypothesis is that knowledge flows should be larger among related products, and hence, exports should increase faster when a country exports related products to a destination.

Importer Relatedness helps us capture knowledge flows on how to: (i) import a product from the same origin than a neighbor, or (ii) export to a neighbor of a current destination. In the example of Figure 1(B), Korea exports Product I (Shirts) to Peru and Argentina and that may affect the future volume of exports of Product I (Shirts) to Chile (who is a geographic neighbor of Peru and Argentina). Here, knowledge on how to import from an origin should be flowing

among neighboring importers, or knowledge on how to export to the neighbor's of a country's destinations should be flowing within the exporter.

Exporter Relatedness captures (i) knowledge flows among neighboring exporters on how to export to a destination, or (ii) knowledge flows on how to import from a neighbor of a country from where you currently import. In the example of Figure 1(C), Chile imports Product I (Shirts) from China and Japan, and that may affect the future volume of exports of Product I (Shirts) from Korea (which is a neighbor of the places from where Chile is currently importing Product I). This would be a knowledge flow on how to export to a destination among neighboring exporters, or a knowledge flow within an importer, of how to import from a neighbor of a current origin.

Mathematically, we can construct the three measures of relatedness using a similar formula. The formula is a weighted average of the number of neighbors, or related products, that already have an active trade relationships. In the case of similarity between products, weights are the proximity between products  $p$  and  $p'$ ,  $\phi_{pp'}$ .  $\phi_{pp'}$  is the minimum of the conditional probability that two products are co-exported by multiple countries (see Hidalgo et al. (2007) and Appendix A).  $\phi_{pp'} = 1$  means products  $p$  and  $p'$  are always co-exported and  $\phi_{pp'} = 0$  means no country exports both products. In the case of geographic neighbors, of both an exporter or an importer, weights are given by the inverse of geographic distance ( $1/D_{dd'}$  and  $1/D_{oo'}$ ), where  $D_{dd'}$  is the distance in kilometers between the most populated cities in countries  $d$  and  $d'$ .

Formally, let  $x_{opd}$  be a matrix summarizing the trade flow in US dollars of product  $p$  from exporter  $o$  to destination  $d$ . Then, Product Relatedness is given by:

$$\omega_{opd} = \sum_{p'} \frac{\phi_{pp'}}{\phi_p} \cdot \frac{x_{op'd}}{x_{od}} \quad (1)$$

where  $x_{od}$  is the volume of trade between countries  $o$  and  $d$  ( $x_{od} = \sum_p x_{opd}$ ) and  $\phi_p = \sum_{p'} \phi_{pp'}$ .

Similarly, Importer Relatedness is given by:

$$\Omega_{opd}^{(d)} = \sum_{d'} \frac{1/D_{dd'}}{1/D_d} \cdot \frac{x_{opd'}}{x_{op}}, \quad (2)$$

where  $x_{op}$  is the volume of trade of product  $p$  from origin country  $o$  ( $x_{op} = \sum_d x_{opd}$ ),  $D_{dd'}$  is the geographic distance between destination country  $d$  and its neighbors  $d'$ , and  $1/D_d = \sum_{d'} 1/D_{dd'}$ .

Finally, Exporter Relatedness is given by:

$$\Omega_{opd}^{(o)} = \sum_{o'} \frac{1/D_{oo'}}{1/D_o} \cdot \frac{x_{o'pd}}{x_{pd}}, \quad (3)$$

where  $x_{pd}$  is the volume of trade of product  $p$  to destination country  $d$  ( $x_{pd} = \sum_o x_{opd}$ ),  $D_{oo'}$  is the geographic distance between origin country  $o$  and its neighbors  $o'$ , and  $1/D_o = \sum_{o'} 1/D_{oo'}$ .

Next, we use these three measures of relatedness, together with data on common cultural and geographic factors, to construct an extended gravity model to study the marginal contribution of product, importer, and exporter relatedness, and of shared languages, borders, and colonial past, to the growth of future exports. Our model predicts bilateral trade in a product in two years time while controlling for: (i) initial trade in that product between the same trade partners, (ii) total exports of the product by the exporter, (iii) total imports of the product by the importer,

(iv-vii) the GDP per capita and population of exporters and importers, and (viii) their geographic distance. Formally, our model is given by Equation 4:

$$\begin{aligned}
x_{opd}^{t+2} = & \beta_0 + \beta_1 \omega_{opd}^t + \beta_2 \Omega_{opd}^{(d)t} + \beta_3 \Omega_{opd}^{(o)t} \\
& + \beta_4 x_{opd}^t + \beta_5 x_{op}^t + \beta_6 x_{pd}^t + \beta_7 D_{od} \\
& + \beta_8 gdp_o^t + \beta_9 gdp_d^t + \beta_{10} Population_o^t + \beta_{11} Population_d^t \\
& + \beta_{12} Border_{od} + \beta_{13} Colony_{od} + \beta_{14} Language_{od} \\
& + \beta_{15} Lang.Proximity_{od} \\
& + \epsilon_{opd}^t
\end{aligned} \tag{4}$$

where the dependent variable,  $x_{opd}^{t+2}$ , represents the volume of trade in (US dollar) of product  $p$  from exporter  $o$  to destination  $d$  in year  $t + 2$ . Our main variables of interest are our three measures of relatedness: Product Relatedness  $\omega_{opd}^t$ , Importer Relatedness  $\Omega_{opd}^{(d)t}$ , Exporter Relatedness  $\Omega_{opd}^{(o)t}$ , and shared border ( $Border_{od}$ ), shared language ( $Language_{od}$ ), language proximity (number of bilingual speakers  $Lang.Proximity_{od}$ ), and shared colonial past ( $Colony_{od}$ ).  $Border_{od}$ ,  $Language_{od}$ ,  $Colony_{od}$  are binary (dummy) variables (0 or 1). The other factors in the model  $gdp$  per capita, population, and distance ( $D_{od}$ ), are standard gravity controls (Tinbergen, 1962; Pöyhönen, 1963). Finally, by incorporating the total volume of exports of a country ( $x_{op}$ ), the total imports of a destination ( $x_{pd}$ ), and the present day trade flow for each product between an origin and a destination ( $x_{opd}$ ), we capture the effects of our variable of interest in the change in trade experience in the subsequent two years. In Equation 4, we make all variables comparable (except binary variables) by standardizing them by subtracting their means and dividing them by their standard deviations. (Please see Appendix B and C for summary statistics and correlation among variables)

Table 1 shows our main results divided into three periods: 2000-2006 (pre-financial crisis), 2007-2012 (crisis period), and 2012-2015 (recovery period). Since our result are qualitatively the same for all of these periods we will describe them in unison. (See detailed results, correlation table and summary statistics in Appendix B and C)

First, we find that the three relatedness variables correlate positively with future bilateral trade. This means that a country is likely to experience an increase in their exports of product  $p$  to a destination  $d$  when: (i) the country is exporting related products to that destination, (ii) it is exporting the same product to the neighbors of a destination (confirming Chaney (2014)), and (iii) it has neighbors that are already exporting the same product to that destination. This extends Bahar et al. (2014), who showed that having geographic neighbors increases the probability of exporting a new product, since Bahar et al. (2014) did not look at individual export destinations (they aggregate across all destinations). Our findings, therefore, complement Bahar et al. (2014) by showing that having neighbors that export a product does not only increase the total volume of exports, but the volume of exports to the same destinations that the neighbors were exporting to.

When comparing the effects of product and geographical relatedness (variables are standardized), we find that the role of Product Relatedness ( $\omega_{opd}^t$ ) is on average the largest, while that of Exporter Relatedness ( $\Omega_{opd}^o$ ) is the smallest. In addition to these, we find strong and positive effects for the role of shared borders, colonial past, shared language, and to a lesser extent, number of translations (language proximity). Other standard gravity factors (distance, GDP per capita, and population) behave as expected.

