The Gravity of Intermediate Goods*

Paola Conconi

Université Libre de Bruxelles (ECARES), CEPR and CESifo

Glenn Magerman

Université Libre de Bruxelles (ECARES) and National Bank of Belgium

Afrola Plaku

Université Libre de Bruxelles (ECARES)

*Postal address (all authors): European Center for Advanced Research in Economics and Statistics (ECARES), Université Libre de Bruxelles, Avenue Franklin Roosevelt 42, 1050 Brussels. Paola Conconi (corresponding coauthor): pconconi@ulb.ac.be. Glenn Magerman: glenn.magerman@ulb.ac.be. Afrola Plaku: Afrola.Plaku@ulb.ac.be.

Abstract

One of the puzzles of the gravity literature is the persistent effect of distance on trade flows, despite the dramatic fall in trade costs during the last few decades (Disdier and Head, 2008). A possible reason for the "distance puzzle" is that trade in intermediate goods – which has risen dramatically during this period due to the emergence of global value chains – may be more sensitive to distance than trade in final goods. Using a dataset of bilateral import flows that covers 5000 products and more than 200 countries over the 1998-2011 period, we show that intermediate goods are indeed more sensitive to distance than are final goods and that differentiated inputs exhibit the highest distance elasticity. The results are robust to including different sets of controls, and using different samples and econometric methodologies. They suggest that sourcing inputs from nearby countries helps final good producers to coordinate with their suppliers, monitor their production, and ensure the timely delivery of inputs that need to be tailored to their needs.

JEL classification: F14, F23.

Keywords: Distance, final goods, intermediate goods, product differentiation.

1 Introduction

During the last three decades, advances in information and communication technology and falling trade barriers have made it easier for firms to source key inputs from foreign suppliers and to fragment their production processes across countries. Research and development, design, production of parts, assembly, and marketing and branding – which were previously performed in close proximity and within the same firm – are increasingly fragmented across the globe and across firms. As a result of the emergence of global value chains (GVCs), trade in intermediate inputs now accounts for as much as two-thirds of international trade (Johnson and Noguera, 2012). These trends have led some to announce the "death of distance" (Cairncross, 1997) and to argue that "certainly it is an exaggeration to claim that moving goods is free, but it is becoming an increasingly apt assumption" (Glaeser and Kohlhase, 2004).

At the same time, the gravity literature of international trade has emphasized the persistent effect of distance on bilateral trade. In an influential meta-analysis study, Disdier and Head (2008) investigate the trends in the variation of 1467 distance estimates from 103 papers and provide systematic evidence that "the estimated negative impact of distance on trade rose around the middle of the century and has remained persistently high since then." They find distance effects to be persistent in two senses: they hold up in a very wide range of samples and methodologies, and they are not declining in more recent data.

In this paper, we argue that the emergence of GVCs can help to explain the "distance puzzle." While fragmenting production across firms and countries has become easier, contractual frictions remain a significant obstacle to the globalization of value chains. On top of the inherent difficulties associated with designing richly contingent contracts, international transactions suffer from a disproportionately low level of enforcement of contract clauses and legal remedies (Antràs, 2016). Contracting frictions are particularly severe when they involve relationship-specific inputs, i.e., differentiated/non-standardized inputs that must be tailored to the need of final good producers (Nunn, 2007). Sourcing these inputs from nearby countries can allow producers to coordinate better with suppliers and to monitor their production. Using more distant suppliers can give rise to problems and delays in the production of key inputs, which can disrupt the entire supply chain.

In light of these contracting frictions, it is then not surprising that GVCs are actually regional in nature. As pointed out by Baldwin (2013), trade in intermediate goods is concentrated within "Factory North America," "Factory Europe," and "Factory Asia." This can partly be attributed to the role of regional trade agreements, which can distort input trade towards suppliers in partner countries (e.g., Conconi *et al.*, 2018). In this paper, we argue that the regional nature of supply chains can also be explained by the higher sensitivity to distance of intermediate goods, and in particular of differentiated inputs.

A few studies have already noted that intermediate goods are more sensitive to distance than final goods (Miroudot *et al.*, 2009; Bergstrand and Egger, 2010; Baldwin and Taglioni, 2011). In this paper, we provide systematic evidence for this difference and explore the mechanisms that underlie it.

We construct a product-level panel dataset, which covers more than 5000 products and more than 200 countries and territories over the 1998-2011 period. We employ the United Nations's Broad Economic Categories (BEC) classification to distinguish between final and intermediate goods, and the classification by Rauch (1999) to distinguish between homogeneous and differentiated goods.

Using this dataset, we first show that intermediate goods are indeed more sensitive to distance. This result is robust to including different sets of fixed effects and controls, exploiting cross-sectional and time series variation in trade flows, and using alternative econometric methodologies.

One possible explanation for this result is that some inputs are raw materials, which can be less easily traded across long distances due to their low value/weight ratios (Miroudot *et al.*, 2009). We dismiss this mechanism: we show that the higher sensitivity to distance of intermediate goods is robust to excluding raw materials. In fact, the difference in the distance elasticity between final and intermediate goods is even larger when we exclude raw materials.

Finally, we show that differentiated inputs are the most sensitive to distance. This is in line with the idea that final good producers tend to source relationship-specific inputs, which need to be tailored to their specific needs, from nearby suppliers. Sourcing differentiated inputs from suppliers in more distant countries could give rise to various problems/delays, which would disrupt their supply chain.

2 Related Literature

Our analysis is related to two main streams of studies: the literature on global sourcing and the organization of supply chains, and the literature on the distance puzzle.