Table 1: Bilateral trade volume after two years for periods 2000-2006 (pre-financial crisis), 2007-2012 (crisis period) and 2012-2015 (recovery period)

	Dependent variable: $\log x_{opd}^{t+2}$		
	2000-2006	2007-2012	2012-2015
$\omega_{opd}^t$	0.209*** (0.001)	0.180*** (0.001)	0.152*** (0.001)
$\Omega_{opd}^{(d)}$	0.143*** (0.001)	0.143*** (0.001)	0.151*** (0.001)
$\Omega_{opd}^{(o)}$	0.077*** (0.001)	0.105*** (0.001)	0.107*** (0.001)
$\log x_{opd}^t$	1.371*** (0.001)	1.600*** (0.001)	1.769*** (0.001)
$\log x_{op}^t$	0.961*** (0.001)	0.972*** (0.002)	0.915*** (0.002)
$\log x_{pd}^t$	0.529*** (0.001)	0.421*** (0.001)	0.406*** (0.002)
$\log Distance$	-0.485*** (0.001)	-0.419*** (0.001)	-0.432*** (0.004)
$\log gdp_o^t$	0.165*** (0.001)	0.198*** (0.001)	0.203*** (0.001)
$\log gdp_d^t$	0.226*** (0.001)	0.216*** (0.001)	0.281*** (0.001)
$\log Population_o$	0.472*** (0.001)	0.456*** (0.001)	0.455*** (0.002)
$\log Population_d$	0.344*** (0.001)	0.367*** (0.001)	0.338*** (0.001)
$Border_{od}$	0.712*** (0.003)	0.717*** (0.004)	0.611*** (0.005)
$Colony_{od}$	0.052*** (0.003)	0.127*** (0.004)	0.193*** (0.005)
$Language_{od}$	0.545*** (0.002)	0.441*** (0.003)	0.334*** (0.003)
$\log Lang.Proximity_{od}$	0.032*** (0.001)	0.027*** (0.001)	-0.001 (0.001)
Constant	9.653*** (0.001)	9.830*** (0.001)	9.793*** (0.001)
Observations	10,911,584	7,591,489	5,332,257
Adjusted R <sup>2</sup>	0.494	0.516	0.558
Residual Std. Error	2.568	2.637	2.529

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Together, the finding that relatedness among products, the presence of knowledge among geographic neighbors, language, colonial history, shared borders, and language proximity, all have a positive and significant effect in the increase of trade flows for particular products and countries, are evidence in support of the notion that knowledge on how to trade a specific product between a specific pair of countries needs to flow for that trade to be materialized. If this hypothesis is correct, we should also be able to study the varying importance of knowledge flows for new and experienced exporters (exporters with or without comparative advantage), and also, for products with different levels of technological sophistication.

Next, we test the effects of the exporters' level of competitiveness in the diffusion of the information needed to trade by dividing exporters of each product into new, nascent, and experienced. We do this by calculating the revealed comparative advantage (RCA) of each exporter in each product. RCA is the ratio between the exports of a country in a product, and the exports that

Table 2: Bilateral trade volume after two years for new, nascent, and experienced exporters

	Dependent variable: $\log x_{opd}^{t+2}$		
	New Exporter	Nascent Exporter	Experienced Exporter
$\omega_{opd}^t$	0.194*** (0.003)	0.140*** (0.004)	0.184*** (0.001)
$\Omega_{opd}^{(d)}$	0.149*** (0.003)	0.122*** (0.005)	0.122** (0.001)
$\Omega_{opd}^{(o)}$	0.111*** (0.003)	0.078*** (0.005)	0.069** (0.001)
$\log x'_{opd}$	0.474*** (0.003)	1.410*** (0.005)	1.484*** (0.001)
$\log x'_{op}$	0.964*** (0.005)	0.684*** (0.006)	0.804*** (0.001)
$\log x'_{pd}$	0.525*** (0.005)	0.518*** (0.008)	0.553** (0.001)
$\log Distance$	-0.532*** (0.004)	-0.504*** (0.006)	-0.454*** (0.001)
$\log gdp_o^t$	0.235*** (0.004)	0.128*** (0.005)	0.100** (0.001)
$\log gdp_d^t$	0.010** (0.005)	0.045*** (0.007)	0.264*** (0.001)
$\log Population_o$	0.331*** (0.003)	0.388*** (0.005)	0.446** (0.001)
$\log Population_d$	0.190*** (0.004)	0.213*** (0.006)	0.362** (0.001)
$Border_{od}$	0.651*** (0.009)	0.817*** (0.016)	0.713** (0.004)
$Colony_{od}$	0.218*** (0.012)	0.259*** (0.020)	0.097** (0.004)
$Language_{od}$	0.557*** (0.007)	0.782*** (0.011)	0.481** (0.003)
$\log Lang.Proximity_{od}$	0.008*** (0.003)	0.016*** (0.004)	0.029*** (0.001)
Constant	7.776*** (0.004)	9.213*** (0.004)	10.124*** (0.001)
Observations	922,092	463,388	8,045,262
Adjusted R <sup>2</sup>	0.281	0.456	0.526
Residual Std. Error	2.641	2.535	2.487

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

are expected based on a country's total export market and the size of the global market for that product. We classify as new exporters all countries with an RCA below 0.2 in a product (countries that export less than 20% of what they are expected to export by chance). We classify as nascent exporters, all countries with an RCA between 0.2 and 1. We classify as the experienced exporters of a product, all countries that have revealed comparative advantage in it ( $RCA > 1$ ).

Table 2 divides country-product pairs into new, nascent, and experienced exporters. The results are consistent with those presented in Table 1, but also, reveal two important distinctions. First, the effects of product and geographic relatedness, especially Exporter Relatedness, are stronger for new exporters, suggesting that knowledge and information frictions impose larger constraints for countries that are not experienced in the export of a product. Second, the overall explanatory power of the model is considerably larger for experienced exporters ( $R^2 \approx 53\%$  vs  $R^2 \approx 46\%$  for nascent exporters and  $R^2 \approx 28\%$  for new exporters. These are large differences, even considering that the sample sizes are not the same). This suggests that inexperienced exporters face more uncertainty (less predictable because of lower R-square), and hence, benefit more from relatedness (higher relatedness coefficients).

Finally, we explore the interaction between our three measures of relatedness and the technological sophistication of products using Lall (2000)'s five technological categories: primary,

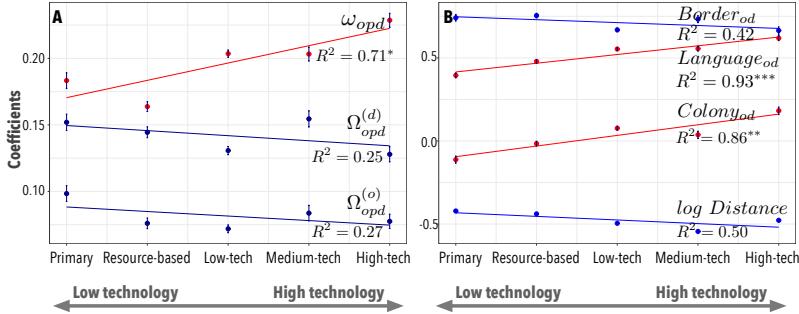


Figure 2: Coefficients of variables by the technological sophistication of products; (A) Coefficients of  $\omega_{opd}$ ,  $\Omega_{opd}^{(d)}$ , and  $\Omega_{opd}^{(o)}$ , (B) Coefficients of  $Border_{od}$ ,  $Language_{od}$ , and  $Colony_{od}$ , and (C) Coefficients of  $\log lang.Proximity$  and  $\log Distance$ . The fitted lines in red are statistically significant, while the lines in blue are not statistically significant.  
\* $p<0.1$ ; \*\* $p<0.05$ ; \*\*\* $p<0.01$

resource-based manufactures, low-tech, medium-tech, and high-tech products. Since Lall's classification is based on the 3-digit Standard International Trade Classification (SITC-3) rev 2, we match products to our data using the conversion table provided by the UN Trade Statistics site<sup>1</sup>. Following Lall (2000), we also exclude "special transactions" such as electric current, cinema film, printed matter, fold, coins, and pets.

We present the coefficients for all of the variables we had in the previous model in the Appendix E (Table A15) and summarize the main results graphically, by plotting the coefficients and their errors as a function of the technological sophistication of products in Figure 2. Trends that increase significantly with technological sophistication ( $p < 0.1$ ) are presented in red, non-significant trends are shown in blue.