First, the emergence of GVCs has motivated a stream of studies on firms' sourcing decisions. A growing body of literature studies the causes and consequences of increasing production fragmentation across countries and the global sourcing of inputs. Several studies emphasize the productivity-enhancing effects of input trade: they show that it allows firms to reduce production costs by accessing novel, cheaper, or higher-quality foreign inputs (e.g., Amiti and Konings, 2007; Goldberg *et al.*, 2010). Other studies examine the determinants of firm boundary choices along value chains (Antràs and Chor, 2013; Alfaro *et al.*, 2019). Another stream of the literature emphasizes selection effects. Bernard *et al.* (2007) show that US importers are on average more than twice as large than non-importers. Antràs *et al.* (2017) develop a theoretical model that can rationalize this heterogeneity. They examine the margins of global sourcing in a multi-country environment. In their model, a firm can add one country to the set of countries from which it is able to import, but this requires incurring a market-specific fixed cost. As a result, relatively unproductive firms opt out of importing from countries that are not particularly attractive sources of inputs. The global sourcing strategy of a firm is to determine the set of countries from which to source inputs, based on cross-country differences in technology, trade costs, and wages. A related study by Blaum *et al.* (2018) develop a multi-country quantitative model to study the effect of imported inputs on firm-level and aggregate productivity.

Second, many studies have provided econometric evidence suggesting that the elasticity of bilateral trade with respect to distance has increased over time. Leamer and Levinsohn (1995) were the first to draw attention to this trend. They conclude that – contrary to popular notions of globalization – the world is not "getting smaller." They argue, in line with the gravity approach, that the driving force that underlies globalization is not lower effective distance-barriers, but increased dispersion of economic mass around the globe. Combes *et al.* (2008) estimate year-by-year distance coefficients. They document a pattern of rising coefficients since the 1950s. Berthelon and Freund (2004) study industry-level trade and find that 75 percent of industries do not exhibit significant changes in the distance effect. The significant changes are almost all in the direction of a larger distance effect over the 1985-2000 period. Many other studies on the "distance puzzle" are reviewed in the meta-analysis study by Disdier and Head (2008). Our paper exploits product-level data on bilateral trade flows to show that inputs are more sensitive to distance than are final goods – particularly when the inputs are differentiated and thus need to be tailored to the needs of final good producers.[1]

3 Data

We perform our analysis at the product level, with the use of disaggregated data on bilateral trade flows from the United Nations's Comtrade dataset. The trade data are defined at the 6-digit level of the Harmonized System classification (HS6), which covers more than 5000 products. Our sample covers 212 countries and territories over the period 1998-2011.

Given our interest in GVCs and sourcing decisions, we focus on imports rather than

[1] Our analysis builds on earlier studies that have estimated gravity regressions at the product or sectoral level (e.g., Anderson and Yotov, 2010; Imbs and Mejean, 2017).

exports. The dependent variable in our regressions is the log of $Imports_{kijt}$, the value of imports of HS6 product k of country i from country j in year t (in current US dollars).

To distinguish between final and intermediate goods, we use the United Nations's BEC classification, in line with previous studies (e.g., Bergstrand and Egger 2010; Alfaro *et al.*, 2019). The BEC is a subjective expert judgment classification that distinguishes products according to their end-use. The original data are based on the SITC classification, but the UN Trade Statistics Division provides a concordance table that allows conversions at the HS6 level. The BEC distinguishes between intermediate, capital, and consumption goods. As an example, roasted, decaffeinated coffee (HS 090121) is classified as a consumption good, while its non-roasted version (HS 090112) is classified as an intermediate. The variable $Intermediate_k$ is a dummy variable that is equal to 1 if product k is classified as intermediate according to the BEC classification and is zero otherwise.[2]

To distinguish between homogeneous and differentiated goods, we rely on the well-known classification by Rauch (1999). This is based on the SITC rev. 2 classification, so we employ the UN Concordance Tables to convert the data at the HS-6 level. Rauch classifies products in three different categories: homogeneous goods, which are traded in organized exchanges; goods that are not traded in organized exchanges, but for which a published reference price can be found; and differentiated goods, which do not fall in either of the two previous categories. We use the conservative version of the Rauch classification. The variable $Differentiated_k$ is a dummy variable equal to 1 if good k is "branded" goods, which does not fall in the two homogeneous goods categories (traded in organized exchanges and reference priced).

In some specifications, we exclude raw materials. Conversions from HS-6 products into raw and non-raw materials can be found in the WITS Classification tables.[3]

Finally, in some regressions we include country-pair variables that are taken from the

[2]Final goods thus comprise both capital and consumption goods. We drop from our analysis those goods that are considered "mixed' in the BEC classification.

[3]`https://wits.worldbank.org/referencedata.html`.

CEPII Gravity Dataset. These include the following: $Distance_{ij}$ is the bilateral distance (in kilometers, population weighted) between the capitals of countries i and j; $Contiguous_{ij}$ is a dummy variable that indicates whether i and j share a common border; $Common\ Language_{ij}$ is a dummy variable that is equal to 1 if the two countries share a common language; $Colony_{ij}$ is a dummy variable that indicates whether the two countries have a colonial link; RTA_{ijt} is a dummy variable that is equal to 1 if i and j are members of a regional trade agreement during in year t.

Descriptive statistics of the main variables that are used in our empirical analysis are reported in Table 1 below.

Table 1
Descriptive statistics of main variables

Variable	N	Mean	Median	Min	Max	Standard deviation
Imports_{ijkt}	77,577,263	1,424,317	27,329	1,000	1.68×10^{11}	55.3×10^{6}
Distance_{ij}	77,577,263	5,449	1,542	60.77	19,888.66	4,432
Intermediate_{k}	77,577,263	0.529	1	0	1	0.500
$\text{Differentiated}_{k}$	77,577,263	0.760	1	0	1	0.427
Contiguous_{ij}	77,577,263	0.097	0	0	1	0.295
$\text{Common Language}_{ij}$	77,577,263	0.175	0	0	1	0.380
Colony_{ij}	77,577,263	0.076	0	0	1	0.265
RTA_{ijt}	77,577,263	0.357	0	0	1	0.479

The table provides descriptive statistics for the main variables used in our empirical analysis, for the 1998-2011 period. See Section 3 for the definition and sources of all variables.

4 Empirical Methodology and Results

The empirical analysis is divided in three parts. First, we systematically examine whether intermediate goods are more sensitive to distance than are final goods, as it has been suggested by some previous studies.