Figure 2 A shows that the effect of Product Relatedness, but not that of Importer Relatedness or Exporter Relatedness, increases with technological sophistication. This suggests that product relatedness captures channels of knowledge and information flow that are relevant for the export of sophisticated products. Also, Figure 2 B shows that the effect of sharing a language and a colonial past, but not those of sharing a border, are larger for more technologically sophisticated products. Once again, this reiterates the idea that borders and geographic distance affect knowledge flows by limiting social interactions (Singh, 2005; Breschi and Lissoni, 2009), so we do not see much of a geographic effect once we take cultural and linguistic similarity into account. The negative effect of distance is slightly larger for technologically sophisticated products, but the effect is not strong enough to be significant. Together, these findings support the idea that trade is driven partly by the diffusion of knowledge and information on how to export each product to each destination.

#### 4. Discussion

During the last decades two ideas have re-framed our understanding of international trade. The first idea is that information and knowledge frictions, not just differences in transportation

<sup>1</sup> Available at <https://unstats.un.org/unsd/trade/classifications/correspondence-tables.asp>

costs, factor endowments, and differences in productivity, shape global trade. The second idea is that countries need to learn how to produce the products they export, and hence, evolve their productive structures in a path dependent manner that is constrained by knowledge flows. Here, we use bilateral trade data, together with various measures of economic size, culture, and geographic proximity, to put these two ideas together. Our findings confirm many existing theories involving the role of language, and culture, but also, add to the body of knowledge by showing that relatedness among products and countries shape future trade volumes. In particular, we showed that relatedness among products, and among geographic neighbors, explains a substantial fraction of future bilateral trade: trade volumes increase when countries export related products to a destination, but also, when they share neighbors who export to that destination, or when they are already exporting to a destination's neighbors. When comparing these three forms of relatedness, we found that relatedness among products is the strongest, suggesting that there may be product or industry specific learning channels that play an important role in the diffusion of the knowledge needed to establish or increase trade relationships. Moreover, we found the effects of relatedness to be stronger for new exporters, and the effects of product relatedness to be stronger for more technologically sophisticated products. These additional considerations support the idea that the presence of related activities facilitates the knowledge flows that countries need to learn how to produce and export products to specific destinations.

Yet, our results leave unanswered many questions about the mechanisms underlying these knowledge and information flows. The two channels we observed among geographic neighbors (Importer and Exporter Relatedness) could further be disaggregated into four channels: the knowledge flows among neighboring importers or exporters, or the knowledge flows within an exporter or within an importer. From there, we may be able to start learning about the specific mechanisms that underlie each of these knowledge flows. Also, the interpretation of relatedness has a similar problem. On the one hand, one cannot know if the flow of knowledge is among product lines, within the same firm or industry, among industries, or the result of spin-offs, foreign direct investment, or migrations.

Nevertheless, our results do provide some light in the long quest to understand how social networks, culture, and knowledge flows, shape international trade. They tell us that product relatedness plays an important role since the size of its effect is larger than the one observed among geographic neighbors. This suggests that looking at knowledge flows among product lines, and among industries, should be an avenue of inquiry for improving our understanding of the social and economic forces that shape global trade.

### Acknowledgement

We thank Mauricio (Pacha) Vargas and Alex Simoes for help with the data. We also thank Cristian Jara Figueroa, Flávio Pinheiro, Tarik Roukny, and Dogyoon Song for helpful comments. This project is funded by the MIT Skoltech Program and by the Cooperative Agreement between the Masdar Institute of Science and Technology and the MIT Media Lab Consortia.

## References

- Anderson, J. E., 1979. A theoretical foundation for the gravity equation. *American Economic Review* 69 (1), 106–116.
- Anderson, J. E., Marcouiller, D., 2002. Insecurity and the pattern of trade: An empirical investigation. *Review of Economics and Statistics* 84 (2), 342–352.
- Bahar, D., Hausmann, R., 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.
- Bailey, M., Cao, R. R., Kuchler, T., Stroebel, J., Wong, A., 2017. Measuring social connectedness. NBER Working Paper No. 23608.
- Balassa, B., 1965. Trade liberalisation and “revealed” comparative advantage. *Manchester School* 33 (2), 99–123.
- Boschma, R., Balland, P.-A., Kogler, D. F., 2014. Relatedness and technological change in cities: The rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. *Industrial and Corporate Change* 24 (1), 223–250.
- Boschma, R., Minondo, A., Navarro, M., 2013. The emergence of new industries at the regional level in Spain: A proximity approach based on product relatedness. *Economic Geography* 89 (1), 29–51.
- Breschi, S., Lissoni, F., 2009. Mobility of skilled workers and co-invention networks: An anatomy of localized knowledge flows. *Journal of Economic Geography* 9 (4), 439–468.
- Casella, A., Rauch, J. E., 2002. Anonymous market and group ties in international trade. *Journal of International Economics* 58 (1), 19–47.
- Chaney, T., 2014. The network structure of international trade. *American Economic Review* 104 (11), 3600–3634.
- Combes, P.-P., Lafourcade, M., Mayer, T., 2005. The trade-creating effects of business and social networks: Evidence from France. *Journal of International Economics* 66 (1), 1–29.
- Feenstra, R. C., Lipsey, R. E., Deng, H., Ma, A. C., Mo, H., 2005. World trade flows: 1962-2000. NBER Working Paper No. 11040.
- Gao, J., Jun, B., Pentland, A., Zhou, T., Hidalgo, C. A., 2017. Collective learning in China’s regional economic development. arXiv:1703.01369.
- Garmendia, A., Llano, C., Minondo, A., Requena, F., 2012. Networks and the disappearance of the intranational home bias. *Economics Letters* 116 (2), 178–182.
- Guevara, M. R., Hartmann, D., Aristarán, M., Mendoza, M., Hidalgo, C. A., 2016. The research space: Using career paths to predict the evolution of the research output of individuals, institutions, and nations. *Scientometrics* 109 (3), 1695–1709.
- Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Simoes, A., Yildirim, M. A., 2014. The atlas of economic complexity: Mapping paths to prosperity. MIT Press.
- Heckscher, E. F., Ohlin, B. G., 1991. Heckscher-Ohlin trade theory. MIT Press.
- Helpman, E., 1987. Imperfect competition and international trade: Evidence from fourteen industrial countries. *Journal of the Japanese and International Economies* 1 (1), 62–81.
- Hidalgo, C. A., 2015. Why information grows: The evolution of order, from atoms to economies. Basic Books.
- Hidalgo, C. A., Hausmann, R., 2009. The building blocks of economic complexity. *Proceedings of the National Academy of Sciences of the United States of America* 106 (26), 10570–10575.

- Hidalgo, C. A., Klinger, B., Barabási, A.-L., Hausmann, R., 2007. The product space conditions the development of nations. *Science* 317 (5837), 482.
- Kogler, D. F., Rigby, D. L., Tucker, I., 2013. Mapping knowledge space and technological relatedness in US cities. *European Planning Studies* 21 (9), 1374–1391.
- Krugman, P., 1979. Increasing returns, monopolistic competition, and international trade. *Journal of International Economics* 9 (4), 469–479.
- Krugman, P., 1991. Increasing returns and economic geography. *Journal of Political Economy* 99 (3), 483–499.
- Lall, S., 2000. The technological structure and performance of developing country manufactured exports, 1985–98. *Oxford Development Studies* 28 (3), 337–369.
- Mayer, T., Zignago, S., 2011. Notes on CEPII's distances measures: The geodist database. CEPII Working Paper No. 2011-25.
- McCallum, J., 1995. National borders matter: Canada-US regional trade patterns. *American Economic Review* 85 (3), 615–623.
- Neffke, F., Henning, M., Boschma, R., 2011. How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic Geography* 87 (3), 237–265.
- Parsons, C., Vézina, P.-L., 2017. Migrant networks and trade: The vietnamese boat people as a natural experiment. IZA Discussion Paper No. 10112.
- Petropoulou, D., 2008. Information costs, networks and intermediation in international trade. CEP Discussion Paper No. 848.
- Portes, R., Rey, H., 2005. The determinants of cross-border equity flows. *Journal of International Economics* 65 (2), 269–296.
- Pöyhönen, P., 1963. A tentative model for the volume of trade between countries. *Weltwirtschaftliches Archiv*, 93–100.
- Rauch, J. E., 1999. Networks versus markets in international trade. *Journal of International Economics* 48 (1), 7–35.
- Rauch, J. E., 2001. Business and social networks in international trade. *Journal of Economic Literature* 39 (4), 1177–1203.
- Rauch, J. E., Trindade, V., 2002. Ethnic Chinese networks in international trade. *Review of Economics and Statistics* 84 (1), 116–130.
- Ronen, S., Gonçalves, B., Hu, K. Z., Vespignani, A., Pinker, S., Hidalgo, C. A., 2014. Links that speak: The global language network and its association with global fame. *Proceedings of the National Academy of Sciences of the United States of America* 111 (52), E5616–E5622.
- Simoes, A. J. G., Hidalgo, C. A., 2011. The economic complexity observatory: An analytical tool for understanding the dynamics of economic development. In: Proceedings of the 17th AAAI Conference on Scalable Integration of Analytics and Visualization. AAAIWS'11-17. AAAI Press, pp. 39–42.
- Singh, J., 2005. Collaborative networks as determinants of knowledge diffusion patterns. *Management Science* 51 (5), 756–770.
- Tinbergen, J., 1962. Shaping the World Economy. New York: Twentieth Century Fund.