In our benchmark regressions, we address this question by exploiting only cross-sectoral variation in bilateral trade flows. For different years in our panel, we estimate the following regression:

$$\begin{aligned} ln(Imports_{ijk}) &= \alpha + \beta_1 \ Intermediate_k \times ln(Distance_{ij}) + \beta_2 \ Intermediate_k \\ &\quad + \beta_3 \ ln(Distance_{ij}) + \beta_4 \ \mathbf{X}_{ij} + \delta_i + \delta_j + \delta_s + \epsilon_{ijk}, \end{aligned} \tag{1}$$

where $Imports_{kij}$ is the value of imports of HS6 product k of country i from country j, $\mathbf{X}_{ij}$ is a vector of bilateral variables (contiguous, common language, colonial relationship, membership in trade agreements), δ_i and δ_j denote fixed effects at the Importing-country and exporting-country level, and the δ_s are fixed effects that identify the broad sector (at the HS2 level) to which product k belongs. If intermediate goods are indeed more sensitive to distance than final goods (the omitted category), the coefficient β_1 should be negative and significant.

In the most demanding specifications, we include country-pair fixed effects δ_{ij} in (1) to account for the role of all bilateral determinants of trade flows. In these specifications, the variables $Distance_{ij}$ and $\mathbf{X}_{ij}$ are absorbed by the fixed effects.

Second, we re-estimate (1) but exclude raw materials. This allows us to rule out one of the possible mechanisms for the higher sensitivity to distance of intermediate goods that has been suggested by previous studies.

Third, we examine the role of product differentiation. If the mechanism that underlies the higher sensitivity to distance of intermediate goods is the complexity of global supply chains and the difficulty of sourcing inputs from suppliers that are located in far away countries, we would expect differentiated intermediate goods to be the most sensitive to distance. These are inputs that need to be tailored to the needs of final good producers, for which contracting frictions are more severe (Nunn, 2007). To verify this, we estimate

$$\begin{aligned} ln(Imports_{ijk}) &= \alpha + \beta_1 \textit{ Intermediate Differentiated}_k \times ln(Distance_{ij}) \\ &+\beta_2 \textit{ Final Differentiated}_k \times ln(Distance_{ij}) \\ &+\beta_3 \textit{ Intermediate Homogeneous}_k \times ln(Distance_{ij}) \\ &+\beta_4 \textit{ Intermediate Differentiated}_k + \beta_5 \textit{ Final Differentiated}_k \\ &+\beta_6 \textit{ Intermediate Homogeneous}_k + \delta_i + \delta_j + \delta_{ij} + \delta_s + \delta_s + \epsilon_{ijk}, \quad (2) \end{aligned}$$

where goods classified as *Final Homogeneous*$_k$ are the omitted category. We expect the coefficient β_1 to be negative and significant and to be larger in absolute terms than the coefficients β_2 and β_3.

4.1 Benchmark Results

In this section, we report the results of our benchmark regressions, in which we estimate (1) and (2) for three different years in our sample period (1998, 2004, 2010). In these regressions, we only exploit cross-sectional variation in bilateral trade flows to identify differences in the sensitivity to distance between intermediate and final goods and the role of product differentiation.

Table 2 reports the results of estimating (1). The key finding is that the coefficient of the interaction term *Intermediate*$_k$ $\times$ *ln(Distance*$_{ij}$) is negative and significant: imports of intermediate goods are more sensitive to distance than are imports of final goods. In terms of magnitude, the results of Table 2 show that the distance elasticity is between -0.017 and 0.029 percentage points more negative for intermediate goods as compared to final goods. Based on the specifications in which we omit country-country pair fixed effects (columns 1, 3, and 5), our estimates imply that the distance elasticity is between 5.84% and 7.75% larger for intermediate goods relative to final goods.

Table 2

Sensitivity to distance: imports of intermediate versus final goods

	(1)	(2)	(3)	(4)	(5)	(6)
$\text{Intermediate}_k \times \ln(\text{Distance}_{ij})$	-0.023***	-0.017*	-0.029***	-0.023***	-0.025***	-0.018**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)
Intermediate_k	0.180***	0.134*	0.167**	0.119*	0.193***	0.136*
	(0.054)	(0.055)	(0.054)	(0.055)	(0.053)	(0.054)
$\ln(\text{Distance}_{ij})$	-0.341***		-0.374***		-0.428***	
	(0.015)		(0.015)		(0.014)	
Contiguous_{ij}	0.440***		0.379***		0.401***	
	(0.050)		(0.045)		(0.043)	
$\text{Common Language}_{ij}$	0.109***		0.092***		0.122***	
	(0.032)		(0.026)		(0.027)	
Colony_{ij}	0.130***		0.118***		0.134***	
	(0.034)		(0.033)		(0.035)	
RTA_{ijt}	0.230***		0.153***		0.122***	
	(0.032)		(0.028)		(0.023)	
Exporting-country FE	Yes	Yes	Yes	Yes	Yes	Yes
Importing-country FE	Yes	Yes	Yes	Yes	Yes	Yes
Country-pair FE	No	Yes	No	Yes	No	Yes
HS2 FE	Yes	Yes	Yes	Yes	Yes	Yes
Year	1998	1998	2004	2004	2010	2010
N	4,785,880	4,783,424	5,544,965	5,541,660	6,098,518	6,095,428
R^2	0.197	0.231	0.189	0.222	0.206	0.240

The table reports the coefficients of OLS regressions. The dependent variable is the log of $Imports_{ijk}$, the value of imports of HS6 product k of country i from country j. Standard errors in parentheses are clustered at the country-pair level. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The results also suggest that intermediate goods are more traded than final goods (in terms of the value of imports), but only between countries that are relatively close to each other. The overall difference in the propensity to trade intermediate vs final goods can be seen by summing up the coefficient of the dummy variable $Intermediate_k$ with the coefficient of the interaction term $Intermediate_k \times ln(Distance_{ij})$.[4]

[4] From equation (1), the effect of intermediates on imports is $\frac{\partial ln(Imports_{ijk})}{Intermediate_k} = \beta_1 \times ln(Distance)_{ij} + \beta_2$. It is straightforward to verify that this effect becomes negative when the distance between countries i and j is above a certain threshold. In the specification of column 1 of Table 2, this threshold is around 2,500 km, implying that countries that are further apart trade more in final goods than in intermediate goods.