## Appendix A. Building a product space for 2000-2015

To calculate the  $\omega_{opd}$ , we need firstly build a product space. We define the product space by looking at all proximity measures between products (Hidalgo et al., 2007) after aggregating all the data that covers from 2000 to 2015. To capture the significant trade flow, we calculate the revealed comparative advantage (RCA) following Balassa (1965):

$$RCA_{o,i} = \frac{x_{o,i}}{\sum_i x_{o,i}} / \left( \frac{\sum_o x_{o,i}}{\sum_{o,i} x_{o,i}} \right) \quad (\text{A1})$$

Based on the result of RCA, we measure the proximity between product by calculating  $\phi_{i,j}$  between product  $i$  and  $j$  (Hidalgo et al., 2007).

$$\phi_{i,j} = \min\{P(RCA_i|RCA_j), P(RCA_j|RCA_i)\} \quad (\text{A2})$$

Using this significant trade flow over 2000-2015, we can create  $1242 \times 1242$  matrix whose entities are the proximity between products. Figure A1 shows the product space of world market in the period from 2000 to 2015.

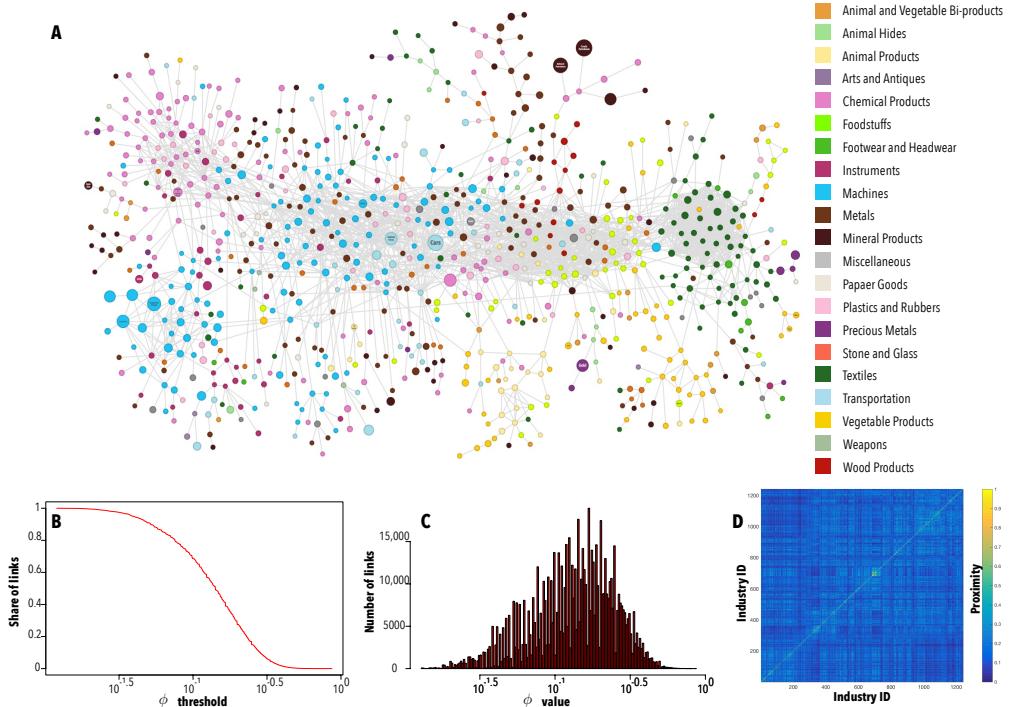


Figure A1: Product space over 2000 to 2015: (A) Network representation of product space, (B) Cumulative distribution of proximity values, (C) Density distribution of proximity values, and (D) the product space matrix sorted in increasing order of the is numerical code.

## Appendix B. Regression Results

Table A1: Bilateral trade volume after two years for periods 2000-2006

	Dependent variable: $\log x_{opd}^{t+2}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\omega_{opd}^t$		0.122*** (0.001)			0.116*** (0.001)	0.202*** (0.001)	0.209*** (0.001)
$\Omega_{opd}^{(d)}$			0.118*** (0.001)		0.109*** (0.001)	0.139*** (0.001)	0.143*** (0.001)
$\Omega_{opd}^{(o)}$				0.041*** (0.001)	0.027*** (0.001)	0.042*** (0.001)	0.077*** (0.001)
$\log x_{opd}^t$	0.441*** (0.0003)	0.439*** (0.0003)	0.441*** (0.0003)	0.441*** (0.0003)	0.440*** (0.0003)	0.396*** (0.0003)	0.377*** (0.0003)
$\log x_{op}^t$	0.335*** (0.0003)	0.326*** (0.0003)	0.338*** (0.0003)	0.333*** (0.0003)	0.327*** (0.0003)	0.318*** (0.0004)	0.344*** (0.0005)
$\log x_{pd}^t$	0.222*** (0.0003)	0.225*** (0.0003)	0.220*** (0.0003)	0.222*** (0.0003)	0.223*** (0.0003)	0.201*** (0.0005)	0.206*** (0.0005)
$\log Distance$	-1.073*** (0.002)	-1.055*** (0.002)	-0.941*** (0.002)	-1.029*** (0.002)	-0.903*** (0.002)	-1.283*** (0.002)	-1.046*** (0.003)
$\log gdp_o^t$						0.247*** (0.002)	0.272*** (0.002)
$\log gdp_d^t$						0.285*** (0.002)	0.334*** (0.002)
$\log Population_o$						0.691*** (0.002)	0.680*** (0.002)
$\log Population_d$						0.521*** (0.002)	0.525*** (0.002)
$Border_{od}$							0.712** (0.003)
$Colony_{od}$							0.052*** (0.003)
$Language_{od}$							0.545*** (0.002)
$\log Lang.Proximity_{od}$							0.031*** (0.001)
Constant	0.407*** (0.008)	0.465*** (0.008)	-0.080*** (0.009)	0.279*** (0.009)	-0.077*** (0.009)	-8.804*** (0.018)	-10.354*** (0.019)
Observations	10,911,584	10,911,584	10,911,584	10,911,584	10,911,584	10,911,584	10,911,584
Adjusted R <sup>2</sup>	0.469	0.470	0.470	0.469	0.471	0.489	0.494
Residual Std. Error	2.632	2.629	2.630	2.632	2.628	2.583	2.568

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table A2: Bilateral trade volume after two years for periods 2000-2006 (pre-financial crisis), 2007-2012(crisis period) and 2012-2015 (recovery period)

	Dependent variable: $\log x_{opd}^{t+2}$					
	2000-2006		2007-2012		2012-2015	
	(1)	(2)	(3)	(4)	(5)	(6)
$\omega_{opd}^t$	0.116*** (0.001)	0.209*** (0.001)	0.082*** (0.001)	0.180*** (0.001)	0.054*** (0.001)	0.152*** (0.001)
$\Omega_{opd}^{(d)}$	0.109*** (0.001)	0.143*** (0.001)	0.105*** (0.001)	0.143*** (0.001)	0.123*** (0.001)	0.152*** (0.001)
$\Omega_{opd}^{(o)}$	0.027*** (0.001)	0.077*** (0.001)	0.073*** (0.001)	0.105*** (0.001)	0.087*** (0.001)	0.107*** (0.001)
$\log x_{opd}^t$	0.440*** (0.0003)	0.377*** (0.0003)	0.485*** (0.0003)	0.422*** (0.0003)	0.526*** (0.0004)	0.462*** (0.0004)
$\log x_{op}^t$	0.327*** (0.0003)	0.344*** (0.0005)	0.327*** (0.0004)	0.342*** (0.001)	0.310*** (0.0005)	0.325*** (0.001)
$\log x_{pd}^t$	0.223*** (0.0003)	0.206*** (0.0005)	0.178*** (0.0004)	0.165*** (0.001)	0.177*** (0.0005)	0.159*** (0.001)
$\log Distance$	-0.903*** (0.002)	-1.046*** (0.003)	-0.758*** (0.003)	-0.919*** (0.003)	-0.728*** (0.003)	-0.954*** (0.004)
$\log gdp_o^t$		0.272*** (0.002)		0.363*** (0.002)		0.406*** (0.003)
$\log gdp_d^t$		0.334*** (0.002)		0.355*** (0.002)		0.480*** (0.002)
$\log Population_o$		0.680*** (0.002)		0.654*** (0.002)		0.654*** (0.002)
$\log Population_d$		0.525*** (0.002)		0.560*** (0.002)		0.518*** (0.002)
$Border_{od}$		0.712*** (0.003)		0.717*** (0.004)		0.611*** (0.005)
$Colony_{od}$		0.052*** (0.003)		0.127*** (0.004)		0.193*** (0.005)
$Language_{od}$		0.545*** (0.002)		0.441*** (0.003)		0.334*** (0.003)
$\log Lang.Proximity_{od}$		0.031*** (0.001)		0.027*** (0.001)		-0.001 (0.001)
Constant	-0.077*** (0.009)	-10.354*** (0.019)	-0.585*** (0.012)	-11.482*** (0.025)	-1.355*** (0.015)	-12.762*** (0.030)
Observations	10,911,584	10,911,584	7,591,489	7,591,489	5,332,257	5,332,257
Adjusted R <sup>2</sup>	0.471	0.494	0.496	0.516	0.539	0.558
Residual Std. Error	2.628	2.568	2.691	2.637	2.582	2.529

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Appendix C. Summary statistics and correlation table

Table A3: Summary statistics: 2000-2006

Statistic	N	Mean	St. Dev.	Min	Max
$\omega_{opd}^t$	10,911,584	0.000	1.000	-3.366	26.699
$\Omega_{opd}^{(d)}$	10,911,584	0.000	1.000	-0.959	93.647
$\Omega_{opd}^{(o)}$	10,911,584	0.000	1.000	-1.172	51.786
$\log x_{opd}^t$	10,911,584	0.000	1.000	-1.339	4.086
$\log x_{op}^t$	10,911,584	0.000	1.000	-2.547	3.172
$\log x_{pd}^t$	10,911,584	0.000	1.000	-2.544	3.758
$\log Distance$	10,911,584	0.000	1.000	-3.855	1.578
$\log gdp_o^t$	10,911,584	0.000	1.000	-3.118	1.339
$\log gdp_d^t$	10,911,584	0.000	1.000	-2.477	1.541
$\log Population_o$	10,911,584	0.000	1.000	-2.474	2.344
$\log Population_d$	10,911,584	0.000	1.000	-2.324	2.768
$Border_{od}$	10,911,584	0.089	0.284	0	1
$Colony_{od}$	10,911,584	0.060	0.237	0	1
$Language_{od}$	10,911,584	0.155	0.362	0	1
$\log Lang.Proximity_{od}$	10,911,584	0.000	1.000	-0.312	5.050

Table A4: Summary statistics: 2007-2012

Statistic	N	Mean	St. Dev.	Min	Max
$\omega_{opd}^t$	7,591,489	0.000	1.000	-3.347	26.234
$\Omega_{opd}^{(d)}$	7,591,489	0.000	1.000	-0.915	91.287
$\Omega_{opd}^{(o)}$	7,591,489	0.000	1.000	-1.171	25.338
$\log x_{opd}^t$	7,591,489	0.000	1.000	-1.374	3.977
$\log x_{op}^t$	7,591,489	0.000	1.000	-2.673	3.094
$\log x_{pd}^t$	7,591,489	0.000	1.000	-2.830	3.745
$\log Distance$	7,591,489	0.000	1.000	-3.951	1.580
$\log gdp_o^t$	7,591,489	0.000	1.000	-3.285	1.479
$\log gdp_d^t$	7,591,489	0.000	1.000	-2.633	1.637
$\log Population_o$	7,591,489	0.000	1.000	-1.990	2.328
$\log Population_d$	7,591,489	0.000	1.000	-1.916	2.749
$Border_{od}$	7,591,489	0.082	0.274	0	1
$Colony_{od}$	7,591,489	0.053	0.224	0	1
$Language_{od}$	7,591,489	0.153	0.360	0	1
$\log Lang.Proximity_{od}$	7,591,489	-0.000	1.000	-0.298	5.349

Table A5: Summary statistics: 2012-2015

Statistic	N	Mean	St. Dev.	Min	Max
$\omega_{opd}^t$	5,332,257	0.000	1.000	-3.373	24.428
$\Omega_{opd}^{(d)}$	5,332,257	0.000	1.000	-0.965	98.891
$\Omega_{opd}^{(o)}$	5,332,257	0.000	1.000	-1.185	34.256
$\log x_{opd}^t$	5,332,257	0.000	1.000	-1.377	3.976
$\log x_{op}^t$	5,332,257	0.000	1.000	-2.787	3.116
$\log x_{pd}^t$	5,332,257	0.000	1.000	-2.900	3.591
$\log Distance$	5,332,257	0.000	1.000	-3.993	1.581
$\log gdp_o^t$	5,332,257	0.000	1.000	-3.193	1.527
$\log gdp_d^t$	5,332,257	0.000	1.000	-2.393	1.652
$\log Population_o$	5,332,257	0.000	1.000	-1.994	2.340
$\log Population_d$	5,332,257	0.000	1.000	-1.861	2.754
$Border_{od}$	5,332,257	0.078	0.268	0	1
$Colony_{od}$	5,332,257	0.052	0.222	0	1
$Language_{od}$	5,332,257	0.141	0.348	0	1
$\log Lang. Proximity_{od}$	5,332,257	0.000	1.000	-0.298	5.383

Table A6: Correlation Matrix: 2000-2006

	$\omega_{opd}^l$	$\Omega_{opd}^{(d)}$	$\Omega_{opd}^{(o)}$	$\log x_{opd}^l$	$\log x_{op}^l$	$\log x_{opd}^l$	$\log Distance$	$\log gdp_o^l$	$\log gdp_o^l$	$\log Population_o$	$\log Population_d$	$Border_{ad}$	$Colony_{ad}$	$Language_{ad}$	$\log LangProximity_{ad}$
$\omega_{opd}^l$	1	0.038	0.080	0.132	0.215	-0.020	-0.026	-0.006	-0.055	-0.060	-0.081	-0.017	-0.025	-0.035	-0.018
$\Omega_{opd}^{(d)}$	0.038	1	0.287	0.049	-0.182	-0.009	-0.536	-0.037	0.091	-0.188	-0.150	0.229	0.013	0.044	0.095
$\Omega_{opd}^{(o)}$	0.080	0.287	1	0.064	-0.003	-0.035	-0.457	0.123	0.025	-0.176	-0.123	0.126	0.005	-0.068	0.100
$\log x_{opd}^l$	0.132	0.049	0.064	1	0.459	0.344	-0.117	0.170	0.159	0.135	0.104	0.110	0.063	0.019	0.132
$\log x_{op}^l$	0.215	-0.182	-0.003	0.459	1	0.171	0.239	0.345	-0.120	0.288	-0.083	-0.190	-0.028	-0.163	0.015
$\log x_{pd}^l$	-0.020	-0.009	-0.035	0.344	0.171	1	0.078	-0.038	0.486	-0.140	0.331	-0.072	-0.026	-0.083	0.085
$\log Distance$	-0.026	-0.536	-0.457	-0.117	0.239	0.078	1	-0.015	-0.066	0.285	0.211	-0.486	-0.042	-0.044	-0.217
$\log gdp_o^l$	-0.006	-0.057	0.123	0.170	0.345	-0.038	-0.015	1	0.016	-0.338	-0.041	-0.115	0.029	-0.072	0.110
$\log gdp_d^l$	-0.055	0.091	0.025	0.159	-0.120	0.486	-0.066	0.016	1	-0.084	-0.148	-0.032	-0.001	-0.065	0.138
$\log Population_o$	-0.060	-0.188	-0.176	0.135	0.288	-0.140	0.285	-0.338	-0.084	1	-0.013	-0.052	0.027	-0.044	0.004
$\log Population_d$	-0.081	-0.150	-0.123	0.104	-0.083	0.331	0.211	-0.041	-0.148	-0.013	1	-0.022	0.019	-0.022	0.058
$Border_{ad}$	-0.017	0.229	0.126	0.110	-0.190	-0.072	-0.486	-0.115	-0.032	-0.052	-0.022	1	0.130	0.170	0.114
$Colony_{ad}$	-0.025	0.013	0.005	0.063	-0.028	-0.026	-0.042	0.029	-0.001	0.027	0.019	0.130	1	0.264	0.071
$Language_{ad}$	-0.035	0.044	-0.068	0.019	-0.163	-0.083	-0.044	-0.072	-0.065	-0.044	-0.022	0.170	0.264	1	-0.103
$\log LangProximity_{ad}$	-0.018	0.095	0.100	0.132	0.015	0.085	-0.217	0.110	0.138	0.004	0.058	0.114	0.071	-0.103	1