Our estimates also confirm standard results in the gravity literature on the role of other bilateral determinants of trade flows: countries trade more when they share a common border or a common language, have a historical colonial relationship, or are members of regional trade agreements.

In Table 3, we reproduce Table 2, but we exclude products that are classified as raw materials from our sample of HS6 goods. As we mention above, some earlier studies point out that many intermediate inputs are raw materials and that the bulky nature of these goods could explain why they are less easily traded across long distances. Raw materials represent 16% of all bilateral trade flows. The majority of HS6 products that are classified as raw materials are indeed intermediate goods.[5]

The coefficient of $Intermediate_k \times ln(Distance_{ij})$ remains negative and significant: the higher sensitivity to distance of intermediate goods is not driven by the bulky nature of many inputs. In fact, the exclusion of raw materials increases the gap between final and intermediate goods (the coefficient of $Intermediate_k \times ln(Distance_{ij})$ is uniformly larger in absolute terms and more significant than in Table 2: based on the specifications in columns 1, 3, and 5, the estimates imply that the distance elasticity is between 7.06% and 9.11% greater for intermediate goods relative to final goods.

Comparing Tables 2 and 3 suggests that, although most raw materials are intermediate inputs, these are generally less sensitive to distance, contrary to what has been suggested in previous studies (Miroudot *et al.*, 2009). This could be because many goods classified as raw materials are natural resources (e.g., oil and minerals), which can only be sourced only from a few countries in the world. Moreover, contracting frictions are less likely to be a serious

[5]Raw materials correspond to 575 HS6 codes in our sample. Of these, 379 are intermediate goods: some are homogeneous (e.g., oil, fertilizers, copper), while others are differentiated (e.g., glass containers, live animals, plants and parts, including seeds and fruits). The remaining products are final goods: some are homogeneous (e.g., frozen fish fillets, frozen shrimps and prawns), while others are differentiated (e.g., fish meat, mackerel).

problems in sourcing raw materials, since most of them are homogeneous intermediate goods, which do not need to be tailored to the specific needs of final good producers.

Table 3

Sensitivity to distance: imports of intermediate versus final goods, excluding raw materials

	(1)	(2)	(3)	(4)	(5)	(6)
$\text{Intermediate}_k \times \ln(\text{Distance}_{ij})$	-0.031***	-0.024***	-0.036***	-0.030***	-0.032***	-0.025***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Intermediate_k	0.231***	0.188***	0.219***	0.172**	0.243***	0.186***
	(0.055)	(0.056)	(0.054)	(0.055)	(0.054)	(0.055)
$\ln(\text{Distance}_{ij})$	-0.363***		-0.395***		-0.453***	
	(0.016)		(0.016)		(0.015)	
Contiguous_{ij}	0.450***		0.383***		0.402***	
	(0.052)		(0.046)		(0.044)	
$\text{Common Language}_{ij}$	0.131***		0.115***		0.149***	
	(0.033)		(0.027)		(0.027)	
Colony_{ij}	0.139***		0.124***		0.139***	
	(0.035)		(0.034)		(0.036)	
RTA_{ijt}	0.240***		0.164***		0.125***	
	(0.033)		(0.029)		(0.024)	
Exporting-country FE	Yes	Yes	Yes	Yes	Yes	Yes
Importing-country FE	Yes	Yes	Yes	Yes	Yes	Yes
Country-pair FE	No	Yes	No	Yes	No	Yes
HS2 FE	Yes	Yes	Yes	Yes	Yes	Yes
Year	1998	1998	2004	2004	2010	2010
N	4,517,900	4,515,509	5,226,087	5,222,769	5,749,171	5,746,025
R^2	0.207	0.242	0.198	0.233	0.216	0.252

The table reports the coefficients of OLS regressions. The dependent variable is the log of $Imports_{ijk}$, the value of imports of HS6 product k of country i from country j. Standard errors in parentheses are clustered at the country-pair level. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

ln

Table 4

Sensitivity to distance: imports of intermediate versus final goods, the role of product differentiation

	(1)	(2)	(3)	(4)	(5)	(6)
Intermediate Differentiated$_k$ × ln(Distance$_{ij}$)	-0.117***	-0.101***	-0.118***	-0.112***	-0.094***	-0.078***
	(0.018)	(0.018)	(0.014)	(0.014)	(0.015)	(0.015)
Final Differentiated$_k$ × Distance$_{ij}$	-0.061***	-0.051**	-0.049***	-0.047***	-0.030*	-0.02
	(0.018)	(0.018)	(0.014)	(0.014)	(0.014)	(0.014)
Intermediate Homogeneous$_k$ × Distance$_{ij}$	-0.002	0.012	0.009	0.02	0.026*	0.043***
	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)	(0.012)
Intermediate Differentiated$_k$	0.785***	0.641***	0.734***	0.663***	0.555***	0.416***
	(0.148)	(0.148)	(0.112)	(0.112)	(0.121)	(0.12)
Final Differentiated$_k$	0.357*	0.252	0.269*	0.228*	0.072	-0.026
	(0.142)	(0.142)	(0.113)	(0.114)	(0.113)	(0.114)
Intermediate Homogeneous$_k$	0.119	-0.035	0.020	-0.114	-0.101	-0.279**
	(0.106)	(0.105)	(0.096)	(0.095)	(0.098)	(0.097)
ln(Distance$_{ij}$)	-0.287***		-0.329***		-0.400***	
	(0.020)		(0.018)		(0.018)	
Contiguous$_{ij}$	0.448***		0.387***		0.408***	
	(0.051)		(0.045)		(0.043)	
Common Language$_{ij}$	0.109***		0.092***		0.123***	
	(0.032)		(0.026)		(0.027)	
Colony$_{ij}$	0.127***		0.114***		0.131***	
	(0.034)		(0.033)		(0.035)	
RTA$_{ijt}$	0.230***		0.152***		0.121***	
	(0.032)		(0.028)		(0.023)	
Test Intermediate Differentiated$_k$ × ln(Distance$_{ij}$) = Final Differentiated$_k$ × Distance$_{ij}$ (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Test Intermediate Differentiated$_k$ × ln(Distance$_{ij}$) = Intermediate Homogeneous$_k$ × Distance$_{ij}$ (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Test Final Differentiated$_k$ × Distance$_{ij}$ = Intermediate Homogeneous$_k$ × Distance$_{ij}$ (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Exporting-country FE	yes	yes	yes	yes	yes	
Importing-country FE	yes	yes	yes	yes	yes	
Country-pair FE	no	yes	no	yes	no	yes
HS2 FE	yes	yes	yes	yes	yes	yes
Year	1998	1998	2004	2004	2010	2010
N	4,785,880	4,783,424	5,544,965	5,541,660	6,098,518	6,095,428
R^2	0.198	0.232	0.191	0.224	0.207	0.241