Table A7: Correlation Matrix: 2007-2012

	$\omega_{opd}^l$	$\Omega_{opd}^{(d)}$	$\Omega_{opd}^{(o)}$	$\log x_{opd}^l$	$\log x_{op}^l$	$\log x_{opd}^l$	$\log Distance$	$\log gdp_o^l$	$\log gdp_o^l$	$\log Population_o$	$\log Population_d$	$Border_{ad}$	$Colony_{ad}$	$Language_{ad}$	$\log LangProximity_{ad}$
$\omega_{opd}^l$	1	0.036	0.083	0.139	0.233	-0.012	-0.030	-0.001	-0.049	-0.066	-0.086	-0.018	-0.024	-0.027	-0.015
$\Omega_{opd}^{(d)}$	0.036	1	0.273	0.059	-0.161	-0.022	-0.512	-0.024	0.093	-0.179	-0.153	0.217	0.013	0.053	0.087
$\Omega_{opd}^{(o)}$	0.083	0.273	1	0.086	-0.004	-0.036	-0.452	0.139	0.061	-0.167	-0.135	0.129	0.004	-0.050	0.101
$\log x_{opd}^l$	0.139	0.059	0.086	1	0.481	0.343	-0.128	0.137	0.138	0.166	0.100	0.126	0.057	0.004	0.132
$\log x_{op}^l$	0.233	-0.161	-0.004	0.481	1	0.206	0.219	0.306	-0.107	0.298	-0.088	-0.168	-0.028	-0.182	0.021
$\log x_{pd}^l$	-0.012	-0.022	-0.036	0.343	0.206	1	0.078	-0.036	0.414	-0.149	0.342	-0.063	-0.017	-0.091	0.083
$\log Distance$	-0.030	-0.512	-0.452	-0.128	0.219	0.078	1	-0.043	-0.104	0.290	0.224	-0.477	-0.046	-0.053	-0.213
$\log gdp_o^l$	-0.001	-0.024	0.139	0.137	0.306	-0.036	-0.043	1	0.034	-0.349	-0.053	-0.090	0.038	-0.117	0.131
$\log gdp_d^l$	-0.049	0.093	0.061	0.138	-0.107	0.414	-0.104	0.034	1	-0.097	-0.177	-0.014	0.012	-0.096	0.160
$\log Population_o$	-0.066	-0.179	-0.167	0.166	0.298	-0.149	0.290	-0.349	-0.097	1	-0.012	-0.046	0.025	-0.034	0.001
$\log Population_d$	-0.086	-0.153	-0.135	0.100	-0.088	0.342	0.224	-0.053	-0.177	-0.012	1	-0.020	0.022	-0.007	0.045
$Border_{ad}$	-0.018	0.217	0.129	0.126	-0.168	-0.063	-0.477	-0.090	-0.014	-0.046	-0.020	1	0.126	0.169	0.111
$Colony_{ad}$	-0.024	0.013	0.004	0.057	-0.028	-0.017	-0.046	0.038	0.012	0.025	0.022	0.126	1	0.252	0.070
$Language_{ad}$	-0.027	0.053	-0.050	0.004	-0.182	-0.091	-0.053	-0.117	-0.096	-0.034	-0.007	0.169	0.252	1	-0.099
$\log LangProximity_{ad}$	-0.015	0.087	0.101	0.132	0.021	0.083	-0.213	0.131	0.160	0.001	0.045	0.111	0.070	-0.099	1

Table A8: Correlation Matrix: 2012-2015

	$\omega_{opd}^t$	$\Omega_{opd}^{(t)}$	$\Omega_{opd}^{(o)}$	$\log x_{opd}^t$	$\log x_{opd}^o$	$\log Distance$	$\log gdp_o^t$	$\log gdp_d^t$	$\log Population_o$	$\log Population_d$	$Colony_{od}$	$Border_{od}$	$Language_{od}$	$LangProximity_{od}$
$\omega_{opd}^t$	1	0.033	0.084	0.143	0.235	-0.006	-0.025	-0.012	-0.044	-0.064	-0.084	-0.021	-0.023	-0.014
$\Omega_{opd}^{(t)}$	0.033	1	0.283	0.069	-0.167	-0.023	-0.542	-0.036	0.095	-0.180	-0.154	0.236	0.017	0.044
$\Omega_{opd}^{(o)}$	0.084	0.283	1	0.082	-0.015	-0.042	-0.440	0.118	0.059	-0.161	-0.132	0.125	0.002	-0.053
$\log x_{opd}^t$	0.143	0.069	0.082	1	0.480	0.350	-0.125	0.109	0.128	0.185	0.108	0.138	0.061	0.022
$\log x_{opd}^o$	0.235	-0.167	-0.015	0.480	1	0.213	0.222	0.271	-0.124	0.309	-0.083	-0.157	-0.031	-0.162
$\log x_{opd}^t$	-0.006	-0.023	-0.042	0.350	0.213	1	0.085	-0.055	0.387	-0.148	0.349	-0.054	-0.021	0.076
$\log Distance$	-0.025	-0.542	-0.440	-0.125	0.222	0.085	1	-0.010	-0.093	0.283	0.227	-0.472	-0.051	-0.042
$\log gdp_o^t$	-0.012	-0.036	0.118	0.109	0.271	-0.055	-0.010	1	0.019	-0.332	-0.048	-0.089	0.030	-0.104
$\log gdp_d^t$	-0.044	0.095	0.059	0.128	-0.124	0.387	-0.093	0.019	1	-0.101	-0.169	-0.006	0.008	-0.075
$\log Population_o$	-0.064	-0.180	-0.161	0.185	0.309	-0.148	0.283	-0.332	-0.101	1	-0.007	-0.035	0.028	-0.005
$\log Population_d$	-0.084	-0.154	-0.132	0.108	-0.083	0.349	0.227	-0.048	-0.169	-0.007	1	-0.020	0.021	0.003
$Border_{od}$	-0.021	0.236	0.125	0.138	-0.157	-0.054	-0.472	-0.089	-0.006	-0.035	-0.020	1	0.136	0.165
$Colony_{od}$	-0.023	0.017	0.002	0.061	-0.031	-0.021	-0.051	0.030	0.008	0.028	0.021	0.136	1	0.260
$Language_{od}$	-0.026	0.044	-0.053	0.022	-0.162	-0.073	-0.042	-0.104	-0.075	-0.005	0.003	0.165	0.260	1
$\log LangProximity_{od}$	-0.014	0.095	0.101	0.124	0.009	0.076	-0.215	0.116	0.158	-0.001	0.042	0.114	0.070	-0.092
														1

**Appendix D. Relationship between bilateral trade volume after two years and the three learning channels by products' competitiveness**