The table reports the coefficients of OLS regressions. The dependent variable is the log of $Imports_{ijk}$, the value of imports of HS6 product k of country i from country j. Standard errors in parentheses are clustered at the country-pair level. At the bottom of the tables are p-values of t-tests. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Finally, in Table 4 we report the results of estimating (2). As expected, differentiated intermediate goods are the most sensitive to distance: the coefficient of the interaction term *Intermediate Differentiated*$_k$ $\times$ *ln(Distance*$_{ij}$*)* is negative and significantly larger in absolute terms than the coefficient of the other interaction terms (see the p-value of the tests reported at the bottom of the table). The results of Table 4 support the idea that contracting frictions along supply chains can lead final good producers to source inputs from nearby suppliers – particularly when the inputs need to be tailored to their specific needs.

In terms of magnitude, taking into account the role of product differentiation increases the gap in the sensitivity to distance of intermediate and final goods. For example, the estimates that are reported in column 1 of Table 4 imply that the distance elasticity is 0.117% percentage points more negative for differentiated intermediate goods relative to homogeneous final goods, which implies that the distance elasticity is 40.76% higher.

4.2 Robustness Checks

In the remainder of this section, we discuss the results of a series of additional estimations that we have carried out to verify the robustness of our results.

Panel regressions

In the benchmark regressions of Tables 2-4, the gap in the sensitivity to distance between intermediate and final goods and the role of product differentiation are identified by exploiting only cross-sectional variation in bilateral trade flows.

In what follows, we verify that the results continue to hold when we pool all of the years in our sample period (1998-2011), including year fixed effects in (1) and (2) to account for macroeconomic conditions that may affect imports.

The results are reported in Tables 5-7. The negative and significant coefficient on the interaction between *Intermediate*$_k$ and *ln(Distance*$_{ij}$*)* in Table 5 confirms that intermediate goods are more sensitive to distance than final goods. The results of Table 6 show that the gap between intermediate and final goods in their sensitivity to distance is larger and more significant if we exclude raw materials from the sample of products that are included in our analysis. Finally, Table 7 confirms that differentiated imports of intermediate goods are the most sensitive to distance: contracting frictions play a key role in shaping sourcing decisions.

Table 5

Sensitivity to distance: imports of intermediate versus final goods (panel regressions)

	(1)	(2)	(3)	(4)
$\text{Intermediate}_k \times \ln(\text{Distance}_{ij})$	-0.021***	-0.024***	-0.015*	-0.017**
	(0.006)	(0.006)	(0.006)	(0.006)
Intermediate_k	0.186***	0.153**	0.123*	0.102*
	(0.049)	(0.049)	(0.052)	(0.051)
$\ln(\text{Distance})_{ij}$	-0.380***	-0.384***		
	(0.013)	(0.014)		
Contiguous_{ij}	0.402***	0.402***		
	(0.044)	(0.044)		
$\text{Common Language}_{ij}$	0.096***	0.103***		
	(0.025)	(0.026)		
Colony_{ij}	0.106***	0.124***		
	(0.032)	(0.032)		
RTA_{ijt}	0.151***	0.153***		
	(0.022)	(0.023)		
Exporter x year FE	Yes	Yes	Yes	Yes
Importer x year FE	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	No
HS 2-digit x year FE	No	Yes	No	Yes
Country pair x year FE	No	No	Yes	Yes
N	77,577,263	77,496,670	77,533,984	77,453,869
R^2	0.171	0.202	0.205	0.236

The table reports the coefficients of OLS regressions. The dependent variable is the log of $Imports_{k,i,j,t}$, the value of imports of HS6 product k of country i from country j in year t. The panel covers the 1998-2011 period. Standard errors in parentheses are clustered at the country-pair level. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 6

Sensitivity to distance: imports of intermediate versus final goods, excluding raw materials (panel regressions)

	(1)	(2)	(3)	(4)
Intermediate$_k$ × ln(Distance$_{ij}$)	-0.028***	-0.031***	-0.023***	-0.024***
	(0.006)	(0.006)	(0.006)	(0.006)
Intermediate$_k$	0.246***	0.204***	0.179***	0.154**
	(0.05)	(0.049)	(0.052)	(0.051)
ln(Distance$_{ij}$)	-0.400***	-0.406***		
	(0.014)	(0.014)		
Contiguous$_{ij}$	0.406***	0.407***		
versus	(0.044)	(0.045)		
Common Language$_{ij}$	0.120***	0.127***		
	(0.026)	(0.026)		
Colony$_{ij}$	0.110***	0.130***		
	(0.033)	(0.034)		
RTA$_{ijt}$	0.157***	0.160***		
	(0.023)	(0.023)		
Exporter x year FE	Yes	Yes	Yes	Yes
Importer x year FE	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	No
HS 2-digit x year FE	No	Yes	No	Yes
Country pair x year FE	No	No	Yes	Yes
N	73,173,076	73,092,483	73,129,720	73,049,703
R^2	0.18	0.211	0.215	0.247

The table reports the coefficients of OLS regressions. The dependent variable is the log of $Imports_{ijkt}$, the value of imports of HS6 product k of country i from country j in year t. The panel covers the 1998-2011 period. Standard errors in parentheses are clustered at the country-pair level. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 7

Sensitivity to distance: imports of intermediate versus final goods, the role of product differentiation (panel regressions)

	(1)	(2)	(3)	(4)
Intermediate Differentiated$_k$ $\times$ ln(Distance$_{ij}$)	-0.100***	-0.109***	-0.088***	-0.097***
	(0.014)	(0.014)	(0.014)	(0.014)
Final Differentiated$_k$ $\times$ Distance$_{ij}$	-0.048***	-0.048***	-0.040**	-0.041**
	(0.014)	(0.014)	(0.014)	(0.014)
Intermediate Homogeneous$_k$ $\times$ Distance$_{ij}$	0.004	0.013	0.016	0.026*
	(0.011)	(0.011)	(0.011)	(0.011)
Intermediate Differentiated$_k$	0.624***	0.683***	0.584***	0.570***
	(0.113)	(0.113)	(0.114)	(0.113)
Final Differentiated$_k$	0.217	0.248*	0.212	0.175
	(0.113)	(0.110)	(0.114)	(0.111)
Intermediate Homogeneous$_k$	-0.037	-0.007	-0.146	-0.155
	(0.091)	(0.09)	(0.09)	(0.089)
ln(Distance$_{ij}$)	-0.332***	-0.340***		
	(0.016)	(0.017)		
Contiguous$_{ij}$	0.408***	0.409***		
	(0.044)	(0.044)		
Common Language$_{ij}$	0.097***	0.103***		
	(0.025)	(0.026)		
Colony$_{ij}$	0.103**	0.121***		
	(0.032)	(0.032)		
RTA$_{ijt}$	0.150***	0.153***		
	(0.022)	(0.023)		
Test Intermediate Differentiated$_k$ $\times$ ln(Distance$_{ij}$) = Final Differentiated$_k$ $\times$ Distance$_{ij}$ (p-value)	0.000	0.000	0.000	0.000
Test Intermediate Differentiated$_k$ $\times$ ln(Distance$_{ij}$) = Intermediate Homogeneous$_k$ $\times$ Distance$_{ij}$ (p-value)	0.000	0.000	0.000	0.000
Test Final Differentiated$_k$ $\times$ Distance$_{ij}$ = Intermediate Homogeneous$_k$ $\times$ Distance$_{ij}$ (p-value)	0.000	0.000	0.000	0.000
Exporter x year FE	Yes	Yes	Yes	Yes
Importer x year FE	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	No
HS 2-digit x year FE	No	Yes	No	Yes
Country pair x year FE	No	No	Yes	Yes
N	77,496,671	77,496,670	77,453,870	77,453,869
R^2	0.173	0.203	0.206	0.237

The table reports the coefficients of OLS regressions. The dependent variable is the log of $Imports_{ijk}$, the value of imports of HS6 product k of country i from country j. The panel covers the 1998-2011 period. Standard errors in parentheses are clustered at the country-pair level. At the bottom of the tables are p-values of t-tests. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Alternative econometric model

The results above are estimated with the use of ordinary least squares (OLS) regressions. Santos Silva and Tenreyro (2006) point out that log-linearized models estimated with OLS can be misleading in the presence of heteroskedasticity. They proposed as a solution to use a Poisson pseudo-maximum likelihood (PPML) estimator. In Tables 8-10 we reproduce Tables 2-4 using the PPML estimator.

In Table 8 the coefficient of $\mathit{Intermediate}_k \times \mathit{ln(Distance}_{ij})$ is negative and statistically significant only in the first two columns. The estimates in column 1 indicate that the distance elasticity is 20.46% higher intermediate goods relative to final goods.

Table 9 confirms that excluding raw materials increases the gap between final and intermediate goods (the coefficient of $\mathit{Intermediate}_k \times \mathit{ln(Distance}_{ij})$ is larger in absolute value and more significant than in Table 2). Based on the specifications in columns 1, 3, and 5, the estimates imply that the distance elasticity is between 26.67% and 40.23% higher for intermediate goods relative to final goods.

Finally, the results of Table 10 confirm that differentiated intermediate goods are more sensitive to distance as compared to homogeneous final goods (the coefficient of the interaction $\mathit{Intermediate\ Differentiated}_k \times \mathit{ln(Distance}_{ij})$ is negative and significant), differentiated final goods (see p-value of the first test at the bottom of the table), and homogeneous intermediate goods (see p-value of the second test at the bottom of the table). These results suggest that contracting frictions along value chains may deter final good producers from sourcing intermediate inputs from distance suppliers, particularly when the inputs need to be adjusted to their needs.

In our analysis, we have focused on positive import values. Santos Silva and Tenreyro (2011) show that, besides being consistent in the presence of heteroskedasticity, the PPML method can also deal with zero values of the dependent variable. However, in their Monte Carlo simulations, they allow zeros to account for between 62 and 83 percent of the observations. In our product-level regressions, zeros would instead account for almost the totality of the observations in our sample.[6]

Table 8

Sensitivity to distance: imports of intermediate versus final goods (PPML)

$\text{Intermediate}_k \times \ln(\text{Distance}_{ij})$	-0.053*	-0.064*	-0.010	-0.031	0.013	-0.006
	(0.027)	(0.027)	(0.047)	(0.048)	(0.052)	(0.054)
Intermediate_k	0.530*	0.619**	-0.039	0.122	-0.022	0.136
	(0.215)	(0.216)	(0.361)	(0.371)	(0.404)	(0.415)
$\ln(\text{Distance}_{ij})$	-0.259***		-0.342***		-0.367***	
	(0.035)		(0.036)		(0.041)	
Contiguous_{ij}	0.507***		0.378***		0.386***	
	(0.059)		(0.058)		(0.071)	
$\text{Common Language}_{ij}$	0.083		-0.049		-0.092	
	(0.057)		(0.069)		(0.085)	
Colony_{ij}	-0.010		-0.032		0.016	
	(0.068)		(0.073)		(0.077)	
RTA_{ijt}	0.614***		0.471***		0.282***	
	(0.063)		(0.060)		(0.062)	
Exporting-country FE	Yes	Yes	Yes	Yes	Yes	Yes
Importing-country FE	Yes	Yes	Yes	Yes	Yes	Yes
Country-pair FE	No	Yes	No	Yes	No	Yes
HS2 FE	Yes	Yes	Yes	Yes	Yes	Yes
Year	1998	1998	2004	2004	2010	2010
N	4,785,880	4,783,424	5,544,965	5,541,660	6,098,518	6,095,428
Pseudo R^2	0.312	0.342	0.326	0.364	0.348	0.392