Table A9: Summary statistics: New exporters

Statistic	N	Mean	St. Dev.	Min	Max
$\omega_{opd}^t$	922,092	0.000	1.000	-2.520	15.486
$\Omega_{opd}^{(d)}$	922,092	0.000	1.000	-0.720	55.607
$\Omega_{opd}^{(o)}$	922,092	0.000	1.000	-1.167	25.297
$\log x_{opd}^t$	922,092	0.000	1.000	-1.000	4.508
$\log x_{op}^t$	922,092	0.000	1.000	-1.415	3.296
$\log x_{pd}^t$	922,092	0.000	1.000	-2.918	2.886
$\log Distance$	922,092	0.000	1.000	-3.309	1.785
$\log gdp_o^t$	922,092	0.000	1.000	-2.429	2.330
$\log gdp_d^t$	922,092	0.000	1.000	-2.477	1.455
$\log Population_o$	922,092	0.000	1.000	-2.220	3.552
$\log Population_d$	922,092	0.000	1.000	-2.383	2.676
$Border_{od}$	922,092	0.170	0.376	0	1
$Colony_{od}$	922,092	0.058	0.234	0	1
$Language_{od}$	922,092	0.292	0.455	0	1
$\log Lang.Proximity_{od}$	922,092	0.000	1.000	-0.239	7.600

Table A10: Summary statistics: Nascent exporters

Statistic	N	Mean	St. Dev.	Min	Max
$\omega_{opd}^t$	463,388	0.000	1.000	-3.112	15.288
$\Omega_{opd}^{(d)}$	463,388	0.000	1.000	-1.050	91.632
$\Omega_{opd}^{(o)}$	463,388	0.000	1.000	-1.248	25.615
$\log x_{opd}^t$	463,388	0.000	1.000	-1.256	4.077
$\log x_{op}^t$	463,388	0.000	1.000	-2.905	3.434
$\log x_{pd}^t$	463,388	0.000	1.000	-3.336	3.460
$\log Distance$	463,388	0.000	1.000	-3.659	1.581
$\log gdp_o^t$	463,388	0.000	1.000	-3.222	1.739
$\log gdp_d^t$	463,388	0.000	1.000	-2.493	1.514
$\log Population_o$	463,388	0.000	1.000	-2.137	2.852
$\log Population_d$	463,388	0.000	1.000	-2.327	2.811
$Border_{od}$	463,388	0.091	0.287	0	1
$Colony_{od}$	463,388	0.042	0.200	0	1
$Language_{od}$	463,388	0.160	0.366	0	1
$\log Lang.Proximity_{od}$	463,388	0.000	1.000	-0.299	5.712

Table A11: Summary statistics: Experienced exporters

Statistic	N	Mean	St. Dev.	Min	Max
$\omega_{opd}^t$	8,045,262	0.000	1.000	-3.738	28.851
$\Omega_{opd}^{(d)}$	8,045,262	0.000	1.000	-1.106	108.927
$\Omega_{opd}^{(o)}$	8,045,262	0.000	1.000	-1.178	50.902
$\log x_{opd}^t$	8,045,262	0.000	1.000	-1.467	3.950
$\log x_{op}^t$	8,045,262	0.000	1.000	-3.373	3.456
$\log x_{pd}^t$	8,045,262	0.000	1.000	-2.559	4.114
$\log Distance$	8,045,262	0.000	1.000	-4.085	1.552
$\log gdp_o^t$	8,045,262	0.000	1.000	-3.623	1.234
$\log gdp_d^t$	8,045,262	0.000	1.000	-2.472	1.570
$\log Population_o$	8,045,262	0.000	1.000	-2.612	2.237
$\log Population_d$	8,045,262	0.000	1.000	-2.316	2.801
$Border_{od}$	8,045,262	0.069	0.254	0	1
$Colony_{od}$	8,045,262	0.062	0.241	0	1
$Language_{od}$	8,045,262	0.128	0.334	0	1
$\log Lang.Proximity_{od}$	8,045,262	0.000	1.000	-0.325	4.763

Table A12: Correlation Matrix: New exporters

	$\omega_{opd}^t$	$\Omega_{opd}^{(d)}$	$\Omega_{opd}^{(o)}$	$\log x_{opd}^t$	$\log x_{op}^t$	$\log x_{pd}^t$	$\log Distance$	$\log gdp_o^t$	$\log gdp_d^t$	$\log Population_o$	$\log Population_d$	$Border_{od}$	$Colony_{od}$	$Language_{od}$	$\log Lang.Proximity_{od}$
$\omega_{opd}^t$	1	0.064	0.127	0.071	0.112	-0.095	-0.108	0.065	-0.107	-0.125	-0.117	0.027	-0.033	0.035	-0.033
$\Omega_{opd}^{(d)}$	0.064	1	0.241	-0.001	-0.077	-0.166	-0.387	0.001	-0.076	-0.109	-0.200	0.165	-0.031	0.123	0.015
$\Omega_{opd}^{(o)}$	0.127	0.241	1	0.083	-0.006	-0.179	-0.462	0.083	-0.112	-0.151	-0.189	0.206	0.014	-0.014	0.073
$\log x_{opd}^t$	0.071	-0.001	0.083	1	0.429	0.211	-0.129	0.154	0.019	0.027	0.015	0.114	0.049	0.043	0.069
$\log x_{op}^t$	0.112	-0.077	-0.006	0.429	1	0.344	0.200	0.373	-0.051	0.113	-0.074	-0.193	-0.088	-0.198	0.024
$\log x_{pd}^t$	-0.095	-0.166	-0.179	0.211	0.344	1	0.395	0.068	0.611	0.006	0.438	-0.322	0.083	-0.292	0.079
$\log Distance$	-0.108	-0.387	-0.462	-0.129	0.200	0.395	1	-0.023	0.224	0.250	0.335	-0.563	-0.002	-0.212	-0.106
$\log gdp_o^t$	0.065	0.001	0.083	0.154	0.373	0.068	-0.023	1	0.064	-0.442	-0.083	-0.098	-0.065	-0.080	0.109
$\log gdp_d^t$	-0.107	-0.076	-0.112	0.019	-0.051	0.611	0.224	0.064	1	-0.057	0.013	-0.278	0.086	-0.202	0.079
$\log Population_o$	-0.125	-0.109	-0.151	0.027	0.113	0.006	0.250	-0.442	-0.057	1	0.041	-0.031	0.010	-0.007	0.006
$\log Population_d$	-0.117	-0.200	-0.189	0.015	-0.074	0.438	0.335	-0.083	0.013	0.041	1	-0.093	0.135	-0.158	0.067
$Border_{od}$	0.027	0.165	0.206	0.114	-0.193	-0.322	-0.563	-0.098	-0.278	-0.031	-0.093	1	0.084	0.229	0.061
$Colony_{od}$	-0.033	-0.031	0.014	0.049	-0.088	0.083	-0.002	-0.065	0.086	0.010	0.135	0.084	1	0.131	0.167
$Language_{od}$	0.035	0.123	-0.014	0.043	-0.198	-0.292	-0.212	-0.080	-0.202	-0.007	-0.158	-0.229	0.131	1	-0.113
$\log Lang.Proximity_{od}$	-0.033	0.015	0.073	0.069	0.024	0.079	-0.106	0.109	0.079	0.006	0.067	0.061	0.167	-0.113	1