The table reports the coefficients of PPML regressions. The dependent variable is the log of $Imports_{ijk}$, the value of imports of HS6 product k of country i from country j. Standard errors in parentheses are clustered at the country-pair level. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

[6] For example, if we included zeros in column 1 of Table 8, they would account for almost 98% of the sample (the number of observations would increase from 4,785,880 to 222,746,370). We have nevertheless tried to reproduce Tables 8-10, with the inclusion of zeros in the dependent variable. The results confirm that intermediate goods are more sensitive to distance than are final goods – particularly when eliminating raw materials. However, the role of product differentiation is less clear-cut (intermediate goods are more sensitive to distance than only two of the three other categories of products).

Table 9

Sensitivity to distance: imports of intermediate versus final goods, excluding raw materials (PPML)

$\text{Intermediate}_k \times \ln(\text{Distance}_{ij})$	-0.107***	-0.114***	-0.119**	-0.127**	-0.108*	-0.117*
	(0.026)	(0.026)	(0.040)	(0.042)	(0.045)	(0.046)
Intermediate_k	0.967***	1.021***	0.831**	0.894**	0.957**	1.028**
	(0.206)	(0.210)	(0.307)	(0.319)	(0.343)	(0.355)
$\ln(\text{Distance}_{ij})$	-0.266***		-0.368***		-0.405***	
	(0.035)		(0.033)		(0.034)	
Contiguous_{ij}	0.494***		0.350***		0.322***	
	(0.061)		(0.061)		(0.061)	
$\text{Common Language}_{ij}$	0.118*		-0.011		0.006	
	(0.057)		(0.073)		(0.076)	
Colony_{ij}	0.029		-0.004		0.084	
	(0.067)		(0.078)		(0.079)	
RTA_{ijt}	0.641***		0.451***		0.263***	
	(0.065)		(0.061)		(0.051)	
Exporting-country FE	Yes	Yes	Yes	Yes	Yes	Yes
Importing-country FE	Yes	Yes	Yes	Yes	Yes	Yes
Country-pair FE	No	Yes	No	Yes	No	Yes
HS2 FE	Yes	Yes	Yes	Yes	Yes	Yes
Year	1998	1998	2004	2004	2010	2010
N	4,517,900	4,515,509	5,226,087	5,222,769	5,749,171	5,746,025
Pseudo R^2	0.319	0.346	0.303	0.337	0.317	0.355

The table reports the coefficients of PPML regressions. The dependent variable is the log of $Imports_{ijk}$, the value of imports of HS6 product k of country i from country j. Standard errors in parentheses are clustered at the country-pair level. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 10

Sensitivity to distance: imports of intermediate versus final goods, the role of product differentiation (PPML)

	(1)	(2)	(3)	(4)	(5)	(6)
Intermediate Differentiated$_k$ × ln(Distance$_{ij}$)	-0.116*	-0.097*	-0.159***	-0.139***	-0.147***	-0.130**
	(0.048)	(0.048)	(0.040)	(0.040)	(0.040)	(0.040)
Final Differentiated$_k$ × Distance$_{ij}$	0.025	0.048	0.020	0.046	0.041	0.059
	(0.049)	(0.049)	(0.053)	(0.054)	(0.056)	(0.057)
Intermediate Homogeneous$_k$ × Distance$_{ij}$	0.079*	0.076*	0.134**	0.127**	0.212***	0.199***
	(0.038)	(0.039)	(0.047)	(0.049)	(0.049)	(0.053)
Intermediate Differentiated$_k$	0.401	0.246	0.436	0.283	0.477	0.344
	(0.377)	(0.379)	(0.320)	(0.320)	(0.333)	(0.330)
Final Differentiated$_k$	-0.823*	-1.014**	-0.794	-1.000*	-1.025*	-1.175**
	(0.386)	(0.388)	(0.415)	(0.427)	(0.444)	(0.452)
Intermediate Homogeneous$_k$	-0.661*	-0.677*	-1.099**	-1.085**	-1.489***	-1.412***
	(0.301)	(0.303)	(0.370)	(0.383)	(0.382)	(0.407)
ln(Distance$_{ij}$)	-0.278***		-0.354***		-0.400***	
	(0.047)		(0.043)		(0.042)	
Contiguous$_{ij}$	0.514***		0.392***		0.393***	
	(0.059)		(0.059)		(0.071)	
Common Language$_{ij}$	0.082		-0.046		-0.088	
	(0.056)		(0.069)		(0.085)	
Colony$_{ij}$	-0.012		-0.037		0.012	
	(0.068)		(0.073)		(0.077)	
RTA$_{ijt}$	0.607***		0.462***		0.274***	
	(0.063)		(0.060)		(0.062)	
Test Intermediate Differentiated$_k$ × ln(Distance$_{ij}$) = Final Differentiated$_k$ × Distance$_{ij}$ (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Test Intermediate Differentiated$_k$ × ln(Distance$_{ij}$) = Intermediate Homogeneous$_k$ × Distance$_{ij}$ (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Test Final Differentiated$_k$ × Distance$_{ij}$ = Intermediate Homogeneous$_k$ × Distance$_{ij}$ (p-value)	0.271	0.573	0.089	0.257	0.028	0.092
Exporting-country FE	yes	yes	yes	yes	yes	
Importing-country FE	yes	yes	yes	yes	yes	
Country-pair FE	no	yes	no	yes	no	yes
HS2 FE	yes	yes	yes	yes	yes	yes
Year	1998	1998	2004	2004	2010	2010
N	4,785,880	4,783,424	5,544,965	5,541,660	6,098,518	6,095,428
Pseudo R^2	0.314	0.345	0.332	0.370	0.356	0.399

The table reports the coefficients of PPML regressions. The dependent variable is the log of $Imports_{ijk}$, the value of imports of HS6 product k of country i from country j. Standard errors in parentheses are clustered at the country-pair level. At the bottom of the tables are p-values of t-tests. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

5 Conclusions

The last few decades have seen a dramatic reduction in different types of trade costs: transport costs, communication costs, and tariff barriers. These changes have fostered the fragmentation of production processes across countries, which has led to the emergence of GVCs and a rise of trade in intermediate goods. Surprisingly, during this period, trade flows have not become less sensitive to distance.