Table A13: Correlation Matrix: Nascent exporters

	$\omega_{opd}^t$	$\Omega_{opd}^{(d)}$	$\Omega_{opd}^{(o)}$	$\log x_{opd}^t$	$\log x_{op}^t$	$\log x_{pd}^t$	$\log Distance$	$\log gdp_o^t$	$\log gdp_d^t$	$\log Population_o$	$\log Population_d$	$Border_{od}$	$Colony_{od}$	$Language_{od}$	$\log Lang.Proximity_{od}$
$\omega_{opd}^t$	1	0.102	0.122	0.071	0.114	-0.018	-0.135	-0.039	-0.070	-0.175	-0.098	0.009	-0.017	0.024	-0.053
$\Omega_{opd}^{(d)}$	0.102	1	0.405	0.125	-0.187	-0.103	-0.618	-0.011	0.072	-0.202	-0.198	0.267	0.070	0.050	0.109
$\Omega_{opd}^{(o)}$	0.122	0.405	1	0.129	-0.013	-0.073	-0.565	0.126	0.005	-0.234	-0.175	0.200	0.051	-0.075	0.098
$\log x_{opd}^t$	0.071	0.125	0.129	1	0.361	0.360	-0.194	0.056	0.128	0.048	0.094	0.191	0.104	0.099	0.107
$\log x_{op}^t$	0.114	-0.187	-0.013	0.361	1	0.379	0.201	0.157	-0.107	0.149	-0.042	-0.142	-0.063	-0.127	-0.044
$\log x_{pd}^t$	-0.018	-0.103	-0.073	0.360	0.379	1	0.176	0.095	0.590	-0.033	0.391	-0.151	0.003	-0.158	0.103
$\log Distance$	-0.135	-0.618	-0.565	-0.194	0.201	0.176	1	-0.022	-0.026	0.277	0.248	-0.476	-0.118	0.0001	-0.216
$\log gdp_o^t$	-0.039	-0.011	0.126	0.056	0.157	0.095	-0.022	1	0.061	-0.574	-0.001	-0.074	-0.032	-0.095	0.091
$\log gdp_d^t$	-0.070	0.072	0.005	0.128	-0.107	0.590	-0.026	0.061	1	-0.109	-0.115	-0.075	0.030	-0.103	0.124
$\log Population_o$	-0.175	-0.202	-0.234	0.048	0.149	-0.033	0.277	-0.574	-0.109	1	0.009	-0.011	0.036	0.044	-0.006
$\log Population_d$	-0.098	-0.198	-0.175	0.094	-0.042	0.391	0.248	-0.001	-0.115	0.009	1	-0.049	0.026	-0.025	0.060
$Border_{od}$	0.009	0.267	0.200	0.191	-0.142	-0.151	-0.476	-0.074	-0.075	-0.011	-0.049	1	0.231	0.123	0.132
$Colony_{od}$	-0.017	0.070	0.051	0.104	-0.063	0.003	-0.118	-0.032	0.030	0.036	0.026	0.231	1	0.133	0.160
$Language_{od}$	0.024	0.050	-0.075	0.099	-0.127	-0.158	0.0001	-0.095	-0.103	0.044	-0.025	0.123	0.133	1	-0.100
$\log Lang.Proximity_{od}$	-0.053	0.109	0.098	0.107	-0.044	0.103	-0.216	0.091	0.124	-0.006	0.060	0.132	0.160	-0.100	1

Table A14: Correlation Matrix: Experienced exporters

	$\omega_{opd}^f$	$\Omega_{opd}^{(d)}$	$\Omega_{opd}^{(o)}$	$\log x_{opd}^f$	$\log x_{opd}^o$	$\log Distance$	$\log gdp_o^f$	$\log gdp_d^f$	$\log Population_o$	$\log Population_d$	$Colony_{ad}$	$Border_{ad}$	$Language_{ad}$	$LangProximity_{ad}$
$\omega_{opd}^f$	1	0.047	0.066	0.121	0.205	0.036	-0.034	-0.052	-0.048	-0.095	-0.072	-0.006	-0.028	-0.044
$\Omega_{opd}^{(d)}$	0.047	1	0.327	0.113	-0.178	0.033	-0.607	0.004	0.163	-0.191	-0.135	0.236	0.022	-0.026
$\Omega_{opd}^{(o)}$	0.066	0.327	1	0.049	-0.035	0.009	-0.469	0.128	0.053	-0.208	-0.106	0.107	-0.004	-0.071
$\log x_{opd}^f$	0.121	0.113	0.049	1	0.417	0.468	-0.165	0.103	0.209	0.093	0.136	0.154	0.062	0.050
$\log x_{opd}^o$	0.205	-0.178	-0.035	0.417	1	0.303	0.185	0.205	-0.138	0.209	-0.082	-0.131	-0.022	-0.081
$\log x_{pd}^f$	0.036	0.033	0.009	0.468	0.303	1	0.030	0.036	0.469	-0.115	0.311	-0.017	-0.045	-0.062
$\log Distance$	-0.034	-0.607	-0.469	-0.165	0.185	0.030	1	-0.082	-0.132	0.255	0.186	-0.447	-0.039	0.028
$\log gdp_o^f$	-0.052	0.004	0.128	0.103	0.205	0.036	-0.082	1	0.025	-0.468	-0.024	-0.082	0.051	-0.010
$\log gdp_d^f$	-0.048	0.163	0.053	0.209	-0.138	0.469	-0.132	0.025	1	-0.085	-0.185	0.037	-0.018	-0.036
$\log Population_o$	-0.095	-0.191	-0.208	0.093	0.209	-0.115	0.255	-0.468	-0.085	1	-0.017	-0.015	0.026	-0.007
$\log Population_d$	-0.072	-0.135	-0.106	0.136	-0.082	0.311	0.186	-0.024	-0.185	-0.017	1	0.004	-0.002	0.011
$Border_{ad}$	-0.006	0.236	0.107	0.154	-0.131	-0.017	-0.447	-0.082	0.037	-0.015	0.004	1	0.125	0.130
$Colony_{ad}$	-0.028	0.022	-0.004	0.062	-0.022	-0.045	-0.039	0.051	-0.018	0.026	-0.002	0.125	1	0.332
$Language_{ad}$	-0.044	-0.026	-0.071	0.050	-0.081	-0.062	0.028	-0.010	-0.036	-0.007	0.011	0.130	0.332	1
$LangProximity_{ad}$	-0.016	0.138	0.101	0.137	-0.013	0.109	-0.250	0.097	0.153	-0.015	0.062	0.137	0.044	-0.095
														1

## Appendix E. Bilateral trade volume after two years for primary products, resource-based, low-tech, medium-tech, and high-tech manufactures

Following Lall (2000), we exclude "special transactions" such as electric current, cinema film, printed matter, fold, coins, and pets.

Table A15: Lall's classification

	Dependent variable: $\log x_{opd}^{t+2}$				
	Primary Product	Resource-based Manufactures	Low-tech Manufactures	Medium-tech Manufactures	High-tech Manufactures
$\omega_{opd}^t$	0.183*** (0.003)	0.164*** (0.002)	0.204*** (0.001)	0.203*** (0.003)	0.229*** (0.003)
$\Omega_{opd}^{(d)}$	0.152*** (0.003)	0.144*** (0.002)	0.131*** (0.002)	0.154*** (0.003)	0.128*** (0.003)
$\Omega_{opd}^{(o)}$	0.098*** (0.003)	0.076*** (0.002)	0.072*** (0.002)	0.084*** (0.003)	0.078*** (0.003)
$\log x_{opd}^t$	0.343*** (0.001)	0.375*** (0.001)	0.409*** (0.0005)	0.362*** (0.001)	0.387*** (0.001)
$\log x_{op}^t$	0.353*** (0.001)	0.341*** (0.001)	0.325*** (0.001)	0.369*** (0.002)	0.333*** (0.001)
$\log x_{pd}^t$	0.565*** (0.004)	0.490*** (0.003)	0.428*** (0.002)	0.509*** (0.004)	0.615*** (0.004)
$\log Distance$	-0.421*** (0.004)	-0.439*** (0.003)	-0.495*** (0.002)	-0.544*** (0.004)	-0.477*** (0.004)
$\log gdp_o^t$	0.103*** (0.003)	0.082*** (0.002)	0.166*** (0.002)	0.132*** (0.004)	0.296*** (0.004)
$\log gdp_d^t$	0.151*** (0.004)	0.184*** (0.002)	0.292*** (0.002)	0.164*** (0.004)	0.271*** (0.004)
$\log Population_o$	0.326*** (0.003)	0.419*** (0.002)	0.482*** (0.002)	0.511*** (0.004)	0.517*** (0.004)
$\log Population_d$	0.298*** (0.003)	0.323*** (0.002)	0.339*** (0.002)	0.349*** (0.003)	0.379*** (0.003)
$Border_{od}$	0.741*** (0.010)	0.754*** (0.007)	0.668*** (0.006)	0.733*** (0.011)	0.664*** (0.011)
$Colony_{od}$	-0.112*** (0.011)	-0.016** (0.007)	0.077*** (0.006)	0.037*** (0.011)	0.182*** (0.012)
$Language_{od}$	0.395*** (0.008)	0.479*** (0.005)	0.553*** (0.004)	0.556*** (0.008)	0.619*** (0.008)
$\log Lang.Proximity_{od}$	0.043*** (0.003)	0.038*** (0.002)	0.031*** (0.001)	0.036*** (0.003)	0.021*** (0.003)
Constant	0.653*** (0.020)	0.484*** (0.015)	0.332*** (0.012)	0.274*** (0.025)	0.529*** (0.024)
Observations	1,127,670	2,241,432	3,314,246	1,110,342	1,049,765
Adjusted R <sup>2</sup>	0.399	0.446	0.527	0.471	0.565
Residual Std. Error	2.874	2.658	2.371	2.689	2.477

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01