In this paper, we show that the emergence of GVCs can help to explain the distance puzzle. We show that intermediate goods – and particularly differentiated inputs – are more sensitive to distance than are final goods. These results are robust to including different sets of fixed effects and controls, exploiting cross-sectional and time series variation in trade flows, and using alternative econometric methodologies.

Our analysis suggests that – although fragmenting production across firms and countries has become easier – contractual frictions remain a significant obstacle to the globalization of supply chains. Sourcing inputs from nearby countries can allow producers to coordinate better with suppliers and monitor their production. The use of more distant suppliers can give rise to problems and delays in the production of key inputs, which can disrupt the entire supply chain.

Our results are based on a panel dataset of bilateral trade flows, which covers around 5,000 products and a large number of countries over the 1998-2011 period. An important avenue for future research is to examine how distance and other bilateral determinants of trade flows – e.g., common language and membership in trade agreements – shape the sourcing decisions of individual firms.

Acknowledgements: We are grateful for the comments and suggestions of Peter Egger, Mathieu Parenti, André Sapir, and seminar and conference participants at ECARES, EUI, and Warsaw School of Economics. Paola Conconi gratefully acknowledges financial support from the FNRS and the ERC (project 834253 - TRASC) and the European Union's Horizon 2020 research and innovation programme (Agreement Grant no. 770680).

References

Alfaro, L., Antràs, P., Conconi, P., & Chor, D. (2019). Internalizing global value chains: A firm-level analysis. Journal of Political Economy, 127(2), 508-559.

Amiti, M., & Konings, J. (2007). Trade liberalization, intermediate inputs, and productivity: Evidence from Indonesia. American Economic Review, 97(5), 1611-1638.

Anderson, J. E., & Yotov, Y. (2010). The changing incidence of geography. American Economic Review, 100(5), 2157-2186.

Antràs, P., & Chor, D. (2013). Organizing the global value chain. Econometrica, 81(6), 2127-2204.

Antràs, P. (2016). Global production: Firms, contracts and trade structure. Princeton University Press.

Antràs, P., Fort, T. C., & Tintelnot, F. (2017). The Margins of global sourcing: Theory and evidence from US firms. American Economic Review, 107(9), 2514-64.

Baldwin, R. (2013). Global supply chains: Why they emerged, why they matter, and where they are going. In D. K. Elms & P. Low (Eds.), Global value chains in a changing world. World Trade Organization.

Baldwin, R., & Taglioni, D. (2011). Gravity chains: Estimating bilateral trade flows when parts and components trade is important. NBER Working Paper No. 16672.

Bergstrand, J., & Egger, P. (2010). A general equilibrium theory for estimating gravity equations of bilateral FDI, final goods trade and intermediate goods trade. In S.

Brakman & P. Van Bergeijk (Eds.), The gravity model in international trade: Advances and applications. Cambridge University Press, New York.

Bernard, A. B., Jensen, J. B., Redding, S. J., & Schott, P. K. (2007). Firms in international trade. Journal of Economic Perspectives, 21(3), 105-130.

Blaum, J., Lelarge, C., & Peters M. (2018). The gains from input trade with heterogeneous importers. American Economic Journal: Macroeconomics, 10(4), 77-127.

Cairncross, F. (1997). The death of distance: How the communications revolution will change our lives. Orion Business Books, London, U.K.

Combes, P-P., Mayer, T., & Thisse, J.-F. (2008). Economic geography: The integration of regions and nations. Princeton University Press, Princeton.

Conconi, P., Garcìa-Santana, M., Puccio, L. & Venturini, R. (2018). From final goods to inputs: The protectionist effect of rules of origin. American Economic Review, 108(8), 2335-2365.

Disdier, A.-C., & Head, K. (2008). The puzzling persistence of the distance effect on bilateral trade. Review of Economics and Statistics, 90(1), 37-48.

Glaeser, E.L., & Kohlhase, J.E. (2003). Cities, regions and the decline of transport costs. Papers in Regional Science, 83(1), 197-228.

Goldberg, P. K., Khandelwal, A. K., Pavcnik, N., & Topalova, P. (2010). Imported intermediate inputs and domestic product growth: Evidence from India. Quarterly Journal of Economics, 125(4), 1727-1767.

Imbs, J., & Mejean, I. (2017). Trade elasticities. Review of International Economics, 25(2), 383-402.

Johnson, R. C., & Noguera, G. (2012). Accounting for intermediates: Production sharing and trade in value added. Journal of International Economics, 86(2), 224-236.

Leamer, E.E., & Levinsohn, J. (1995). International trade theory: The evidence. In G. Grossman & K. Rogoff (Eds.), Handbook of international economics Volume 3. North-Holland, Amsterdam.

Miroudot, S., Lanz, R., & Ragoussis, A. (2009). Trade in intermediate goods and services. OECD Trade Policy Papers No. 93.

Nunn, N. (2007). Relationship-specificity, incomplete contracts and the pattern of trade. Quarterly Journal of Economics, 122(2), 569-600.

Rauch, J. E. (1999). Networks versus markets in international trade. Journal of International Economics, 48(1), 7-35.

Santos Silva, J., & Tenreyro, S. (2006). The log of gravity. Review of Economics and Statistics, 88(4), 641-658.

Santos Silva, J., & Tenreyro, S. (2011). Further simulation evidence on the performance of the poisson-PML estimator. Economics Letters, 112(2), 220-222.